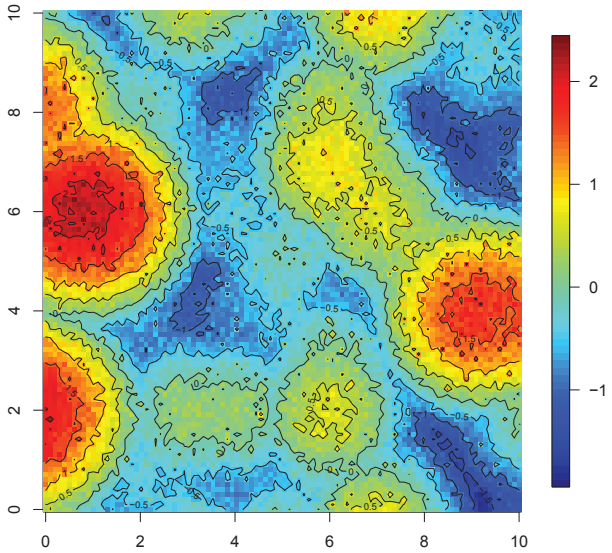
Modélisation spatiale de valeurs extrêmes : application à l'étude de précipitations en France

Quentin Sebille

Institut Camille Jordan
UMR 5208 du CNRS

Modélisation spatiale de valeurs extrêmes
Application à l'étude de précipitations en France



Quentin Sebille

Thèse de doctorat

Université Claude Bernard Lyon 1

Université Claude Bernard Lyon 1
École doctorale **InfoMath**, ED 512
Spécialité : **Mathématiques appliquées**
N. d'ordre 2016LYSE1244

Modélisation spatiale de valeurs extrêmes
Application à l'étude de précipitations en France

Thèse de doctorat

Soutenue publiquement le 1er décembre 2016 par

Quentin Sebille

devant le Jury composé de :

Mme Liliane Bel	Université Paris Sud	Rapportrice
M. Peter Craigmile	Ohio State University	Rapporteur
Mme Anne Dutfoy	EDF	Examinatrice
Mme Anne-Laure Fougères	Institut Camille Jordan (Lyon 1)	Directrice de thèse
M. Stéphane Girard	INRIA Grenoble Rhône-Alpes	Examinateur
Mme Cécile Mercadier	Institut Camille Jordan (Lyon 1)	Directrice de thèse
M. Benjamin Renard	IRSTEA Lyon	Examinateur

Remerciements

Cette thèse n'aurait jamais vu le jour sans le soutien et les conseils de mes deux directrices. Aussi je tiens à vous remercier très chaleureusement toutes les deux pour votre bienveillance et vos encouragements, tant du point de vue personnel que professionnel. Cécile, je suis venu te voir pour te demander si tu me pensais capable de faire une thèse et tu as su m'orienter vers ce travail qui me correspondait. Avec Anne-Laure, vous m'avez ensuite très rapidement convaincu de me lancer dans cette belle aventure et m'avez accompagné avec attention tout le long de ce projet. Vous avez été d'excellentes tutrices : vous me manquerez et je ne vous en remercierai jamais assez !

Je souhaite aussi exprimer ma gratitude envers les membres du jury qui ont accepté d'évaluer mon travail avec attention. Je suis d'autant plus reconnaissant envers Liliane et Peter d'avoir été les rapporteurs de mon manuscrit. Vos appréciations et vos suggestions ont été très valorisantes. Peter, merci beaucoup pour l'effort de traduction que tu as dû faire.

Lors de ces trois années de thèse, j'ai pu rencontrer plusieurs personnes qui m'ont aidé à avancer grâce leur conseils et/ou leurs encouragements. Je pense notamment aux chercheurs et chercheuses d'EDF : Anne, Sylvie, Marie, Thi-Thu, Pietro et Rémy. Mais aussi à celles et ceux que j'ai pu croiser lors de conférences : Anne, Mathieu, Aymeric, Jenny, Anne-Catherine, Dan, Roberto, Juliette, Nicolas et Davide. Je remercie également Brian Reich et Ben Shaby qui m'ont encouragé à écrire mon package et ont su répondre à mes questions concernant leur modèle. Enfin, je tiens à saluer et remercier plusieurs collègues de l'ICJ : Thierry et Violaine pour leur aide sur mes problèmes de code, Laurent et Vincent pour avoir résolu mes soucis d'ordre informatique, Aurélie et Céline (Laurent) pour l'aspect administratif, Sylvie pour m'avoir confié la gestion des commandes de café, ainsi que Gilles avec qui j'ai partagé le stage de Master. Mais aussi Céline (Vial), Véronique, Clément, Gabriela, Thierry, Anne, François et Ivan, et les doctorant-e-s : Luigia, Antoine, Benoît, Michele, Manaf, Simon, Cécilia, Colin, les Maxime, les François, Nils, Agathe, Mathias, Corentin, Niccolo et tous ceux que j'aurais pu oublier !

J'ai également reçu beaucoup de soutien de la part de mes amis, qui méritent bien évidemment d'être cités ici : mes anciens collocs Yoann et Augustin pour toutes ces soirées qui m'ont permis de déconnecter, Nico qui redescend de sa capitale pour partager whisky et cigares à chaque vacance scolaire, Armand, Angélique et Jib pour les soirées running (qui n'ont pas duré tant que ça), Vincent et Paul pour s'occuper de notre Yoann depuis qu'il ne vit plus avec nous ; Luigia et Guillaume (qui ont toujours notre escabeau !) ; Florence, Steven (et Benjamin !), Laurie, Camille, Nonor, Pef (à qui je dois une revanche à la coinche) ; Cynthia et Pauline (pour les bons souvenirs de Master en cours d'opti avec Augustin, la verveine et les babés), Annabelle, Elsa et Amandine ; et enfin mes amis de lycée Kevin, Corentin et Joëlle.

Merci également à mes parents, qui ont cru en moi pendant tout ce temps et qui m'ont encouragé à me lancer dans cette thèse. Merci aussi à mon frère Hugo et à ma sœur Hélène, mais aussi à mes grands-parents, à Marie-Do, Thierry mon parrain, Ambroise et Flo, chacun pour leur intéressement à ce que je faisais et pour leurs encouragements. Merci aussi à ma petite cousine Sarah, que j'embrasse très fort ! Et aussi un grand merci à mes cousins Benoît et Morgane pour m'avoir accueilli à Paris quand j'avais un séminaire, en partageant avec moi un bon petit plat végétarien accompagné d'un verre d'eau (lol). Merci aussi à Alain et Fabienne de venir écouter ma soutenance et de partager ce moment avec moi !
Il me faut aussi remercier Bizkit, qui m'a écouté répéter mes exposés et ma soutenance, qui a su me réconforter quand le moral était au plus bas, en utilisant force câlins et ronronnements (sous réserve qu'il y ait assez de croquettes dans la gamelle...) !

Et pour terminer par le meilleur : Pauline, tu as partagé ma vie pendant ces trois ans et tu as vécu cette thèse au moins autant que moi, tout en m'apportant un énorme soutien. Je te remercie du fond du cœur d'avoir été à mes côtés durant les meilleurs et les pires moments de ces trois ans qui auront marqué notre vie : ta présence m'a donné la force de finir ce beau projet :) Je t'aime.

Table des matières

Table des figures

Liste des tableaux

Introduction

Événements extrêmes de précipitations en France

Les catastrophes naturelles causent chaque année à travers le monde la perte de vies humaines ainsi que de nombreux dégâts en infrastructures. La plupart de ces événements sont liés à des phénomènes naturels *extrêmes*, c'est-à-dire d'intensité inhabituellement forte voire inédite. C'est par exemple le cas des canicules, des inondations, des tempêtes, des tsunamis ou même des pics de pollution.

Du fait du réchauffement climatique, ces catastrophes naturelles sont susceptibles de devenir de plus en plus fréquentes et/ou de plus en plus violentes.

En France, les pluies extrêmes sont à l'origine d'inondations qui peuvent provoquer des dégâts catastrophiques aussi bien dans les zones habitées que dans les campagnes.

On peut citer trois catastrophes liées aux précipitations extrêmes qui ont eu lieu pendant la réalisation de cette thèse :

1. Du 17 au 30 novembre 2014, plusieurs épisodes orageux extrêmes dans le sud de la France ont provoqué une longue série d'inondations qui ont touché les départements des Pyrénées Orientales, de l'Hérault, du Gard, du Var et des Alpes-Maritimes et ont provoqué la mort de 17 personnes. En particulier, pendant cette série de catastrophes, le massif des Cévennes a été touché par un fort épisode pluvieux qui a duré du 9 octobre au 13 novembre 2014. La figure 1 montre sur un schéma de Météo France le cumul de précipitations durant cet épisode. Jusqu'à 559mm d'eau sont tombés à Barnas dans l'Ardèche pendant ces quatre jours. Il faut faire remarquer que ce type d'enregistrement n'est pas rare pour la région : il s'agit des *épisodes cévenols*. C'est l'accumulation des épisodes pluvieux qui relèvent dans ce cas du caractère extrême.
2. Le samedi 3 octobre 2015, les Alpes-Maritimes ont été touchées par de violents orages extrêmement pluvieux qui se sont formés en l'espace de deux heures avec un total de 174mm de précipitations à Cannes, record enregistré par la station météorologique en place depuis 1949. Le bilan humain de cette catastrophe est de 20 morts et 2 disparus. Au total, 600 millions d'euros ont été estimés en dommages matériels.
3. En 2016, un épisode de pluies intenses et d'inondations ont frappé l'Europe centrale entre le 28 mai et le 7 juin. En France, ces fortes pluies ont concerné les départements du Loiret, du Cher, de l'Essonne, de la Seine-et-Marne et de l'Yonne et ont provoqué la mort de 4 personnes au total. Selon l'Association Française de l'Assurance (AFA), il s'agit de l'inondation la plus coûteuse de l'histoire du régime d'assurance des catastrophes naturelles, avec des pertes estimées entre 900 millions et 1.4 milliards d'euros.

Le phénomène de pluies peut être considéré comme les réalisations d'une variable aléatoire qui possède à la fois un aspect temporel et un aspect spatial. Avec des outils statistiques adaptés, il est alors possible de prédire de manière probabiliste le type d'événements catastrophiques listés ci-dessus, afin d'anticiper un scénario de catastrophe naturelle et se protéger de ses conséquences, en construisant par exemple des infrastructures capables de résister aux intempéries extrêmes.

En pratique cependant, il est nécessaire d'observer des événements similaires afin d'estimer la probabilité que le même scénario se reproduise. Cette condition est en contradiction directe avec le caractère rare de l'événement extrême.

Pour cela, on fait appel à la théorie statistique des valeurs extrêmes, qui permet d'*extrapoler* le comportement extrémal d'une variable aléatoire à partir des observations faites au quotidien. Grâce aux outils proposés par ce domaine de la statistique relativement récent et en pleine activité du point de vue de la recherche scientifique, il est donc possible de donner une estimation probabiliste d'un risque associé à un phénomène aléatoire tel que les catastrophes naturelles.

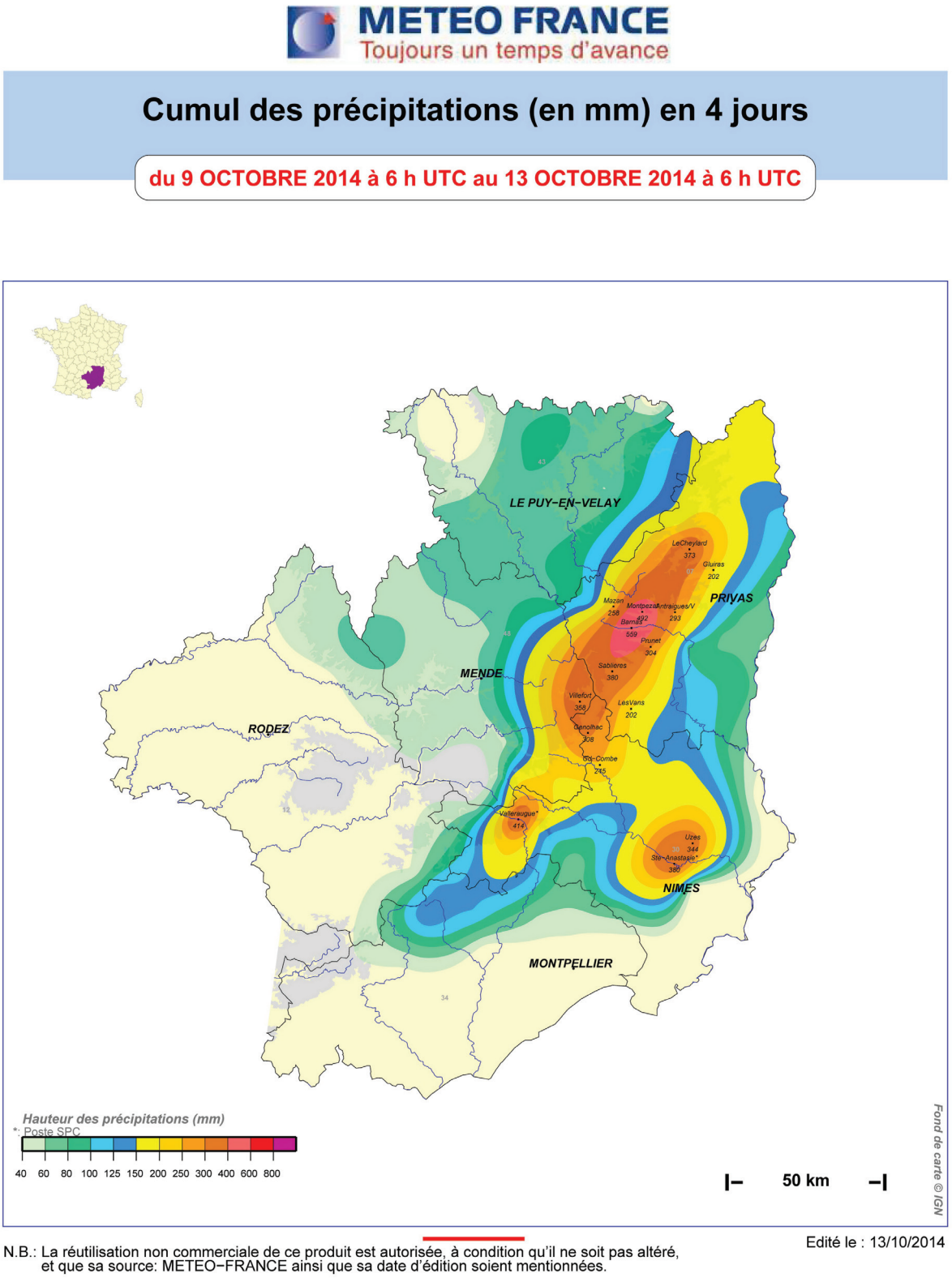


FIGURE 1 – Cumul de pluies de l'épisode cévenol enregistré du 9 au 13 octobre 2014. Source : Météo France.

Points développés dans la thèse

Cette thèse est financée par EDF R&D dans le cadre du projet MADONE (Méthodes pour les Agressions D'Origine Naturelle Externe). Depuis fin 2011, EDF s'est engagée sur le sujet de la modélisation des valeurs extrêmes multivariées et spatiales de phénomènes naturels telles que les précipitations ou encore les températures. Un des intérêts d'EDF est l'étude de risque associé aux précipitations extrêmes en France sur les structures telles que les barrages ou les centrales nucléaires.

Ces travaux ont donc de multiples objectifs liés à la modélisation statistique des précipitations extrêmes en France.

Modélisation de la dépendance spatiale des précipitations extrêmes

Une première problématique concerne la modélisation des données de précipitations extrêmes observées sur plusieurs stations météorologiques, c'est-à-dire les lieux où les données sont récoltées.

Les précipitations possèdent un caractère spatial, c'est-à-dire en particulier que les observations entre plusieurs stations sont corrélées en fonction de la distance qui les sépare. Plus cette distance est grande, moins la dépendance entre les deux stations est forte.

Il est important de sélectionner des modèles issus de la théorie des valeurs extrêmes qui tiennent compte de la dépendance spatiale entre les différentes positions des stations. En effet, ignorer la dépendance spatiale entre les observations entraîne généralement une mauvaise estimation du risque de catastrophe. Le choix d'un modèle statistique adapté pour décrire les précipitations extrêmes est donc une étape importante pour la prédiction.

Les travaux de la thèse se sont concentrés sur plusieurs problèmes associés à des mesures de risques multivariés pour des précipitations extrêmes. Deux questions en particulier ont suscité la mise en place de ces études :

1. Quelle est la probabilité que plusieurs stations météorologiques soient touchées, de façon simultanée ou partielle, par un fort cumul de pluies ?
2. Sachant qu'une ou plusieurs stations sont frappées par un événement extrême, comment évaluer le risque de propagation du phénomène extrême en d'autres positions ?

Le risque de précipitations extrêmes est généralement mesuré à travers une valeur appelée le *niveau de retour* à T-années, qui correspond à la valeur dépassée en moyenne une fois toutes les T années par le maximum annuel du phénomène observé. Dans le cas qui nous intéresse, c'est un niveau de retour centennal ($T = 100$) qui est évalué. Une définition plus formelle de cette quantité est donnée dans le chapitre 1. Les deux questions précédentes peuvent être reformulées en utilisant cette mesure :

1. Estimer la probabilité que sur un ensemble de stations météorologiques, un événement soit observé dont l'intensité dépasse celle du niveau de retour centennal.
2. Estimer la probabilité qu'une ou plusieurs positions connaissent un dépassement du niveau de retour centennal sur un jour donné, sachant que c'est le cas sur une partie du réseau de stations météorologiques.

Le premier objectif est étudié dans la partie II de la thèse, avec l'utilisation des processus max-stables. Le second objectif est investigué grâce aux modèles de processus Pareto dans la partie III.

Extrapolation spatiale du comportement extrême

La deuxième problématique posée par la thèse concerne la situation dans laquelle la position où l'on souhaite évaluer un risque de fort cumul de précipitations n'est pas jaugé. Autrement dit, aucune donnée n'est enregistrée en ce point.

En pratique, il est en effet impossible de récolter des données de précipitations sur l'ensemble d'une région et encore moins sur tout le territoire français et on dispose donc d'un réseau de stations météorologiques qui enregistrent des données à l'endroit précis où elles sont positionnées. Cependant, il peut être requis de fournir une mesure de risque en des endroits qui ne correspondent pas à une station météorologique.

Si le réseau de stations météorologiques est suffisamment dense et qu'une station météorologique est proche du site d'intérêt, une solution possible est d'analyser les valeurs extrêmes de cette station et de supposer que les mesures de risque sont égales entre ces deux positions.

L'utilisation d'un modèle spatial permet généralement de considérer l'information sur toute une région autour de la position non jaugée concernée et de réaliser une prédiction de la mesure de risque. Des outils de prédiction spatiale, issus par exemple du domaine de la géostatistique peuvent être envisagés. On peut également prédire la valeur du phénomène observé ayant un caractère spatial en des points qui ne correspondent pas aux données observées en conditionnant par les observations obtenues sur les stations météorologiques.

Plan de la thèse

Afin de répondre aux différentes problématiques, la thèse se présente en trois parties.

La partie I introduit les outils nécessaires à l'analyse des extrêmes de précipitations journalières utilisés dans la thèse en présentant plusieurs éléments théoriques appartenant aux domaines statistiques :

- de la théorie des valeurs extrêmes (Chapitre 1), dans le cas univarié (Section 1.1) et multivarié (Section 1.2),
- de la géostatistique (Chapitre 2),
- de l'approche inférentielle bayésienne (Chapitre 3) avec une présentation des algorithmes MCMC (Section 3.2).

Le dernier pan (Chapitre 4) présente les données de précipitations journalières françaises utilisées dans la thèse en analysant le caractère non-stationnaire (du point de vue temporel et spatial) du phénomène aléatoire associé.

La partie II se concentre sur la modélisation des valeurs maximales annuelles d'un processus spatial. Plusieurs modèles issus de la littérature sont d'abord présentés avec leurs différentes caractéristiques, puis sont comparés dans une étude sur simulations (Chapitre 5). Un modèle en particulier fait l'objet d'une analyse plus approfondie dans le chapitre 6, avec une description plus avancée de ses caractéristiques, de ses avantages et inconvénients, ainsi que de l'algorithme permettant son ajustement.

Enfin, la partie III présente des méthodes récentes permettant de modéliser les dépassements de seuil par un processus spatial. L'objectif de cette partie est notamment l'estimation d'une probabilité conditionnelle liée aux valeurs extrêmes. Les approches traitant des excès de seuil dans un cadre spatial sont décrites dans un premier temps dans le chapitre 7 puis, dans un second temps, plusieurs estimateurs de la probabilité conditionnelle d'intérêt sont définis dans le chapitre 8. Ils sont ensuite testés à travers plusieurs études sur simulations dans le chapitre 9. Une procédure permettant de prendre en compte la dépendance temporelle des observations journalières est également proposée et analysée dans ces deux chapitres.

Le programme utilisé pour la modélisation statistique tout au long de la thèse est le logiciel libre R (R Core Team, 2013). Une des contributions de cette thèse est l'implémentation d'un package (ensemble de fonctions pré-codées) dédié à un modèle spatial particulier de la littérature et aux améliorations qui lui ont été apportées. Ce package peut être trouvé sous le nom `hkevp` (Sebille, 2016).

Première partie

Outils méthodologiques et données de précipitations

Chapitre 1

Théorie des valeurs extrêmes

Soit X une variable aléatoire, un événement A est dit rare pour X si la probabilité $\Pr(X \in A)$ est petite. En statistique, la théorie des valeurs extrêmes (TVE) s'intéresse aux événements rares intervenant dans la queue de distribution de X. L'ensemble A est donc généralement noté : $A = (z, \infty)$ ou $A = (-\infty, -z)$, pour z une valeur extrême pour la variable aléatoire X. Afin d'étudier ces événements, la TVE se base sur deux théorèmes de convergence asymptotique : le théorème de Fisher-Tippett-Gnedenko et celui de Pickands-Balkema-de Haan.

La théorie est aujourd'hui bien développée pour des variables aléatoires univariées ou multivariées. Les références classiques sur la question sont Resnick (1987), Beirlant *et al.* (2004) et de Haan et Ferreira (2006), Embrechts *et al.* (1997) et Finkenstädt et Rootzén (2004) pour des applications en finance, ou encore Coles (2001) pour un aspect plus important accordé à la partie appliquée.

Plusieurs packages associés au logiciel R (R Core Team, 2013) sont également portés sur les méthodes d'estimations de la TVE. On peut citer entre autres `evd` (Stephenson et Ferro, 2015), `ismev` (Heffernan *et al.*, 2016), `texmex` (Southworth et Heffernan, 2013), `extRemes` (Gilleland et Katz, 2011) ou encore `SpatialExtremes` (Ribatet, 2015) et `RandomFields` (Schlather *et al.*, 2016) pour l'étude de valeurs extrêmes de processus spatiaux.

Dans ce chapitre, les principales notions liées à la TVE sont rappelées. Dans un premier temps, le cas univarié est traité avec trois approches possibles utilisées dans la littérature scientifique sur les valeurs extrêmes : l'étude des maxima par blocs, des excès de seuil et des processus ponctuels. Dans un second temps, on s'intéresse au cas de variables multivariées en dissociant les cas de dépendance asymptotique (AD) et d'indépendance asymptotique (AI).

L'extension de la TVE au cas fonctionnel (pour des processus stochastiques) n'est pas traitée dans cette partie, mais fera l'objet d'une présentation dans le chapitre 5 en définissant notamment plusieurs modèles spatiaux pour les valeurs extrêmes.

1.1 Cas univarié

1.1.1 Maxima par blocs

La théorie des valeurs extrêmes (TVE) trouve ses origines dans les résultats de Fisher et Tippett (1928) et Gnedenko (1943) sur la convergence en loi de la valeur maximale d'une série de répliques iid d'une variable aléatoire X. Le théorème suivant condense les résultats de ces deux articles.

Théorème 1 (Fisher-Tippett-Gnedenko)**.** *Soit X une variable aléatoire continue, soient $\{X_1, X_2, \ldots\}$ des répliques iid de X et $M_n = \max(X_1, \ldots, X_n)$, $\forall n \in \mathbb{N}$.*

S'il existe des suites $\{a_n\} > 0$ et $\{b_n\} \in \mathbb{R}$ et une fonction de répartition G non-dégénérée telles que :

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant z\right) \xrightarrow[n\to\infty]{} G(z) \ ,$$

alors G est nécessairement de loi $\mathrm{GEV}(\mu, \sigma, \xi)$*, définie sur* $\left\{z \in \mathbb{R} \ : \ 1 + \xi(z-\mu)/\sigma > 0\right\}$.

La fonction de répartition (f.d.r) de la loi GEV (*Generalized Extreme Value*) s'écrit :

$$G(z) = \exp\left(-\left[1 + \xi\frac{z-\mu}{\sigma}\right]_+^{-1/\xi}\right) \ ,$$

où $\mu \in \mathbb{R}$ est le paramètre de localisation, $\sigma > 0$ est le paramètre d'échelle, $\xi \in \mathbb{R}$ est le paramètre de forme et la notation a_+ correspond à $\max(a, 0)$ pour tout $a \in \mathbb{R}$.

Si le paramètre de forme ξ est nul, $G(z)$ s'écrit comme la limite continue en 0 :

$$G(z) = \exp\left(-\exp\left[-\frac{z-\mu}{\sigma}\right]\right) .$$

Si le théorème de Fisher-Tippett-Gnedenko est vérifié pour une variable aléatoire X de f.d.r F, on dit qu'elle appartient au domaine d'attraction de G et on note $X \sim F \in \mathrm{DA}(G)$.

La loi $\mathrm{GEV}(\mu, \sigma, \xi)$ généralise trois ensembles de lois déterminées par le paramètre de forme ξ. Selon le signe de ce dernier, X appartient au domaine d'attraction de Fréchet, de Gumbel ou de Weibull inverse :

- Si $\xi > 0$, X appartient au domaine d'attraction de la loi Fréchet. C'est le cas si F est une loi "à queue lourde", comme par exemple la loi de Pareto ou de Cauchy.
- Si $\xi = 0$, X appartient au domaine d'attraction de la loi Gumbel. C'est le cas si F est une loi "à queue légère", comme par exemple la loi normale ou exponentielle.
- Si $\xi < 0$, X appartient au domaine d'attraction de la loi Weibull-inverse. C'est le cas si F est une loi bornée à droite, comme par exemple la loi uniforme.

1.1.2 Excès de seuil

Une seconde approche à la théorie des valeurs extrêmes utilise le théorème de Pickands-Balkema-de Haan sur la convergence des excès de seuil. Soit $\{X_1, X_2, \ldots\}$ des répliques iid d'une variable aléatoire X et u un quantile élevé pour X. Les excès du seuil u par la variable aléatoire X sont définis par :

$$X - u \mid X > u .$$

Le théorème suivant, issu des résultats de Pickands (1975) et Balkema et De Haan (1974), décrit la convergence des excès du seuil u lorsque u tend vers le point limite $u^* \in \{\mathbb{R} \cup \infty\}$ défini par :

$$u^* = \sup\left\{u \in \mathbb{R} \ : \ F(u) < 1\right\} .$$

Théorème 2 (Pickands-Balkema-de Haan)**.** *Si $X \sim F \in \mathrm{DA}(G)$, alors :*

$$\Pr\big(X \leqslant u + z \mid X > u\big) \xrightarrow[u \to u^*]{} H(z) ,$$

et H est de loi $\mathrm{GP}(\tau, \xi)$ *:*

$$H(z) = 1 - \left(1 + \xi\frac{z}{\tau}\right)_+^{-1/\xi} \quad , \quad z \geqslant 0 .$$

La loi GP (*Generalized Pareto*) est définie par un paramètre d'échelle $\tau > 0$ et un paramètre de forme $\xi \in \mathbb{R}$.

Autrement dit, si X appartient au domaine d'attraction d'une loi $\mathrm{GEV}(\mu, \sigma, \xi)$, les excès de seuil

$$X - u \mid X > u$$

suivent approximativement une loi $\mathrm{GP}(\tau, \xi)$ et on a de plus la relation $\tau = \sigma + \xi(u - \mu)$.

Il est important de noter que le seuil u n'est pas un paramètre de la loi limite H, mais qu'en pratique, il est nécessaire de choisir ce seuil de telle façon que la valeur estimée de τ soit stable. De façon analogue, pour l'approche des maxima par blocs, la taille des blocs influe sur les valeurs de μ et σ (mais pas sur ξ) et doit donc être choisie "assez grande" en pratique lorsque la série des maxima annuels est approchée par une loi GEV.

1.1.3 Processus ponctuel

Convergence vers un processus de Poisson

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

$$N(A) = \sum_{i \geqslant 1} \mathbb{1}_{\{X_i \in A\}} ,$$

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

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

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

1. $N(A) \sim \text{Poi}\big(\Lambda(A)\big)$,
2. $N(A_1), \ldots, N(A_k)$ *sont mutuellement indépendants.*

Le théorème suivant (cf. Théorème 7.1.1 de Coles, 2001) permet de définir une troisième représentation des valeurs extrêmes d'une variable aléatoire X qui utilise le cadre d'un processus ponctuel de Poisson.

Théorème 3 (Convergence vers un processus de Poisson)**.** *Soit* X *une variable aléatoire et* $\{X_1, X_2, \ldots\}$ *des répliques iid de* X*. Si* X *appartient au domaine d'attraction d'une loi GEV*(μ, σ, ξ)*, alors la suite de processus ponctuels* $\{\mathcal{P}_n\}_n$ *définie sur* $\mathcal{A} = [0,1] \times \mathbb{R}$ *par :*

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

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

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

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

Lien avec les deux premières approches

En suivant le chapitre 7 de Coles (2001), il est possible de faire le lien entre la représentation des valeurs extrêmes par un processus de Poisson et les deux approches décrites dans la section précédente, à savoir les maxima par blocs et les excès de seuil.

Si on définit l'ensemble $A_z = (0,1) \times (z, \infty)$, le théorème 3 permet d'écrire :

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

On retrouve donc bien la forme de la fonction de répartition de la loi GEV de paramètres μ, σ et ξ.

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

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

Pour $z \geqslant u$, on obtient alors, grâce au théorème 3 :

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

On retrouve la fonction de répartition de la loi GP de paramètres τ et ξ.

Modélisation avec les excès de seuils

Les liens faits précédemment avec la méthode des excès de seuil permettent de reformuler le théorème 3 de la façon suivante (cf. Théorème 7.1.1 de Coles, 2001).

Théorème 4. *Soient* $\{X_1, X_2, \ldots\}$ *des répliques iid d'une variable aléatoire* X *et*

$$N_n = \left\{ \left(\frac{i}{n+1}, X_i \right) \ : \ i = 1, \ldots, n \right\} \ .$$

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

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi \frac{z - \mu}{\sigma} \right]^{-1/\xi} \tag{1.1}$$

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

En pratique, si on observe n_{ann} années d'observations, le théorème précédent fait correspondre les paramètres μ, σ et ξ de la formule (1.1) avec la loi GEV associée au maximum sur n_{ann} années d'observations. Il est donc préférable de travailler sur l'échelle annuelle en appliquant à (1.1) la modification :

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

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

1.1.4 Calcul d'un niveau de retour

On appelle *niveau de retour à* T *années* pour une variable aléatoire X le quantile qui est dépassé en moyenne une fois toutes les T années. Avec $M_n = \max(X_1, \ldots, X_n)$, le niveau de retour à T années noté z_T est donc défini par :

$$\Pr(M_n > z_T) = \frac{1}{T} \ . \tag{1.2}$$

Le niveau de retour à T années est une mesure de risque courante dans les applications de la théorie des valeurs extrêmes. Cette valeur sert en particulier à définir les dimensions d'une structure en fonction du risque encouru. Plus la gravité associée à une rupture de la construction est importante, plus la période de retour fixée T associée au phénomène naturel est grande. Par exemple, $T = 1000$ pour une digue et $T = 10000$ pour une centrale nucléaire.

Les trois approches de la théorie des valeurs extrêmes décrites dans la section précédente permettent d'estimer le niveau de retour à T années associé à une variable aléatoire X dans le cas où l'on dispose d'observations iid de cette variable aléatoire. Si les observations sont corrélées, des conditions supplémentaires sont nécessaires (voir Section 1.1.5).

Niveau de retour calculé à partir des maxima par blocs

En utilisant la définition (1.2) du niveau de retour et la forme explicite de la loi GEV associée aux maxima annuels, un estimateur du niveau de retour peut être directement construit par cette méthode :

$$\widehat{z_T}^{(1)} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left[-\log \left\{ 1 - \frac{1}{T} \right\} \right]^{-\hat{\xi}} - 1 \right) \ , \quad T > 1 \ , \tag{1.3}$$

où $\hat{\mu}$, $\hat{\sigma}$ et $\hat{\xi}$ sont les estimations respectives des paramètres μ, σ et ξ de la loi GEV, obtenus par maximum de vraisemblance ou par la méthode des moments (cf. de Haan and Ferreira, 2006, page 139).

Niveau de retour calculé à partir des excès de seuil

Si on définit z_m tel que $\Pr(X > z_m) = 1/m$, on obtient le niveau de retour à m-observations d'après la notation (1.2). En notant n_y le nombre d'observations par années (en pratique on a $n_y = 365$ jours en supprimant les 29 février), le niveau de retour à m-observations devient équivalent au niveau de retour à m/n_y-années.

En choisissant un seuil $u > z_m$ assez élevé, on a :

$$\Pr(X > z_m) = \zeta_u \Pr(X > z_m \mid X > u) = \frac{1}{m} \ ,$$

où $\zeta_u = \Pr(X > u)$ peut être estimé empiriquement par n_u/n, où n_u est le nombre d'observations parmi les réalisations $(X_1, \ldots, X_n)$ qui dépassent le seuil u.

L'approximation de $\Pr(X > z_m \mid X > u)$ par une loi GP de paramètres τ et ξ permet donc d'obtenir un second estimateur du niveau de retour à T-années :

$$\widehat{z_T}^{(2)} = u + \frac{\hat{\tau}}{\hat{\xi}} \left(\left(\frac{n_u T n_y}{n} \right)^{\hat{\xi}} - 1 \right) , \tag{1.4}$$

où $\hat{\tau}$ et $\hat{\xi}$ sont les estimations des paramètres τ et ξ de la loi GP.

Niveau de retour calculé à partir des processus ponctuels

Pour tout $z > u$, l'espérance du nombre d'excès de z par an est :

$$\mathbb{E}\big[N\big((0,1) \times (z, \infty)\big)\big] = \Lambda\big((0,1) \times (z, \infty)\big) = \left[1 + \xi \frac{z - \mu}{\sigma}\right]^{-1/\xi} ,$$

et le niveau de retour à T-années z_T satisfait donc l'équation

$$\left[1 + \xi \frac{z_T - \mu}{\sigma}\right]^{-1/\xi} = \frac{1}{T} .$$

On en déduit l'estimateur du niveau de retour par la méthode des processus ponctuels :

$$\widehat{z_T}^{(3)} = \hat{\mu} + \hat{\sigma} \frac{T^{\hat{\xi}} - 1}{\hat{\xi}} ,$$

où $\hat{\mu}, \hat{\sigma}$ et $\hat{\xi}$ sont les estimations des paramètres μ, σ et ξ associés à l'intensité du processus de Poisson limite.

1.1.5 Observations stationnaires

Théorèmes de Leadbetter

En pratique, supposer que les observations $\{X_i\}_{i \geqslant 1}$ sont iid n'est pas toujours justifié, mais sous certaines conditions, il est possible d'appliquer les résultats de la théorie des valeurs extrêmes vus précédemment pour estimer un niveau de retour. Une condition requise, connue sous le nom de condition $D(u_n)$ (Leadbetter, 1974), permet d'éviter le cas de dépendance à long terme.

Cette condition est aussi donnée dans Beirlant *et al.* (2004) (cf. Condition 10.1, page 373) :

Condition 1 ($D(u_n)$)**.** *Soit $\mathcal{I}_{j,k}(u_n)$ l'événement défini pour tous entiers $j < k$ par :*

$$\mathcal{I}_{j,k}(u_n) := \big\{ \max(X_j, \ldots, X_k) \leqslant u_n \big\} ,$$

alors une suite stationnaire $\{X_i\}_{i \geqslant 1}$ vérifie la condition $D(u_n)$ si pour tout $A_1 \in \mathcal{I}_{1,\ell}(u_n)$ et $A_2 \in \mathcal{I}_{\ell+s,n}(u_n)$ et $1 \leqslant \ell \leqslant n - s$, on a :

$$\big| \Pr(A_1 \cap A_2) - \Pr(A_1)\Pr(A_2) \big| \leqslant \alpha(n, s) ,$$

où $\alpha(n, s_n) \to 0$ pour une suite $s_n = o(n)$, $n \to \infty$.

La condition $D(u_n)$ impose donc à une série stationnaire que les observations X_i et X_j deviennent quasi-indépendantes lorsqu'elles sont séparées par une distance suffisante. Le théorème suivant (Leadbetter, 1974) explicite la convergence des maxima par blocs d'une série stationnaire vérifiant cette condition.
Ce théorème est donné dans Beirlant *et al.* (2004) (cf. Théorème 10.2, page 373) :

Théorème 5 (Leadbetter, 1974)**.** *Si $\{X_i\}_{i \geqslant 1}$ est une suite stationnaire, si $M_n = \max(X_1, \ldots, X_n)$, s'il existe deux suites $a_n > 0$ et $b_n \in \mathbb{R}$ telles que*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant z \right) \xrightarrow[n \to \infty]{} G(z) ,$$

où G est une loi non-dégénérée et si la condition $D(u_n)$ est vérifiée avec $u_n = a_n z + b_n$ pour tout $z \in \mathbb{R}$ tel que $G(z) > 0$, alors G est de forme GEV.

Les paramètres de la loi limite G ne sont pas identiques à ceux correspondant au cas iid. Pour mettre en relation les paramètres, il faut se référer au théorème de Leadbetter (1983), que l'on peut retrouver dans Beirlant *et al.* (2004) (cf. Théorème 10.4, page 377) :

Théorème 6 (Leadbetter, 1983). *Soient $\{X_i\}_{i\geqslant 1}$ une suite stationnaire et $\{X_i^*\}_{i\geqslant 1}$ une suite iid de même loi que les X_i. Soient $M_n = \max(X_1, \ldots, X_n)$ et $M_n^* = \max(X_1^*, \ldots, X_n^*)$. Sous les hypothèses du théorème de Leadbetter (1974), on a :*

$$\Pr\left(\frac{M_n^* - b_n}{a_n} \leqslant z\right) \xrightarrow[n\to\infty]{} G_*(z) \ ,$$

si et seulement si :

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant z\right) \xrightarrow[n\to\infty]{} G(z) \ ,$$

où $G(z) = G_*^\theta(z)$ et où $\theta \in (0, 1]$ est appelé indice extrémal de la série $\{X_i\}_{i\geqslant 1}$.

Les paramètres (μ, σ, ξ) de G et les paramètres (μ^*, σ^*, ξ^*) de G_* sont reliés par l'ensemble d'équations :

$$\left\{\begin{array}{rcl} \mu^* &=& \mu - \frac{\sigma}{\xi}\left(1 - \theta^\xi\right) \ , \\ \sigma^* &=& \sigma\theta^\xi \ , \\ \xi^* &=& \xi \ . \end{array}\right.$$

Indice extrémal

Leadbetter (1983) montre que $\{X_i\}_{i\geqslant 1}$ a pour indice extrémal $\theta \in (0, 1]$ si pour un $\tau \in (0, \infty)$, il existe une suite croissante $\{u_n\}_{n\geqslant 1}$ telle que :

1. $n\bar{F}(u_n) \xrightarrow[n\to\infty]{} \tau$,
2. $\Pr(M_n \leqslant u_n) \xrightarrow[n\to\infty]{} \exp(-\theta\tau)$,

où $M_n = \max(X_1, \ldots, X_n)$ et $\bar{F} = 1 - F$ est la fonction de survie de la série $\{X_i\}_{i\geqslant 1}$.

Plus l'indice extrémal est faible et plus les excès de la série $\{X_i\}_{i\geqslant 1}$ auront tendance à se regrouper en paquets ou *clusters* de dépassement. En fait, l'indice extrémal θ peut être interprété de deux façons :

– comme l'inverse de la taille moyenne des clusters de dépassement :

$$\theta^{-1} := \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{p_n} 1_{\{X_i > u_n\}} \ \middle| \ M_{p_n} > u_n\right] \ , \quad p_n = o(n) \ ,$$

– comme la probabilité qu'un excès de u_n par la série X soit le dernier :

$$\theta := \lim_{n\to\infty} \Pr\big(\max(X_1, \ldots, X_{p_n}) \leqslant u_n \mid X_1 \geqslant u_n\big) \ , \quad p_n = o(n) \ .$$

Ferro et Segers (2003) construisent un estimateur non-paramétrique de l'indice extrémal θ basé sur les temps d'inter-excès. On note $\{S_k(u)\}_{k=1,\ldots,N_u}$ les temps d'excès de u de la série $\{X_i\}_{i\geqslant 1}$ et $\{T_k(u)\}_{k=1,\ldots,N_u-1}$ les temps d'inter-excès définis par

$$T_i(u) = S_{i+1}(u) - S_i(u) \ ,$$

pour tout $i \in \{1, \ldots, N_u - 1\}$.

Deux estimateurs de θ sont alors proposés :

$$\tilde{\theta}_1(u) = \frac{2\left(\sum_{i=1}^{N_u} T_i(u)\right)^2}{(N_u - 1)\sum_{i=1}^{N_u} T_i^2(u)} \ ,$$

et

$$\tilde{\theta}_2(u) = \frac{2\left(\sum_{i=1}^{N_u} T_i(u) - 1\right)^2}{(N_u - 1)\sum_{i=1}^{N_u}(T_i(u) - 1)(T_i(u) - 2)} \ .$$

Le second estimateur n'est pas défini si les clusters sont tous de taille 1 ou 2 mais permet d'affecter une contribution de 0 aux petits inter-excès (c'est-à-dire, entre les excès apparaissant au sein d'un même cluster). Ferro et Segers (2003) proposent ainsi d'utiliser l'estimateur :

$$\hat{\theta}(h) = \left\{\begin{array}{lll} \max\left\{1, \tilde{\theta}_1(u)\right\} & \text{si} & \max\{T_1, \ldots, T_{N_u-1}\} \leqslant 2 \ , \\ \max\left\{1, \tilde{\theta}_2(u)\right\} & \text{si} & \max\{T_1, \ldots, T_{N_u-1}\} > 2 \ . \end{array}\right.$$

Calcul d'un niveau de retour pour des données stationnaires

Cai *et al.* (2013) proposent deux méthodes pour calculer le niveau de retour à T-années sur des données corrélées.

La première approche, appelée *méthode de l'indice extrémal*, utilise l'approche des dépassements de seuil avec l'estimation de θ. Cette méthode est constituée de trois étapes :

1. Estimer l'indice extrémal θ sur la série observée $\{X_i\}_{i=1,\dots,n}$ en utilisant par exemple la méthode de Ferro et Segers (2003).
2. Ajuster la queue de distribution de la série observée par une loi GP comme si la série observée était indépendante et estimer par maximum de vraisemblance les paramètres τ et ξ.
3. En choisissant un ordre k assez grand par rapport au nombre d'observations n, le niveau de retour à T années peut être estimé par :
$$\hat{x}_T := x_{n-k+1,n} + \hat{\tau}\frac{\left(\frac{k}{\hat{\alpha}n}\right)^{\hat{\xi}} - 1}{\hat{\xi}} \ ,$$
où $\hat{\alpha} = 1 - (1 - 1/T)^{1/(n_y\hat{\theta})}$, n_y est le nombre d'observations de $\{X_i\}_{i\geqslant 1}$ par an ($n_y = 365$) et $x_{n-k+1,n}$ est la k-ième plus grande valeur observée de la série.

La deuxième approche, notée *méthode de declustering*, se sert du Théorème 4.5 de Hsing (1987), qui montre que les clusters de dépassements peuvent être asymptotiquement considérés comme indépendants. Les valeurs maximales de chaque cluster de dépassement sont donc considérées comme indépendantes et il est alors possible d'ajuster une loi GP sur cette série de dépassements (plus courte que celles des dépassements bruts) en se ramenant au cas iid et en utilisant l'estimateur (1.4).

1.2 Cas multivarié

Dans le cas d'une variable aléatoire univariée, les valeurs extrêmes correspondent aux valeurs élevées définies grâce à la relation d'ordre sur $\mathbb{R}$. Pour un vecteur $X = (X_1, \dots, X_d)$ de $\mathbb{R}^d$, la relation d'ordre n'existe plus et l'étude des valeurs extrêmes peut se faire en utilisant plusieurs approches. Plusieurs méthodes possibles pour définir les valeurs extrêmes multivariées sont proposées par Barnett (1976), parmi lesquelles

1. prendre le vecteur de maxima "composantes par composantes" $M_n = (\vee_{i=1}^n X_{i,1}, \dots, \vee_{i=1}^n X_{i,d})$,
2. définir les excès selon une fonction $f : \mathbb{R}^d \longrightarrow \mathbb{R}$,
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.

1.2.1 Séparer les marges de la structure de dépendance

Soit $X = (X_1, \dots, X_d) \sim F$ un vecteur multivarié à marges continues de $\mathbb{R}^d$. La loi de X est séparée en deux aspects :

- les lois marginales $F_1, \dots, F_d$ qui décrivent le comportement de chaque composante,
- la structure de dépendance entre les composantes expliquée par la *copule* C_F :
$$F(x) = C_F\big(F_1(x_1), \dots, F_d(x_d)\big) \ .$$

Les valeurs extrêmes du vecteur multivarié X sont donc regardées en deux temps. D'abord, les lois marginales sont analysées en utilisant la théorie des valeurs extrêmes univariées (voir section 1.1). Ensuite, le comportement joint de la copule C_F est regardé près du point limite $(1, \dots, 1)$.

Dans la section suivante, deux mesures complémentaires de dépendance extrémale sont présentées.

1.2.2 Mesures de dépendance asymptotique

Afin d'évaluer la dépendance extrémale du vecteur $X = (X_1, \dots, X_d)$, deux mesures de dépendance extrémale (ou dépendance asymptotique) sont définies dans cette section. Ces mesures ne dépendent que de la copule C_F associée à la fonction de répartition F et se concentrent sur la dépendance dans les extrêmes qui, en général, est différente de celle observée sur les valeurs moyennes de X.

Ces deux mesures sont définies en dimension $d = 2$ par simplicité mais elles peuvent être généralisées au cas $d \geqslant 3$. Coles *et al.* (1999) recensent ces mesures avec des illustrations sur des données réelles maritimes et de précipitations. Bacro et Toulemonde (2013) les utilisent pour des processus spatiaux en se concentrant sur les phénomènes asymptotiquement indépendants.

Coefficient de dépendance caudale

Le coefficient de dépendance caudale, plus connu sous le terme anglais *upper-tail dependence coefficient*, est le nombre $\chi \in [0,1]$ défini par :

$$\begin{aligned}\chi &= \lim_{u\to 1} \Pr\big(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)\big) \\ &= \lim_{u\to 1} \frac{1-2u+C_F(u,u)}{1-u} \ ,\end{aligned}$$

où C_F est la copule associée au vecteur $X = (X_1, X_2)$.

Le coefficient de dépendance caudale permet de dissocier deux cas de dépendance asymptotique :
– lorsque $\chi = 0$, les variables X_1 et X_2 sont dites *asymptotiquement indépendantes*,
– lorsque $\chi > 0$, les variables X_1 et X_2 sont dites *asymptotiquement dépendantes*.

Si $\chi = 0$, plusieurs types de dépendance sont encore possibles. Par exemple, si (X_1, X_2) suit une loi gaussienne bivariée de corrélation $\rho < 1$, alors les composantes X_1 et X_2 sont asymptotiquement indépendantes (Sibuya, 1960). Ce résultat est vrai pour toute corrélation ρ même très proche de 1 : un vecteur gaussien même très corrélé positivement a des valeurs extrêmes indépendantes.
Si $\chi > 0$, la fonction de survie jointe $\bar{F}$ décroît vers 0 à la même vitesse que les fonctions de survies marginales. Autrement dit,

$$\Pr\big(X_{P,1} > x, X_{P,2} > x\big) \approx \chi x^{-1} \ ,$$

où le vecteur $X_P = (X_{P,1}, X_{P,2})$ correspond à la transformation du vecteur X en marges Pareto standard grâce à la formule :

$$X_{P,j} = \frac{1}{1-F_j(X_j)} \ , \quad j = 1,2 \ . \tag{1.5}$$

Estimer le coefficient χ est possible en utilisant les probabilités empiriques :

$$\widehat{\chi}(u) = \frac{\sum_{i=1}^n \mathbb{1}_{\left\{\min\left(\hat{F}_1(X_{i,1}), \hat{F}_2(X_{i,2})\right) > u\right\}}}{\sum_{i=1}^n \mathbb{1}_{\left\{\hat{F}_2(X_{i,2}) > u\right\}}} \ ,$$

pour u proche de 1, où $\hat{F}_j$ est la fonction de répartition marginale empirique de X_j pour $j \in \{1,2\}$.

Indice de dépendance résiduelle

Une mesure complémentaire au coefficient χ est l'indice de dépendance résiduelle $\eta \in (0,1]$ de Ledford et Tawn (1996), nommé en anglais *residual dependence index* (de Haan et Ferreira, 2006, p. 236). Ce coefficient est défini en supposant que :

$$\Pr\big(X_{P,1} > x, X_{P,2} > x\big) = \mathcal{L}(x) x^{-1/\eta} \ ,$$

où $\mathcal{L}$ est une fonction à variation lente, c'est-à-dire que pour tout $\lambda > 0$, $\mathcal{L}(\lambda x)/\mathcal{L}(x) \xrightarrow[x\to\infty]{} 1$.

Les coefficients χ et η sont complémentaires dans le sens où :
– si X est un vecteur asymptotiquement dépendant, $\chi > 0$, $\eta = 1$ et $\mathcal{L}(x) \xrightarrow[x\to\infty]{} \chi$,
– si X est asymptotiquement indépendant, $\chi = 0$ et $\eta < 1$.

L'indice η permet de décrire la dépendance associée à des variables asymptotiquement indépendantes. Cai *et al.* (2013) utilisent l'indice η et la variation régulière de X pour estimer la probabilité que le couple (X_1, X_2) apparaisse dans une *région d'échec* élevée.

Le degré d'association entre les composantes du vecteur X est indiqué par η :
– si $\eta \in (0, 1/2)$, les composantes sont associées négativement,
– si $\eta = 1/2$, les composantes sont indépendantes,
– si $\eta \in (1/2, 1]$, les composants sont associées positivement.

Par exemple, si (X_1, X_2) suivent une loi normale bivariée de coefficient de corrélation $\rho < 1$, le coefficient de dépendance caudale χ est nul mais l'indice de dépendance résiduelle η est donné par :

$$\eta = (1+\rho)/2 \ .$$

Heffernan (2000) fournit, pour plusieurs lois usuelles asymptotiquement indépendantes, la formule permettant de calculer l'indice η en fonction des paramètres de la loi.

L'estimation de l'indice η est proposée par Draisma *et al.* (2004), qui généralise l'approche de Peng (1999). Sous certaines conditions, les auteurs montrent que la fonction de répartition F_T de la variable aléatoire

$$T := \min\left(\frac{1}{1-F_1(X_1)}, \frac{1}{1-F_2(X_2)}\right)$$

est à variation régulière d'indice $1/\eta$.

L'indice η est donc estimé sur l'échantillon $\{T_i\}_{i=1,\dots,n}$, défini pour tout $i \in \{1,\dots,n\}$ par :

$$T_i := \min\left(\frac{n+1}{n+1-R(X_{i,1})}, \frac{n+1}{n+1-R(X_{i,2})}\right) ,$$

où $R(X_{i,j})$ correspond au rang de la variable X_j pour $j \in \{1,2\}$. Un estimateur est construit à partir de l'estimateur non-paramétrique de Hill (1975) et des m plus grandes statistiques d'ordre :

$$\widehat{\eta} := \frac{1}{m}\sum_{i=1}^{m} \log\left(\frac{T_{n,n-i+1}}{T_{n,n-m}}\right) , \tag{1.6}$$

où $T_{n,r-1}$ est la r-ième statistique d'ordre de T.

1.2.3 Convergence des valeurs extrêmes multivariées

Les théorèmes de convergence de Fisher-Tippett-Gnedenko et Pickands-Balkema-de Haan peuvent être généralisés au cas d'un vecteur multivarié $X = (X_1,\dots,X_d)$, mais les propriétés des lois limites sont plus complexes que dans le cas unidimensionnel décrit dans la section 1.1.

Convergence des maxima par blocs

Le théorème 6.1.1 de (de Haan et Ferreira, 2006, p210) montre la convergence des maxima par blocs, en les considérant composante par composante :

Théorème 7. *Soient $(X_1,\dots,X_d)$ un vecteur aléatoire de fonction de répartition F et le vecteur de maximum composante par composante :*

$$M_n = \big(M_{n,1},\dots,M_{n,d}\big) = \left(\bigvee_{i=1}^{n} X_{i,1},\dots,\bigvee_{i=1}^{n} X_{i,d}\right) .$$

S'il existe des suites $a_n \in \mathbb{R}_+^d$ et $b_n \in \mathbb{R}^d$, et une fonction de répartition G non dégénérée telles que :

$$\Pr\left(\frac{M_{n,1}-b_{n,1}}{a_{n,1}} \leqslant z_1,\dots,\frac{M_{n,d}-b_{n,1}}{a_{n,d}} \leqslant z_d\right) \xrightarrow[n\to\infty]{} G(z_1,\dots,z_d) \ , \ \big(z_1,\dots,z_d\big) \in \mathbb{R}^d$$

alors G est nécessairement une loi de valeurs extrêmes multivariée (MEV) et on dit que $F \in \mathrm{DA}(G)$.

Grâce à la transformation (1.5), on peut supposer que les marges de X sont de loi Pareto standard. On obtient ainsi une loi limite de marges Fréchet unitaires GEV(1,1,1). La fonction de répartition G s'écrit alors :

$$G(z) = \exp\big(-V(z_1,\dots,z_d)\big) ,$$

où V est appelée *fonction exposante* de la loi G. Cette mesure contient l'information sur la structure de dépendance entre les maxima des composantes du vecteur X et vérifie les propriétés suivantes :

1. V est homogène d'ordre -1, c'est-à-dire que pour tout réel $\lambda > 0$ on a $V(\lambda\mathbf{z}) = \lambda^{-1}V(\mathbf{z})$,
2. $V(\infty,\dots,\infty,z_j,\infty,\dots,\infty) = 1/z_j$.

La première condition assure que la loi G est max-stable. Précisément, pour tout $\lambda > 0$:

$$G^{\lambda}(\lambda z_1,\dots,\lambda z_d) = G(z_1,\dots,z_d) .$$

La seconde condition assure que les lois marginales G_j de G sont Fréchet unitaires : $G_j(z_j) = \exp(-1/z_j)$.

Les cas limites de l'indépendance asymptotique et de la dépendance complète correspondent respectivement à :

$$V(z_1,\dots,z_d) = \sum_{j=1}^{d}\frac{1}{z_j} \quad \text{et} \quad V(z_1,\dots,z_d) = \bigvee_{j=1}^{d}\frac{1}{z_j} .$$

Dans le cas bivarié $d = 2$, la dépendance peut s'exprimer par la fonction A de Pickands (1981) :

$$G(z_1, z_2) = \exp\left[-\left(\frac{1}{x_1} + \frac{1}{x_2}\right) A\left(\frac{x_1}{x_1 + x_2}\right)\right] ,$$

qui vérifie les propriétés suivantes (Beirlant *et al.*, 2004, Section 8.2.5) :

1. $A : [0, 1] \longrightarrow [1/2, 1]$,
2. $\max(t, 1 - t) \leqslant A(t) \leqslant 1$, pour tout $t \in [0, 1]$,
3. $A(0) = A(1) = 1$.

Vraisemblance d'une loi MEV

Un problème rencontré avec les lois MEV est que le nombre de termes de la fonction de densité g devient rapidement trop grand à mesure que d augmente pour que la vraisemblance liée aux observations de maxima par blocs puisse être utilisée.

En dimension $d = 2$, la densité de G s'écrit :

$$g(z_1, z_2) = \big(\partial_{\{1\}} V(z_1, z_2)\partial_{\{2\}} V(z_1, z_2) - \partial_{\{1,2\}} V(z_1, z_2)\big) \exp\big(-V\big(z_1, z_2\big)\big) ,$$

où $\partial_I V$ représente la dérivée partielle de V selon les indices $I \subset \{1, 2\}$.

La vraisemblance est composée d'autant de termes que le nombre de partitions distinctes de l'ensemble $\{1, \ldots, d\}$. Ce nombre est égal au nombre de Bell. S'il reste raisonnable pour $d = 2$ ou 3, il devient rapidement très élevé : pour $d = 10$, la fonction de densité $g(z_1, \ldots, z_{10})$ s'écrit comme la somme de 115 975 termes, ce qui rend impossible l'inférence classique comme l'estimation par maximum de vraisemblance.

Pour pallier ce problème, il est préférable d'utiliser une vraisemblance composite (Lindsay, 1988; Padoan *et al.*, 2010) qui ne prend en compte par exemple que la vraisemblance entre les paires $g(z_i, z_j)_{1 \leqslant i < j \leqslant d}$. Cette méthode est présentée plus en détails dans la partie II de la thèse sur les processus max-stables.

Queues de distributions multivariées

La modélisation des valeurs extrêmes par le vecteur de maxima composante par composante est un outil relativement simple mais qui présente deux inconvénients :

1. Beaucoup de données sont jetées en ne choisissant qu'une valeur parmi un bloc dont la taille doit être suffisamment grande pour que l'approximation par une loi MEV soit raisonnable.
2. Si le but d'une analyse est d'estimer une probabilité du type $\Pr(X_1 > x_1, X_2 > x_2)$ avec des variables X_1 et X_2 asymptotiquement indépendantes (de coefficient de dépendance caudal $\chi = 0$), l'approche des maxima par blocs ne convient plus.

Pour cela, il est nécessaire de pouvoir étudier les valeurs extrêmes d'un vecteur multivarié X selon une approche alternative. Cette partie concerne l'étude des dépassements de seuil dans le cas multivarié et spatial et est présenté plus en détails dans le chapitre 7.

Dans cette section, la décomposition en coordonnées *pseudo-polaires* est explicitée. On considère le vecteur X_P de marges Pareto standard et on définit les dépassements du seuil u par :

$$X_P \mid f(X_P) > u \ , \quad f(X_P) = \sum_{j=1}^{d} X_{P,j} \ .$$

La décomposition pseudo-polaire de X_P se fait en définissant :
– un pseudo-rayon $R = f(X_P) = \sum_{j=1}^{d} X_{P,j}$,
– un pseudo-angle $W = X_P/R$.
Avec cette écriture, les excès de seuil u s'écrivent :

$$\big\{X_P \mid f(X_P) > u\big\} \Longleftrightarrow \big\{RW \mid R > u\big\} \ .$$

Si $F \in \mathrm{DA}(G)$, alors on a pour $t \to \infty$:

$$\Pr\big(R > tr, W \in A_w \mid R > t\big) \longrightarrow r^{-1} H(A_w) \ , \quad r > 0 \ , \ w \in \mathcal{S}_{d-1} \ ,$$

où $\mathcal{S}_{d-1} = \Big\{w \in \mathbb{R}_+^d \ : \ \sum_{j=1}^{d} w_j = 1\Big\}$ est le simplexe unitaire de $\mathbb{R}^d$ pour la norme L_1.

H est une mesure de probabilité sur $\mathcal{S}_{d-1}$ appelée *mesure spectrale* ou *angulaire*. Elle est liée à la fonction exposante V par la relation :

$$V(z) = d \int_{\mathcal{S}_{d-1}} \max\left\{\frac{w_1}{z_1}, \ldots, \frac{w_d}{z_d}\right\} dH(w) \ . \tag{1.7}$$

En dimension $d = 2$, le simplexe unitaire $\mathcal{S}_{d-1}$ correspond à l'intervalle $[0, 1]$. Si les composantes de X_P sont asymptotiquement indépendantes ($\chi = 0$), alors la mesure spectrale H a toute sa masse sur $\{0, 1\}$:

$$H(w) = \frac{1}{2}\mathbb{1}_{\{w=0\}} + \frac{1}{2}\mathbb{1}_{\{w=1\}} \ .$$

Si en revanche X_P est asymptotiquement dépendant, H aura une masse strictement positive sur $(0, 1)$.

1.2.4 Modèles d'indépendance asymptotique

Lors d'une étude sur les valeurs extrêmes d'un vecteur multivarié, le cas d'indépendance asymptotique peut poser problème, par exemple si le but est l'estimation d'une probabilité :

$$\Pr\big(X \in A\big) \ , \tag{1.8}$$

pour un ensemble "extrême" A (Cai *et al.*, 2013).

En effet, si X_P est le vecteur observé en marges Pareto standard, alors X_P est à variations régulières et on a pour un ensemble $B \subset \mathbb{R}^d$:

$$t\Pr\big(X_P/t \in rB\big) \xrightarrow[t\to\infty]{} \mu(rB) \ , \quad t, r > 0 \ , \tag{1.9}$$

où μ est une mesure homogène d'ordre -1. Autrement dit : $\mu(rB) = r^{-1}\mu(B)$.

Comme la convergence (1.9) est vérifiée pour tout ensemble $B \subset [0, \infty)^d \backslash \{0\}$, on obtient l'approximation :

$$\Pr\big(X_P \in trB\big) \approx r^{-1}\Pr\big(X_P \in tB\big) \ .$$

En calculant de façon empirique la probabilité $\Pr\big(X \in tB\big)$ sur l'ensemble tB, cette relation permet d'estimer la probabilité (1.8) pour A un ensemble plus extrême en posant $A = trB$.

Cette méthode ne marche plus si les composantes du vecteur X_P sont asymptotiquement indépendantes car la mesure μ concentre sa masse sur les axes $\{0\} \cup (0, \infty)$ et $(0, \infty) \cup \{0\}$.

Modèle de Ledford-Tawn

Le modèle de Ledford et Tawn (1996) suppose que le couple asymptotiquement indépendant $(X_{P,1}, X_{P,2})$ vérifie :

$$\Pr\big(X_{P,1} > x, X_{P,2} > x\big) = \mathcal{L}(x)x^{-1/\eta} \ ,$$

où $\mathcal{L}$ est une fonction à variation lente et η est l'indice de dépendance résiduelle. Grâce à cette formulation, on obtient une convergence similaire à (1.9) :

$$\mathcal{L}(t)t^{1/\eta}\Pr\big(X_P/t \in B\big) \longrightarrow \nu(rB) \ ,$$

où cette fois ν est une mesure homogène d'ordre $-1/\eta$. Cette convergence permet d'obtenir l'approximation :

$$\Pr\big(X_P \in trB\big) \approx r^{-1/\eta}\Pr\big(X_P \in tB\big) \ . \tag{1.10}$$

Pour estimer la probabilité d'échec (1.8), la procédure à suivre est donc la suivante :

1. Estimer η en utilisant l'estimateur (1.6) construit par Draisma *et al.* (2004).
2. Calculer la probabilité $\Pr\big(X_P \in tB\big)$, $t \in (0, 1)$ de façon empirique.
3. Utiliser la relation (1.10) pour estimer (1.8).

Cette méthode est efficace mais peut toujours poser problème. En effet, il n'est pas toujours possible de se ramener à un ensemble tA suffisamment proche des données pour pouvoir estimer la probabilité $\Pr\big(X_P \in tB\big)$ et si une composante de X_P est plus élevée que les autres, l'approximation ne sera plus valable. En effet, le modèle de Ledford et Tawn (1996) est construit sur la supposition que toutes les composantes sont extrêmes en même temps.

Modèle conditionnel de Heffernan-Tawn

Une méthode alternative pour étudier le comportement joint des extrêmes et pour estimer la probabilité d'échec (1.8) est le modèle conditionnel de Heffernan et Tawn (2004) et Heffernan et Resnick (2007). Soit $X_E = (X_{E,1}, X_{E,2})$ le vecteur aléatoire X transformé en marges exponentielles standard par la transformation :

$$X_{E,j} = -\log F_j(X_j) \ .$$

Le modèle conditionnel de Heffernan et Tawn (2004) étudie le comportement de la composante $X_{E,2}$ lorsque $X_{E,1} > u_E$, où u_E est une valeur élevée pour la variable exponentielle $X_{E,1}$.

Les auteurs montrent que pour une grande variété de structures de dépendances entre les composantes de X_E, on a la convergence :

$$\Pr\left(\frac{X_{E,2}-\alpha X_{E,1}}{X_{E,1}^{\beta}} \leqslant x, X_{E,1}-u_E > y \;\middle|\; X_{E,1} > u_E\right) \xrightarrow[u_E\to\infty]{} K(x)\exp(-y)\ , \tag{1.11}$$

pour $\alpha \in [0,1]$ et $\beta < 1$ et une fonction de répartition K non dégénérée. En d'autres termes, les variables

$$Z = \frac{X_{E,2}-\alpha X_{E,1}}{X_{E,1}^{\beta}} \quad \text{et} \quad X_{E,1}-u_E$$

deviennent indépendantes quand $u_E \to \infty$ sachant $X_{E,1} > u_E$.
Si le vecteur X_E est asymptotiquement dépendant ($\chi > 0$), alors $\alpha = 1$ et $\beta = 0$, mais s'il est asymptotiquement indépendant ($\chi = 0$), le vecteur de paramètres (α, β) a ses valeurs dans $[0,1] \times (-\infty, 1)\backslash\{0,1\}$.

L'ajustement du modèle de Heffernan et Tawn (2004) permet d'estimer les valeurs de α et β et les paramètres de la fonction de répartition K associée à Z. Ensuite, le modèle permet de produire des simulations de $X_E \mid X_{E,1} > u_E^*$ où $u_E \geqslant u_E$ est un seuil arbitraire. La procédure de simulation est décomposée de la façon suivante :

1. Générer une réalisation de $X_{E,1}^* \sim \text{Exp}(1) + u_E^*$.
2. Générer Z de fonction de répartition K.
3. Définir $X_{E,2}^* = \hat{\alpha} X_{E,1}^* + \left(X_{E,1}^*\right)^{\hat{\beta}} Z$.

En utilisant les simulations X_E^* de $X_E \mid X_{E,1} > u_E^*$, il est possible d'estimer empiriquement une probabilité d'échec du type (1.8).

Cette méthode a pour avantage d'être très flexible car la convergence (1.11) est souvent vérifiée, mais l'estimation empirique de la probabilité d'échec sur les simulations peut poser des problèmes, comme indiqué dans les chapitres 8 et 9, où ce modèle est utilisé dans un cas particulier pour estimer une probabilité d'échec conditionnelle.

Chapitre 2

Élements de géostatistique

Ce chapitre introduit les notions de base de la géostatistique (ou statistique spatiale) utilisées dans la thèse. Parmi les ouvrages de référence traitant de ce domaine de la statistique, on peut citer Wackernagel (2013), Chilès et Delfiner (2009), Cressie (1992) ou encore le cours en ligne de Guyon (2007).

Le type de données considéré est formé par un processus stochastique $\{X(s)\}_{s\in\mathcal{S}}$, où $\mathcal{S}$ est la région spatiale définie comme un sous-ensemble de $\mathbb{R}^p$. On s'intéresse en particulier au cas $p = 2$. Généralement, le processus $\{X(s)\}_{s\in\mathcal{S}}$ est observé sur un ensemble de positions $\{s_1, \dots, s_d\} \subset \mathcal{S}$.

Typiquement, le but d'une approche utilisant la géostatistique est de prédire le phénomène observé en un ensemble de positions *cibles* $\{s_1^*, \dots, s_{d^*}^*\} \subset \mathcal{S}$, avec une erreur de prédiction associée. Un exemple précurseur d'application pour $p = 3$ est l'étude de la porosité de la roche dans un sous-sol, en vue d'une exploitation pétrolière : cette application a notamment donné naissance au terme *géostatistique*.

Les notions théoriques associées à un processus spatial (stationnarité, isotropie, covariance, etc.) sont définies dans la section 2.1. Ensuite, la question de la prédiction spatiale par *krigeage* est étudiée dans la section 2.2. Cette méthode nécessite de connaître la covariance du processus étudié : le problème d'estimation de la covariance est décrit dans la section 2.3.

2.1 Propriétés d'un processus spatial

2.1.1 Stationnarité

Soit $X := \{X(s)\}_{s\in\mathcal{S}}$ un processus spatial, on rappelle que X est entièrement caractérisé par ses lois jointes fini-dimensionnelles, c'est-à-dire les lois de $\big(X(s_1), \dots, X(s_d)\big)$, pour tout ensemble fini de sites $\{s_1, \dots, s_d\} \subset \mathcal{S}$. Le processus spatial X est dit *de second ordre* si pour tout $s \in \mathcal{S}$, le moment d'ordre 2 de la variable $X(s)$ existe : $\mathbb{E}\big[X(s)^2\big] < \infty$. On note alors, pour tout $s, s' \in \mathcal{S}$, la moyenne et la covariance de X respectivement par :

$$\begin{array}{rcccl} \mu_X & : & \mathcal{S} & \longrightarrow & \mathbb{R} \\ & & s & \longmapsto & \mathbb{E}\big[X(s)\big] \end{array} \quad \text{et} \quad \begin{array}{rcccl} C_X & : & \mathcal{S}\times\mathcal{S} & \longrightarrow & \mathbb{R} \\ & & (s,s') & \longmapsto & \mathrm{Cov}\big(X(s), X(s')\big) = \mathbb{E}\big[X(s)X(s')\big] \end{array} .$$

La fonction de covariance C_X est semi-définie positive, c'est-à-dire que pour tout $d \geqslant 1$, $a \in \mathbb{R}^d$ et tout ensemble $\{s_1, \dots, s_d\} \subset \mathcal{S}$:

$$\sum_{i,j=1\dots d} a_i a_j C_X(s_i, s_j) \geqslant 0 .$$

La définition 2 ci-dessous introduit la notion de stationnarité pour un processus spatial X de second ordre.

Définition 2 (Stationnarité d'un processus spatial de second ordre)**.** *Un processus X de second ordre est dit stationnaire au sens fort si sa loi est invariante par translation :*

$$\big(X(s_1 + h), \dots, X(s_d + h)\big) \overset{\mathcal{L}}{=} \big(X(s_1), \dots, X(s_d)\big) \ , \quad \forall s_1, \dots, s_d, h \in \mathcal{S} .$$

Le processus X est dit stationnaire au sens faible (ou d'ordre 2) si sa moyenne et sa covariance sont invariantes par translation :

$$\mu_X(s_1) = \mu_X(s_1 + h) \text{ et } C_X(s_1, s_2) = C_X(s_1 + h, s_2 + h) \ , \quad \forall s_1, s_2, h \in \mathcal{S} .$$

Si X est un processus stationnaire d'ordre 2, alors sa moyenne est constante $\mu_X(s) = \mu$ pour tout $s \in \mathcal{S}$. De plus, il existe une fonction paire définie positive $\kappa_X : \mathcal{S} \longrightarrow \mathbb{R}_+$ telle que pour tous $s, s' \in \mathcal{S}$, on a :

$$C_X(s, s') = \kappa_X(s - s') \ .$$

En particulier, la variance du processus spatial X est constante : $\mathrm{Var}\big(X(s)\big) = \kappa_X(0)$ pour tout $s \in \mathcal{S}$. On peut alors définir la fonction de corrélation ρ_X associée au processus X en fonction de la différence $h = s - s'$:

$$\begin{array}{rccl} \rho_X & : & \mathcal{S} & \longrightarrow & [-1, 1] \ , \\ & & h & \longmapsto & \kappa_X(h)/\kappa_X(0) \ . \end{array}$$

Un processus spatial X stationnaire d'ordre 2 est dit *isotrope* si sa covariance ne dépend que de la distance euclidienne entre les positions (et non pas de l'orientation du vecteur $\overrightarrow{ss'}$). Autrement dit X est un processus isotrope si :

$$\kappa_X(s - s') = \kappa_X^{(\mathrm{iso})}(h) \ , \quad h = ||s - s'|| \ .$$

La condition de stationnarité d'ordre 2 n'est pas toujours vérifiée pour un processus spatial : c'est par exemple le cas du mouvement brownien. Une notion plus fine est introduite dans la définition 3 : celle d'un processus intrinsèque.

Définition 3 (Processus spatial intrinsèque)**.** *Un processus spatial de second ordre est dit intrinsèque (ou à incréments stationnaires) s'il est de moyenne constante et si pour tout $h \in \mathcal{S}$, le processus*

$$D_h = \{X(s + h) - X(s)\}_{s \in \mathcal{S}}$$

est stationnaire d'ordre 2.

Pour un processus intrinsèque X, la structure spatiale est modélisée par le *semivariogramme* γ défini pour tout $s, h \in \mathcal{S}$ par :

$$2\gamma_X(h) := \mathrm{Var}\big(X(s + h) - X(s)\big) \ .$$

Plusieurs propriétés mathématiques caractérisent le semivariogramme :

1. γ_X est une fonction positive, paire et nulle à l'origine.
2. γ_X est une fonction définie-négative.
3. Si le processus X est stationnaire d'ordre 2, alors le semivariogramme γ_X est borné et s'écrit :

$$\gamma_X(h) = \kappa_X(0) - \kappa_X(h) = \kappa_X(0)\big(1 - \rho_X(h)\big) \ ,$$

où ρ_X est le *corrélogramme* de X (ou fonction de corrélation).

2.1.2 Modèles de covariance

Soit X un processus stationnaire d'ordre 2, on s'intéresse dans cette section à des modèles de covariance (et de façon équivalente, à des semivariogrammes) qui sont généralement utilisés dans la littérature.

Si la fonction de corrélation $\rho_X(h)$ tend vers 0 quand $||h|| \to \infty$, X est dit *ergodique*. Dans ce cas, le semivariogramme associé γ_X affichera un *palier* (*sill* en anglais), noté $\delta = \kappa_X(0)$ pour $||h|| \to \infty$. La distance h nécessaire pour atteindre ce palier est appelée la *portée*. Comme cette distance est infinie pour la plupart des modèles de covariance, il est préférable d'utiliser la *portée pratique* λ définie par :

$$\gamma_X(\lambda) = 0.95 \ \delta \ .$$

Parmi les modèles paramétriques de corrélogrammes ρ_X employés usuellement dans la littérature, on peut citer :

– le modèle exponentiel généralisé (*Powered Exponential* en anglais) :

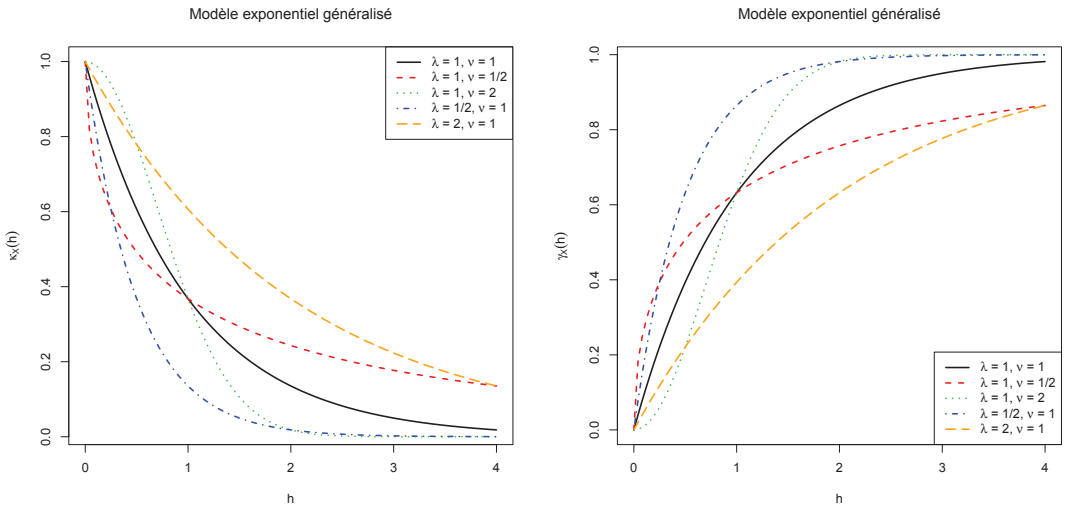
$$\rho_X(h) = \exp\left(-\left(\frac{||h||}{\lambda}\right)^{\nu}\right) \ ,$$

où ν est appelé paramètre de *lissage*.

– le modèle de *Whittle-Matérn* (Matérn, 1986) :

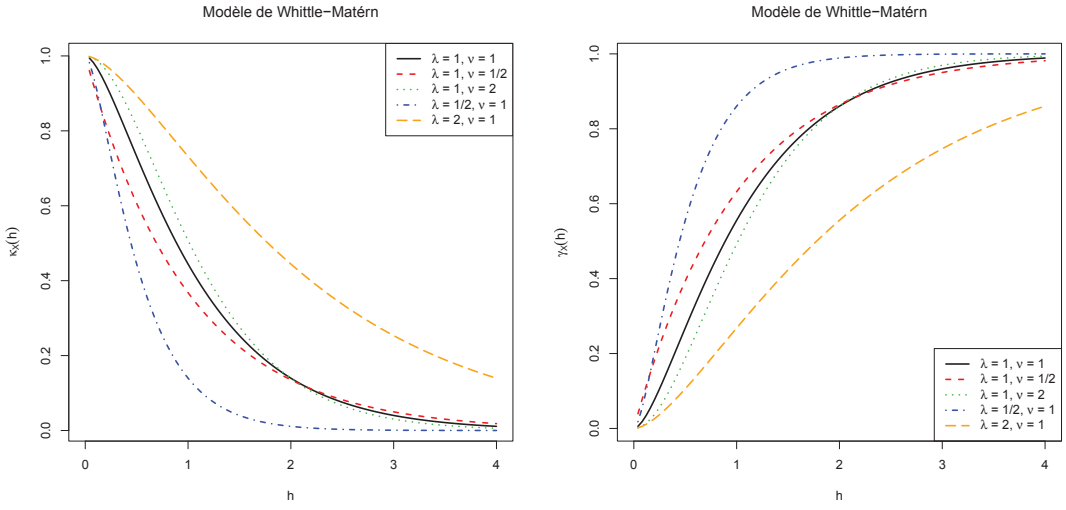
$$\rho_X(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{||h||}{\lambda}\sqrt{2\nu}\right)^{\nu} K_\nu\left(\frac{||h||}{\lambda}\sqrt{2\nu}\right) \ ,$$

où ν est le paramètre de lissage, Γ est la fonction gamma et K_ν la fonction de Bessel de seconde espèce (cf. Abramowitz et Stegun, 1964, Chapitre 9).



(a) Fonction de corrélation ρ_X.

(b) Semivariogramme γ_X.



(c) Fonction de corrélation ρ_X.

(d) Semivariogramme γ_X.

FIGURE 2.1 – Exemples de tracé de fonctions de corrélation et de semivariogrammes pour deux modèles paramétriques : exponentiel généralisé et Whittle-Matérn.

Le paramètre de lissage ν a une grande influence sur la qualité de prédiction spatiale (Stein, 1999). Il contrôle la régularité du semivariogramme en 0 ainsi que celle du processus spatial X (cf. Guyon, 2007, Section 1.3).

Les deux modèles de fonction de corrélation sont illustrés sur la figure 2.1 pour plusieurs valeurs des paramètres λ et ν. Les figures de gauche représentent les corrélogrammes ρ_X et les figures de droite affichent les semivariogrammes γ_X équivalents.

Les modèles de fonctions de corrélation et de semivariogrammes présentés sur la figure 2.1 sont isotropes (i.e ils ne dépendent que de la distance $||h||$ et non de l'orientation). Pour obtenir un corrélogramme anisotrope dans $\mathbb{R}^p$, on peut poser :

$$\rho_X(h) = \prod_{i=1}^{p} \exp\left(-\frac{1}{2}(h_i, \lambda_i)^2\right) ,$$

en prenant pour exemple le modèle exponentiel généralisé, avec un paramètre de portée λ_i associé à chaque direction h_i.

2.2 Le krigeage

Le but classique de la géostatistique est la prédiction spatiale du processus X en un point $s^* \in \mathcal{S}$ sachant qu'on dispose d'observations en $\{s_1, \ldots, s_d\}$. En supposant que X est stationnaire de second ordre avec un modèle de covariance connu, l'outil de prédiction utilisé en géostatistique est *le krigeage*.

Afin de simplifier l'écriture dans les paragraphes suivants, on utilise les notations suivantes :
- $X_{1:d} = \big(X(s_1), \dots, X(s_d)\big)$ pour désigner les observations du processus X,
- $\mu_s = \mathbb{E}\big[X(s)\big]$ pour tout $s \in \mathcal{S}$, et $\mu_{1:d} = (\mu_{s_1}, \dots, \mu_{s_d})$ pour désigner la moyenne du processus X sur l'ensemble de sites $\{s_1, \dots, s_d\}$,
- $\Sigma_{1:d} = \big[C_X(s_i, s_j)\big]_{1 \geqslant i,j \geqslant d}$ la matrice de variance-covariance observée et
- $\Sigma_{s^*} = \big(C_X(s^*, s_1), \dots, C_X(s^*, s_d)\big)$ le vecteur de covariances entre le site cible s^* et l'ensemble $\{s_1, \dots, s_d\}$.

2.2.1 Méthodes

Soit X un processus stationnaire d'ordre 2 de moyenne μ_X et covariance C_X. Selon si l'on connaît μ_X et/ou C_X, plusieurs outils de krigeage sont disponibles comme le krigeage *simple* et sa généralisation, le krigeage *universel*. Ces deux méthodes sont expliquées dans cette section.

Krigeage simple

On suppose que la moyenne μ_X et la fonction de covariance C_X du processus observé X sont connues et que la moyenne est constante sur $\mathcal{S}$: $\mu_s = \mu$ pour tout $s \in \mathcal{S}$. Le prédicteur du krigeage simple $\widehat{X}_{\mathrm{KS}}(s^*)$ en une position $s^* \in \mathcal{S}$ est alors défini par :

$$\widehat{X}_{\mathrm{KS}}(s^*) = \mu + \Sigma_{s^*}\Sigma_{1:d}^{-1}(X_{1:d} - \mu \mathbb{1}_d)^T \ ,$$

où $\mathbb{1}_d$ est le vecteur de taille d composé de 1. Le prédicteur $\widehat{X}_{\mathrm{KS}}(s^*)$ est le BLUP (*Best Linear Unbiased Predictor*), ou meilleure prédiction linéaire (au sens des moindres carrés) sans biais.

La variance du krigeage simple correspond à l'erreur quadratique moyenne :

$$\widehat{\sigma}^2_{\mathrm{KS}}(s^*) = C_X(s^*, s^*) - C_{s^*}\Sigma_s^{-1}C_{s^*}^T \ .$$

Entre autres, le package R nommé `DiceKriging` (Roustant *et al.*, 2012) ou encore la fonction `kriging` implémentée dans le package `SpatialExtremes` permettent de calculer la prédiction par krigeage simple $\widehat{X}_{\mathrm{KS}}(s^*)$ et la variance $\widehat{\sigma}^2_{\mathrm{KS}}(s^*)$ associée.

Krigeage universel

Dans la plupart des applications, il est plus réaliste de supposer la moyenne de X non constante et inconnue. Si cette moyenne est décrite par une régression linéaire sur des covariables, la prédiction spatiale utilise la méthode du krigeage universel.

On suppose que $\mu_X(s) = k(s)\beta^T$, où $k(s) = \big(k_1(s), \dots, k_m(s)\big)$ représente un ensemble de m covariables associées à la position $s \in \mathcal{S}$ et $\beta = \big(\beta_1, \dots, \beta_m\big)$ sont les coefficients de la régression linéaire. Par exemple, les covariables usuelles considérées pour des phénomènes climatiques sont la longitude, la latitude, l'altitude et une coordonnée à l'origine.

L'estimateur du krigeage universel $\widehat{X}_{\mathrm{KU}}(s^*)$ de $X(s^*)$ est le BLUP, défini par :

$$\widehat{X}_{\mathrm{KU}}(s^*) = k(s^*)\hat{\beta}^T + \Sigma_{s^*}\Sigma_{1:d}^{-1}\big(X_{1:d}^T - K_{1:d}\hat{\beta}^T\big) \ ,$$

où $\hat{\beta}^T = \big(K_{1:d}^T\Sigma_{1:d}^{-1}K_{1:d}\big)^{-1}K_{1:d}^T\Sigma_{1:d}^{-1}X_{1:d}^T$ est l'estimateur des moindres carrés de β^T, et où

$$K_{1:d} = \begin{bmatrix} k_1(s_1) & \dots & k_m(s_1) \\ \vdots & \ddots & \vdots \\ k_1(s_d) & \dots & k_m(s_d) \end{bmatrix} \text{, supposée de plein rang.}$$

La variance de prédiction pour le krigeage universel est donnée par :

$$\widehat{\sigma}^2_{\mathrm{KU}}(s^*) = C_X(s^*, s^*) - \Sigma_{s^*}\Sigma_{1:d}^{-1}\Sigma_{s^*}^T + \big(k(s^*)^T - K_{1:d}^T\Sigma_{1:d}^{-1}\Sigma_{s^*}^T\big)^T\big(K_{1:d}^T\Sigma_{1:d}^{-1}K_{1:d}\big)^{-1}\big(k(s^*)^T - K_{1:d}^T\Sigma_{1:d}^{-1}\Sigma_{s^*}^T\big) \ .$$

Un cas particulier du krigeage universel, le nom de *krigeage ordinaire*, concerne le cas où la moyenne de X est une constante inconnue.

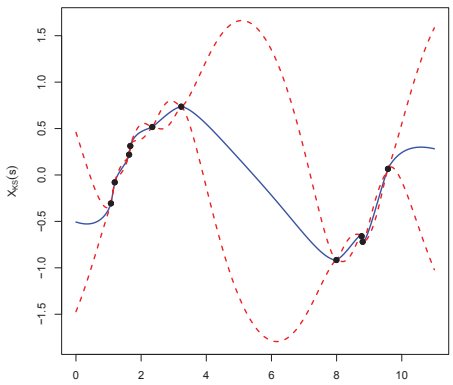


FIGURE 2.2 – Exemple d'estimation par krigeage et intervalle de confiance à 95% associé en dimension $p = 1$.

2.2.2 Illustration

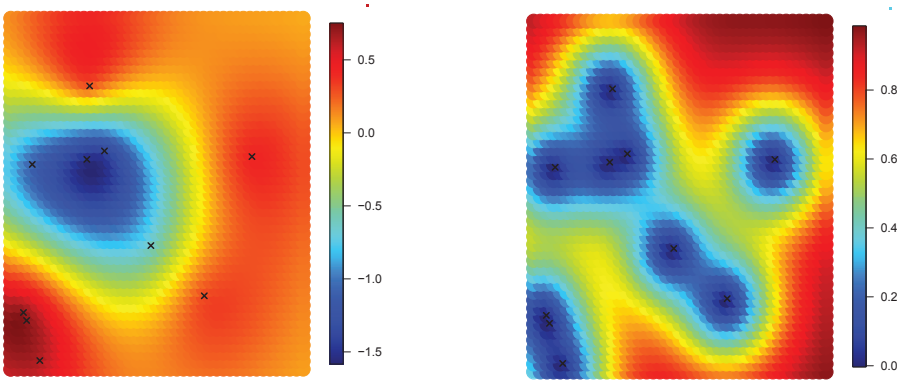
La figure 2.2 illustre le krigeage simple d'un processus gaussien X observé en $d = 10$ positions. Les points noirs correspondent aux observations du processus X, la courbe bleue correspond à la prédiction par krigeage et les courbes rouges pointillées à l'intervalle de confiance à 95% calculé pour tout s^* par :

$$\mathrm{CI}_{95\%}(s^*) = \widehat{X}_{\mathrm{KS}}(s^*) \pm 1.96\ \widehat{\sigma}_{\mathrm{KS}}(s^*)\ ,$$

où 1.96 correspond au quantile à 97.5% de la loi normale centrée réduite.

Le prédicteur du krigeage est *interpolant*, c'est-à-dire que $\widehat{X}_{\mathrm{KS}}(s_j) = X(s_j)$ et $\widehat{\sigma}^2_{\mathrm{KS}}(s_j) = 0$ pour toute position observée $s_j \in \{s_1, \ldots, s_d\}$. On voit en effet sur la figure 2.2 que la courbe bleue passe par les points noirs qui correspondent aux observations. Plus la distance entre une cible donnée s^* et l'ensemble $\{s_1, \ldots, s_d\}$ est grande, plus l'intervalle de confiance associé sera grand. C'est ce qu'on observe sur la figure 2.2 autour de l'abscisse 6. La variance s'agrandit encore plus aux extrémités de la région, jusqu'à atteindre la valeur limite du palier δ : c'est ce qu'on appelle *l'effet de bord*.

La figure 2.3 illustre la même méthode en dimension $p = 2$ où les sites observés $\{s_1, \ldots, s_{10}\}$ sont tirés aléatoirement sur une région spatiale et représentés par les croix noires. La figure 2.3a indique l'estimation moyenne du krigeage et la partie de droite la variance associée à chaque position cible.



(a) Moyenne $\widehat{X}_{\mathrm{KS}}$ du krigeage. (b) Variance σ^2_{KS} du krigeage.

FIGURE 2.3 – Exemple d'estimation par krigeage en dimension $p = 2$.

Sur la figure 2.3b, l'effet de bord est aussi bien représenté par une variance minimale autour des positions observées (les croix noires) et maximale à l'extérieure de l'enveloppe convexe de l'ensemble $\{s_1, \ldots, s_d\}$.

2.3 Estimation de la covariance

Les formules de prédiction présentées dans la section 2.2 supposent que la covariance du processus observé X est connue ainsi que ses paramètres. En pratique, ce n'est pas le cas mais il existe des méthodes paramétriques par exemple permettant de l'estimer dans un des modèles décrits dans la section 2.1.2.

Une approche descriptive possible est celle de la nuée variographique, qui permet de comparer un choix de modèle de covariance avec des estimations empiriques du semivariogramme γ_X. Si on suppose que le processus X est isotrope, on peut afficher le nuage variographique correspondant aux $K = \frac{d(d+1)}{2}$ estimations de γ_X pour toutes les paires de sites :

$$\mathcal{N}_X := \left\{ \left(||s_i - s_j|| \ , \ \frac{1}{2}\big(X(s_i) - X(s_j)\big)^2 \right) \right\}_{1 \leqslant i < j \leqslant d} .$$

Un exemple de nuée variographique pour $d = 50$ positions est donné sur la figure 2.4, où le semivariogramme exact ayant servi à simuler le processus X est indiqué par la ligne rouge.

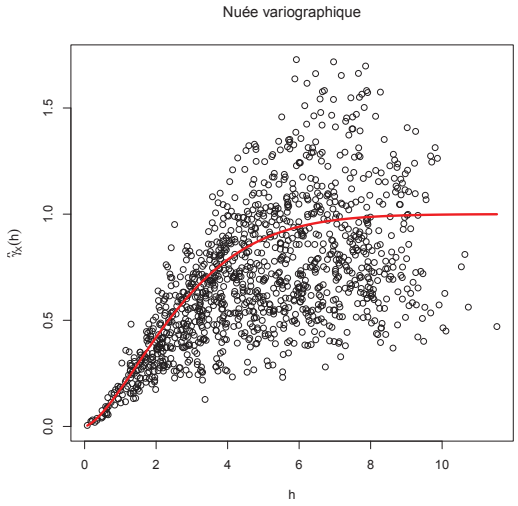


FIGURE 2.4 – Illustration de la nuée variographique avec le semivariogramme γ_X correspondant.

Cette figure permet de valider le choix d'un modèle $\gamma_X(\cdot;\theta)$ pour le semivariogramme mais ne permet pas de donner une estimation du vecteur de paramètres θ de la covariance. Pour cela, comme indiqué par Guyon (2007), il est possible d'utiliser l'estimateur des moindres carrés pondérés (MCP) :

$$\widehat{\theta}_{\text{MCP}} = \arg\min_{\theta \in \Theta} \sum_{k=1}^{K} \frac{|N(h_k)|}{\gamma_X^2(h_k;\theta)} \big(\widehat{\gamma}_{X,n}(h_k) - \gamma_X(h_k;\theta)\big)^2 \tag{2.1}$$

où Θ est l'espace dans lequel est défini le vecteur de paramètres θ, $|N(h)|$ est le cardinal de l'ensemble

$$N(h) = \big\{(s_i, s_j) \ : \ h - \Delta \leqslant s_i - s_j \leqslant h + \Delta \ ; \ i, j = 1, \ldots, d\big\} \ ,$$

est une classe approximante à tolérance Δ de la distance h entre deux positions. Le nombre K dans (2.1) correspond au nombre de classes retenues pour l'estimation empirique $\widehat{\gamma}_{X,n}$ du semivariogramme :

$$\widehat{\gamma}_{X,n}(h) = \frac{1}{2|N(h)|} \sum_{s_i, s_j \in N(h)} \big(X(s_i) - X(s_j)\big)^2 \ .$$

L'incertitude associée à l'estimation du vecteur de paramètres θ n'est pas prise en compte lors de la prédiction par krigeage simple ou krigeage universel. Une possibilité pour pallier ce problème serait de se servir du cadre bayésien (voir chapitre 3 et en particulier la section 3.1.3) en utilisant la loi prédictive de θ.

Chapitre 3

Inférence bayésienne et algorithmes MCMC

Ce chapitre introduit les notions de base de l'approche bayésienne, ainsi que les algorithmes MCMC les plus courants. De nombreux ouvrages traitent de cette approche, parmi lesquels on peut citer Gelman *et al.* (2014) ou Boreux *et al.* (2010).

3.1 L'inférence bayésienne

L'inférence bayésienne est une méthode d'estimation, à l'instar de la méthode des moindres carrés, du maximum de vraisemblance ou de la méthode des moments. Dans le cas d'une estimation au moyen d'un modèle paramétrique, l'approche bayésienne se différencie des autres par le fait qu'elle considère le vecteur de paramètres $\theta \in \Theta \subset \mathbb{R}^p$ comme aléatoire et non pas comme une valeur fixe qu'il faut estimer et à laquelle on associe une erreur standard.

3.1.1 L'équation de Bayes

L'équation de Bayes est au centre du paradigme bayésien. Elle est donnée par la formule :

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\pi(x)} \ , \tag{3.1}$$

où :

- $\pi(x|\theta)$ est la vraisemblance des données par rapport au modèle fixé. Il s'agit de la probabilité d'observer les données x sachant le modèle et le paramètre (ou vecteur de paramètres) θ associé. L'estimation par maximum de vraisemblance cherche à maximiser $\pi(x|\theta)$ et à trouver l'argument θ qui renvoie la valeur maximale :
$$\widehat{\theta} := \arg\max_{\theta \in \Theta} \pi(x|\theta) \ .$$
- $\pi(\theta)$ est la fonction de densité de la loi *a priori*. Il s'agit de l'information que l'on a sur ce paramètre avant d'avoir regardé les données x. Cette information peut être par exemple la connaissance d'un expert sur le phénomène physique étudié. Une loi a priori dite *vague* est une loi qui n'apporte pas ou peu d'information, comme la loi uniforme ou une loi normale ayant une variance élevée.
- $\pi(\theta|x)$ est la fonction de densité de la loi *a posteriori*. C'est la quantité que l'on cherche à estimer dans un modèle paramétrique : il s'agit de la loi du paramètre θ construite à partir de l'a priori de l'expert $\pi(\theta)$ et des données x à travers la vraisemblance $\pi(x|\theta)$.
- $\pi(x)$ est la constante de renormalisation. Pour que la fonction $\pi(\cdot|x)$ soit une fonction de densité à support dans $\Theta \in \mathbb{R}^p$, il faut avoir $\int_\Theta \pi(\theta|x)d\theta = 1$. Cette constante permet d'assurer cette condition. Par la formule des probabilités totales, on a :
$$\pi(x) = \int_\Theta \pi(x|\theta)\pi(\theta)d\theta \ .$$
Si la dimension du paramètre θ est $p = 1$, calculer cette constante de renormalisation est généralement simple, mais cela devient rapidement compliqué en dimension supérieure.

En général, on écrit la formule de Bayes avec le signe "proportionnel à" pour ne pas écrire la constante de renormalisation :

$$\pi(\theta|x) \propto \pi(x|\theta)\pi(\theta) \ .$$

L'approche bayésienne est ainsi basée sur le paradigme de l'apprentissage : on possède une connaissance a priori du phénomène, et en l'associant avec les observations, on obtient une estimation a posteriori. Si la loi a priori est vague (autrement dit non informative, avec une masse équivalente sur tout le domaine de définition de θ), l'estimation repose (presque) entièrement sur la vraisemblance des données. Dans le cas contraire, si l'expert apporte un a priori très informatif, le poids alloué aux observations sera plus faible.

3.1.2 A priori conjugués

Parfois, en choisissant une loi a priori de façon judicieuse, on peut connaître la loi a posteriori de façon explicite. C'est le cas pour les deux exemples illustrés dans cette section.

Premier exemple : loi normale de variance connue

Soit $X \sim \mathcal{N}(\theta, \sigma^2)$, où σ^2 est la variance (connue) et θ est le paramètre que l'on souhaite estimer. Soient $X_1, \ldots, X_n$ des répliques indépendantes de X. On fixe une loi a priori normale pour le paramètre θ :

$$\theta \sim \mathcal{N}\left(m_\theta, \sigma_\theta^2\right) \ ,$$

où m_θ est la moyenne et σ_θ^2 est la variance de θ.

Dans ce cas, la loi a posteriori de θ ayant observé $x = \{x_1, \ldots, x_n\}$ est aussi la loi normale :

$$\pi(\theta|x) \ = \ \mathcal{N}(m_*, \sigma_*^2) \ ,$$

où :

$$\left\{ \begin{array}{rcl} m_* & = & \left(\sigma_\theta^{-2} + \frac{n}{\sigma^2}\right)^{-1} \left(\frac{m_\theta}{\sigma_\theta^2} + \frac{n\bar{x}}{\sigma^2}\right) \\ \sigma_* & = & \left(\sigma_\theta^{-2} + \frac{n}{\sigma^2}\right)^{-1} \end{array} \right. ,$$

et où $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ est la moyenne empirique.

Démonstration. On a $X \sim \mathcal{N}(\theta, \sigma^2)$, donc $\pi(x|\theta) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2\right)$ et comme $\theta \sim \mathcal{N}(m_\theta, \sigma_\theta^2)$, on a $\pi(\theta) \propto \exp\left(-\frac{(\theta - m_\theta)^2}{2\sigma_\theta^2}\right)$.

En appliquant la formule de Bayes (3.1), on obtient :

$$\begin{array}{rcl} \pi(\theta|x) & \propto & \pi(\theta)\pi(x|\theta) \\ & = & \exp\left(-\dfrac{(\theta - m_\theta)^2}{2\sigma_\theta^2}\right) \exp\left(-\dfrac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right) \\ & = & \exp\left(-\dfrac{\theta^2 - 2m_\theta\theta + m_\theta^2}{2\sigma_\theta^2} - \dfrac{\sum_{i=1}^n x_i^2 - 2n\bar{x}\theta + n\theta^2}{2\sigma^2}\right) \\ & \propto & \exp\left(-\dfrac{1}{2}\left[\theta^2\left(\sigma_\theta^{-2} + \dfrac{n}{\sigma^2}\right) - 2\theta\left(\dfrac{m_\theta}{\sigma_\theta^2} + \dfrac{n\bar{x}}{\sigma^2}\right)\right]\right) \\ & \propto & \exp\left(-\dfrac{1}{2}\underbrace{\left(\sigma_\theta^{-2} + \dfrac{n}{\sigma^2}\right)}_{\sigma_*^{-2}}\left(\theta - \underbrace{\left(\sigma_\theta^{-2} + \dfrac{n}{\sigma^2}\right)^{-1}\left(\dfrac{m_\theta}{\sigma_\theta^2} + \dfrac{n\bar{x}}{\sigma^2}\right)}_{m_*}\right)^2\right) \ . \end{array}$$

On en déduit donc que $\pi(\theta|x)$ est la loi normale de paramètres (m_*, σ_*^2). □

Second exemple : loi normale de moyenne connue

Soient $X \sim \mathcal{N}(m, \theta)$, où m est la moyenne (connue) et θ est la variance (inconnue), et $X_1, \ldots, X_n$ des répliques iid de X. On choisit cette fois une loi a priori inverse-gamma pour θ, de paramètre de forme α_θ et de paramètre d'intensité β_θ. Autrement dit :

$$\begin{array}{rcl} \theta & \sim & \text{InvGamma}(\alpha_\theta, \beta_\theta) \ , \\ \pi(\theta) & \propto & (1/\theta)^{\alpha_\theta + 1} \exp\left(-\beta_\theta/\theta\right) \ . \end{array}$$

Dans ce cas, la loi a posteriori de θ, avec les observations $\{x_1, \ldots, x_n\}$ est aussi une loi inverse-gamma, de paramètres de forme α_* et d'intensité β_*, donnés respectivement par :

$$\left\{ \begin{array}{rcl} \alpha_* & = & \alpha_\theta + \frac{n}{2} \\ \beta_* & = & \beta_\theta + \frac{1}{2}\sum_{i=1}^n (x_i - m)^2 \end{array} \right. .$$

Démonstration. En utilisant la formule de Bayes (3.1), on obtient :

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta)\pi(x|\theta) \\ &\propto (1/\theta)^{\alpha_\theta+1}\exp\left(-\beta_\theta/\theta\right)\times\theta^{-n/2}\exp\left(-\frac{\sum_{i=1}^n(x_i-m)^2}{2\theta}\right) \\ &\propto (1/\theta)^{\alpha_\theta-n/2+1}\exp\left(-\frac{\beta_\theta+\frac{1}{2}\sum_{i=1}^n(x_i-m)^2}{\theta}\right)\end{aligned}$$

On en déduit donc la forme inverse-gamma de la loi a posteriori sur θ :

$$\pi(\theta|x) \;=\; \text{InvGamma}\left(\alpha_\theta+\frac{n}{2},\beta_\theta+\frac{1}{2}\sum_{i=1}^n(x_i-m)^2\right)\;.$$

□

3.1.3 Prédiction avec l'approche bayésienne

A la place d'estimations ponctuelles auxquelles on peut associer un intervalle de confiance, l'inférence bayésienne fournit une loi a posteriori $\pi(\theta|x)$, qui contient toute l'information sur l'incertitude d'estimation. Pour différencier l'incertitude obtenue avec les autres approches, on parle d'intervalle de *crédibilité* au lieu d'intervalle de *confiance*.

Il est toujours possible d'obtenir des estimations ponctuelles du paramètre θ, en regardant par exemple la moyenne, la médiane ou encore le mode de la loi a posteriori obtenue. Pour calculer une erreur standard, on peut par exemple regarder l'écart-type a posteriori.

Quantité a posteriori

Souvent, on ne souhaite pas estimer directement les paramètres du modèle mais on s'intéresse plutôt à une quantité qui est fonction de ces paramètres. Le chapitre 5 de la thèse en est un exemple, puisqu'on s'intéresse à un niveau de retour z_T défini à partir des paramètres de la loi GEV μ, σ et ξ (cf. Section 1.1.4).

De façon générale, la quantité que l'on souhaite estimer est $\psi = f(\theta)$, où f est une fonction inversible connue (par exemple la formule (1.3)). En utilisant la formule du changement de variable, la loi a posteriori de la quantité d'intérêt ψ est :

$$\pi(\psi|x) = \frac{\partial}{\partial\psi}\big(f^{-1}(\psi)\big)\pi\big(f^{-1}(\psi)|x\big)\;.$$

Malheureusement, calculer cette quantité peut s'avérer très compliqué en pratique. Une méthode plus simple pour estimer ψ est d'utiliser des algorithmes de simulation MCMC (voir section 3.2).

Loi prédictive a posteriori

En général, le but recherché par la modélisation statistique est la *prédiction* de nouvelles valeurs. Si on observe $X = \{X_1,\ldots,X_n\}$, on veut pouvoir prédire $X^* = \{X_1^*,\ldots,X_m^*\}$. Il peut s'agir de valeurs futures dans le temps : $X^* = \{X_{n+1}^*,\ldots,X_{n+m}^*\}$, ou encore de valeurs observées à des positions différentes dans l'espace (cf. Chapitre 2).

Avec l'approche bayésienne, il est possible d'obtenir la loi prédictive a posteriori. En effet, on a :

$$\begin{aligned}\pi(x^*|x) &= \int_\Theta \pi(x^*,\theta|x)d\theta \\ &= \int_\Theta \pi(x^*|\theta,x)\pi(\theta|x)d\theta \\ &= \int_\Theta \pi(x^*|\theta)\pi(\theta|x)d\theta\;,\end{aligned}$$

où $\pi(x^*|\theta)$ est la vraisemblance des prédictions et $\pi(\theta|x)$ la loi a posteriori basée sur les données observées.

3.2 Algorithmes de simulation MCMC

Les algorithmes MCMC sont un ensemble de procédures qui permettent, entre autres, de simuler des réalisations d'une variable aléatoire selon une loi donnée. La condition nécessaire à ces algorithmes est que la fonction de densité f souhaitée puisse être évaluée en tout point x à un facteur multiplicatif près.

Ces méthodes sont très utiles en grande dimension, ou lorsque la fonction de densité n'a pas une forme simple et n'est connue qu'à une constante multiplicative près. C'est par exemple le cas pour les lois a posteriori obtenues par l'approche bayésienne, lorsqu'aucune loi a priori conjuguée ne peut être utilisée.

On notera ici f la fonction de densité cible qui représente par exemple la loi a posteriori $\pi(\cdot|x)$.

Plusieurs ouvrages traitent de ces algorithmes en détail. Un exemple est le livre de Robert et Casella (2009) (Chapitres 6, 7 et 8) ou encore le cours de Elie et Lapeyre (2001).

Dans cette section, trois méthodes populaires sont présentées : l'algorithme de Metropolis-Hastings, l'échantillonneur de Gibbs et l'algorithme de Metropolis-within-Gibbs qui regroupe les deux premiers.

3.2.1 Chaînes de Markov

Le résultat d'un algorithme MCMC est une suite $X = \{X_1, X_2, \ldots\}$ qui converge en loi, sous certaines conditions, vers la loi cible de densité f. Cette chaîne doit donc vérifier quelques propriétés énoncées dans cette section.

Définition d'une chaîne de Markov

On dit que $\{X_n\}_{n\geqslant 0}$ est une *chaîne de Markov* sur un espace d'états E de noyau de transition K si

$$\mathcal{L}(X_{n+1} \mid X_0, \ldots, X_n) = \mathcal{L}(X_{n+1} \mid X_n) \ ,$$

et si les propriétés suivantes sont vérifiées :

1. $K(x, y) \geqslant 0$ pour tous $x, y \in E$,
2. $\int_E K(x, y)dy = 1$,
3. $\Pr(X_{n+1} = x \mid X_n = x_n) = K(x_n, x)$.

Chaîne de Markov stationnaire

Soit μ une mesure positive sur l'espace d'états E. On dit que la chaîne de Markov $\{X_n\}_{n\geqslant 0}$ de noyau de transition K est stationnaire par μ si $\mu K = \mu$.

Pour montrer qu'une chaîne est stationnaire par μ, on peut montrer qu'elle est *réversible*, c'est-à-dire si pour tous $x, y \in E$ on a :

$$\mu(x)K(x, y) = \mu(y)K(y, x) \ . \tag{3.2}$$

Démonstration. Supposons que la chaîne $\{X_n\}_{n\geqslant 0}$ vérifie l'équation (3.2). Alors :

$$\begin{aligned}
(\mu K)(y) &:= \int_E \mu(x)K(x, y)dx \\
&= \int_E \mu(y)K(y, x)dx \\
&= \mu(y) \underbrace{\int_E K(y, x)dx}_{=1} \\
&= \mu(y) \ .
\end{aligned}$$

□

Les trois algorithmes MCMC présentés dans la section suivante permettent de simuler des chaînes de Markov qui vérifient la condition de réversibilité (3.2) pour la loi cible f. La chaîne obtenue est donc de loi stationnaire f.

3.2.2 Les algorithmes MCMC

L'algorithme de Metropolis-Hastings

L'algorithme MCMC le plus populaire est celui de Metropolis-Hastings (Metropolis *et al.*, 1953; Hastings, 1970; Tierney, 1994)). Sachant la fonction de densité souhaitée f, un état initial x_0 et une loi *de proposition* q, on simule la chaîne de Markov $\{x_1, \ldots, x_N\}$ de taille N en suivant l'algorithme 1.

Un candidat $x_{(c)}$ est proposé grâce à la loi de proposition q et s'il est accepté, il devient le nouvel état de la chaîne. Sinon, on conserve l'état précédent. La probabilité d'acceptation $p\big(x_{(c)}\big)$ est comparée à la réalisation d'une variable u de loi uniforme pour tester si le candidat est accepté ou non.

Algorithm 1 Metropolis-Hastings

for $r = 1, \ldots, N$ **do**
  On simule un état *candidat* $x_{(c)} \sim q$,
  On calcule la *probabilité d'acceptation* $p\big(x_{(c)}\big) = \min\left\{1, \frac{f(x_{(c)})q(x_{r-1})}{f(x_{r-1})q(x_{(c)})}\right\}$,
  On simule $u \sim \text{Unif}(0,1)$
  if $u < p\big(x_{(c)}\big)$ **then**
    $x_r = x_{(c)}$
  else
    $x_r = x_{r-1}$
  end if
end for

On peut remarquer que la loi cible f doit être connue à une constante multiplicative près car celle-ci disparaît dans le calcul de la probabilité d'acceptation $p\big(x_{(c)}\big)$. Ce point permet notamment de résoudre le problème du calcul de la constante de renormalisation $\pi(x)$ dans l'équation (3.1).

La loi de proposition q peut être choisie indépendamment de la loi cible f. Cependant, elle doit satisfaire plusieurs conditions, comme par exemple avoir un support contenant celui de f afin que l'algorithme puisse visiter ce dernier dans son intégralité. La condition $\sup_{x\in\mathbb{R}} f(x)/q(x) < \infty$ permet également de s'assurer que l'algorithme de Metropolis-Hastings soit efficace (Robert et Casella, 2009, Exemple 6.2).

Dans la plupart des cas, la loi de proposition est une marche aléatoire centrée sur l'état actuel de la chaîne de Markov. Par exemple, pour une étape $r \in \{1, \ldots, N\}$ de l'algorithme, on peut choisir la loi normale :

$$q\big(x_{(c)}|x_{r-1}\big) \propto \exp\left(-\frac{\big(x_{(c)} - x_{r-1}\big)^2}{2\sigma_q^2}\right) .$$

Si on choisit une marche aléatoire symétrique, le calcul de la probabilité d'acceptation $p\big(x_{(c)}\big)$ se simplifie et on obtient :

$$p\big(x_{(c)}\big) = \min\left\{1, \frac{f(x_{(c)})}{f(x_{r-1})}\right\} .$$

Même dans ce cas, la loi de proposition ne disparaît pas complètement : elle est toujours utilisée pour générer le candidat $x_{(c)}$. Le choix de la variance σ_q^2 est important. En effet, une trop grande variance a pour conséquence de générer des candidats qui peuvent être parfois trop à l'écart du centre de la loi f ciblée, alors qu'une trop petite variance fait évoluer la chaîne de Markov trop lentement.
Ce dernier point est discuté plus en détail dans la section 3.2.3.

L'échantillonneur de Gibbs

L'échantillonneur de Gibbs (Gelfand et Smith, 1990) est un cas particulier de Metropolis-Hastings, où la chaîne de Markov X que l'on veut simuler est à valeurs dans $\mathbb{R}^d$, pour $d \geqslant 2$. On souhaite donc obtenir des réalisations d'une loi multidimensionnelle f.

Pour cela, il est nécessaire de connaître les lois marginales conditionnelles

$$f_j(\cdot|x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d) ,$$

pour tout $j \in \{1, \ldots, d\}$. A la différence de Metropolis-Hastings, il n'est pas nécessaire de choisir une loi de proposition q.

L'échantillonneur de Gibbs est décrit dans l'algorithme 2 et étudié dans Casella et George (1992).

Cette méthode possède deux avantages : d'abord, elle permet de simuler selon une loi de très grande dimension, ce que l'algorithme de Metropolis-Hastings ne pouvait pas faire. Ensuite, elle est généralement plus rapide car il n'y a pas de test d'acceptation d'un candidat : le nouvel état de la chaîne est directement simulé selon la loi cible f. En revanche, le principal inconvénient est qu'il faut connaître la forme explicite de toutes les lois marginales conditionnelles. En pratique, ce n'est pas toujours le cas, et c'est pourquoi on utilise une méthode alternative : l'algorithme de Metropolis-within-Gibbs.

Algorithm 2 Échantillonneur de Gibbs

Choisir l'état initial de la chaîne $x_0 = (x_{1,0}, \ldots, x_{d,0})$.
for $r = 1, \ldots, N$ **do**
  Simuler $x_{1,r} \sim f_1(\cdot|x_{2,r-1}, \ldots, x_{d,r-1})$.
  Simuler $x_{2,r} \sim f_2(\cdot|x_{1,r}, x_{3,r-1}, \ldots, x_{d,r-1})$.
  $\vdots$
  Simuler $x_{j,r} \sim f_j(\cdot|x_{1,r}, \ldots, x_{j-1,r}, x_{j+1,r-1}, \ldots, x_{d,r-1})$.
  $\vdots$
  Simuler $x_{d,r} \sim f_d(\cdot|x_{1,r}, \ldots, x_{d-1,r})$.
end for

L'algorithme mixte : Metropolis-within-Gibbs

Cet algorithme, aussi appelé *algorithme de Gibbs avec proposition Metropolis* est le plus souvent utilisé dans les modèles statistiques complexes mettant en jeu un vecteur de paramètres θ de dimension assez grande.

Par exemple, c'est le cas dans le chapitre 6 où un algorithme de ce type est utilisé pour estimer les paramètres d'un modèle spatial de valeurs extrêmes possédant une structure hiérarchique assez complexe.

Le principe est similaire à celui de l'échantillonneur de Gibbs, mais il ne requiert pas de connaître les lois marginales conditionnelles $f_j(\cdot|\ldots)$. A la place, on simule un état candidat $x_{j,(c)}$ pour chaque élément marginal du vecteur x en utilisant des lois de propositions q_j. On calcule ensuite une probabilité d'acceptation de façon analogue à l'algorithme de Metropolis-Hastings.
L'algorithme 3 donne les étapes de la méthode de Metropolis-within-Gibbs.

Algorithm 3 Metropolis-within-Gibbs

Choisir un état initial $x_0 = (x_{1,0}, \ldots, x_{d,0})$.
for $r = 1, \ldots, N$ **do**
  for $j = 1, \ldots, d$ **do**
    Simuler $x_{j,(c)} \sim q_j$
    Calculer la probabilité d'acceptation $p_j\big(x_{(c)}\big) = \min\left\{1, \frac{f(x_{1,r},\ldots,x_{j-1,r},x_{j,(c)},x_{j+1,r-1},\ldots,x_{d,r-1})}{f(x_{1,r},\ldots,x_{j-1,r},x_{j,r-1},x_{j+1,r-1},\ldots,x_{d,r-1})} \frac{q_j(x_{j,r-1})}{q_j(x_{j,(c)})}\right\}$.
    Simuler $u_j \sim \text{Unif}(0,1)$.
    if $u_j < p_j\big(x_{(c)}\big)$ **then**
      $x_{j,r} = x_{j,(c)}$
    else
      $x_{j,r} = x_{j,r-1}$
    end if
  end for
end for

3.2.3 Évaluation de la convergence des chaînes

Les trois algorithmes présentés dans la section précédente permettent de simuler une chaîne de Markov $\{X_n\}_{n\geqslant 1}$ où les éléments X_n sont à valeurs dans $\mathbb{R}^d$, pour $d \geqslant 1$. La loi stationnaire de X est la fonction de densité f souhaitée (par exemple, la loi a posteriori), mais en pratique, la convergence peut être trop lente.

El Adlouni *et al.* (2006) recense et compare les méthodes d'évaluation de la convergence des chaînes de Markov (voir aussi le package R `CODA` (Plummer *et al.*, 2006)). Deux approches peuvent être citées :

1. celle de Geweke *et al.* (1991) qui consiste à diviser la chaîne de Markov en sous-parties et à tester si elles sont des réalisations de la même loi,
2. celle de Gelman *et al.* (2014) qui consiste à lancer indépendamment en parallèle plusieurs fois le même algorithme et à comparer les chaînes obtenues par chacun.

Plusieurs problèmes peuvent ainsi être détectés, auxquels on peut appliquer une solution pendant ou après avoir lancé l'algorithme MCMC.

Période de chauffe

Un problème possible rencontré lors de l'utilisation de ces algorithmes vient du choix des valeurs initiales x_0 de la chaîne de Markov. Si ce dernier est mauvais, dans le sens où il n'est pas situé dans le cœur de la loi f, les premières valeurs ne représentent pas la loi stationnaire limite.

Pour pallier ce problème, il est courant de définir une *période de chauffe* (*burn-in period* en anglais) que l'on retire du résultat final pour s'assurer que tous les éléments que l'on regarde sont bien distribués selon la loi cible.

Adaptation des chaînes

Si les lois de propositions sont des marches aléatoires, la variance σ_q^2 associée à ces lois donne l'amplitude des *sauts* réalisés dans le support de f. Le choix de σ_q aura donc un impact sur la vitesse de convergence de la chaîne de Markov.

- Si σ_q est trop petit, les candidats générés ont plus de chances d'être acceptés mais la chaîne de Markov évolue trop rapidement. Le support de f ne sera pas bien visité puisque la chaîne se concentrera sur le cœur de la distribution.
- Si σ_q est trop grand, les candidats sont souvent rejetés et par conséquent, la chaîne évolue trop lentement dans ce cas.

Afin de résoudre ce problème, il est possible d'adapter les sauts lors de la période de chauffe, en modifiant la valeur de l'amplitude des sauts σ_q en fonction du ratio d'acceptation sur un ensemble d'itérations. Si le ratio est trop élevé (resp. trop bas), cela signifie que l'amplitude des sauts est trop faible (resp. trop forte). On l'augmente (resp. diminue) pour obtenir une amplitude de saut "optimale". La littérature sur les algorithmes bayésiens donne la valeur de 0.44 comme valeur optimale du ratio d'acceptation (Roberts et Rosenthal, 2009).

Il est important de noter que la procédure d'adaptation des sauts dans l'algorithme MCMC doit se faire pendant la période de chauffe pour que la chaîne de Markov résultante vérifie la propriété de réversibilité décrite dans la section 3.2.1.

Thinning des chaînes

La chaîne de Markov obtenue par un algorithme MCMC est stationnaire mais est aussi corrélée de par sa construction. Certains éléments de la chaîne peuvent même être égaux d'une étape à l'autre si le candidat a été rejeté. En pratique, même si la quantité souhaitée porte souvent sur une statistique de l'échantillon (moyenne, médiane ou variance par exemple), il est pourtant souhaitable d'obtenir une chaîne indépendante et bien distribuée selon la loi stationnaire souhaitée f. Une méthode permettant d'obtenir un tel échantillon est le *thinning*.

Le principe est simple : on choisit de garder un état toutes les N_{th} itérations. Plus ce nombre est élevé, plus on supprime d'états au résultat de l'algorithme MCMC et plus il faut alors augmenter la valeur de N.

Certains modèles sont dits *non mélangeants* lorsque leur chaîne de Markov estimée par un algorithme MCMC présente systématiquement une évolution lente avec une corrélation fortement marquée. C'est par exemple le cas du modèle spatial de valeurs extrêmes étudié dans cette thèse. Ce point est discuté dans le chapitre 6 à la section 6.2.4.

Chapitre 4

Précipitations extrêmes en France

Ce chapitre présente les données de précipitations étudiées dans la thèse à travers plusieurs études descriptives sur la stationnarité des valeurs extrêmes, d'un point de vue temporel dans un premier temps, puis spatial dans un second temps.

L'objectif est de considérer avant tout l'évolution temporelle des valeurs extrêmes et de déterminer une région d'étude homogène afin d'appliquer par la suite des modèles spatiaux (voir Partie II).

4.1 Généralités

Les données de précipitations correspondent à des cumuls enregistrés (généralement en millimètres) sur un pas de temps donné. Si les données utilisées dans cette thèse utilisent des cumuls journaliers, il est possible de regarder des précipitations sur un pas de temps plus petit (horaire par exemple). Ces données sont enregistrées par des pluviomètres contrôlés quotidiennement : la précision de la mesure est donc généralement assez bonne (de l'ordre de 0.1mm), bien qu'un biais puisse apparaître en présence de vent ou de neige.

Les précipitations sont caractérisées par une proportion importante de jours *secs* avec un cumul enregistré de 0, qui correspondent aux jours non pluvieux. En hydrologie, il est usuel de distinguer en réalité plusieurs types d'enregistrements :

1. les jours *secs*, pour un cumul observé de 0mm,
2. les jours *humides*, pour un cumul positif inférieur à 1mm,
3. les jours *pluvieux*, pour un cumul supérieur (strictement) à 1mm.

Deux types de précipitations peuvent être distinguées : les pluies *frontales* et les pluies *convectives*. Les premières apparaissent généralement en hiver et montrent un champ spatial très étendu, de l'ordre de plusieurs dizaines de kilomètres. Les secondes apparaissent souvent en été lorsqu'un trou se forme entre la basse atmosphère où la masse d'air est chaude, et la haute atmosphère où l'air est plus froid. Ces dernières correspondent aux phénomènes orageux observables en été et en automne en France.

En pratique, différencier les deux types de précipitations à partir des données est très compliqué, surtout si les cumuls de précipitations sont observés à un pas de temps journalier et sur un réseau de stations relativement espacé. Pour identifier le phénomène convectif, il est souvent nécessaire d'associer les données de précipitations avec un autre type de données, comme l'intensité des éclairs frappant le sol (Soriano *et al.*, 2001; Tapia *et al.*, 1998) ou les courants aériens (Maraun *et al.*, 2011).

4.2 Jeux de données

Deux sources de données de précipitations au pas de temps journalier ont été utilisées au cours de la thèse. La première est enregistrée sur un petit nombre de stations réparties de façon espacée sur le territoire français, mais contient plusieurs séries de longueurs importantes (plus de 100 ans). Le second jeu de données concerne un réseau très dense de stations météorologiques sur une partie du pays et contient des séries de longueurs moyennes (de 30 à 57 ans).

4.2.1 European Climate Assessment and Dataset

Le premier jeu de données de précipitations considéré est issu de la base de données publique de ECA&D (European Climate Assessment and Dataset, cf. Klein Tank *et al.*, 2002; Klok et Klein Tank, 2009), disponible à l'adresse :

`eca.knmi.nl/` .

Cette base de données contient 51 séries de précipitations, mais un traitement de ces séries a mis en évidence :
- la présence de plusieurs séries doublons ou situées géographiquement à la même position (ce qui est le cas par exemple des stations nommées Lyon et Bron),
- d'enregistrements sur une période trop courte (environ 10 ans) pour que l'extrapolation de valeurs extrêmes soit raisonnable.

Après la suppression de ces séries, il reste 41 stations réparties sur la France, représentées sur la figure 4.1. Les stations météorologiques sont réparties de façon non uniforme sur le territoire : certaines régions sont vides (par exemple les Alpes, la Bourgogne, les pays de Loire ou les côtes de la Manche) alors que d'autres contiennent un agglomérat de stations (la région parisienne par exemple).

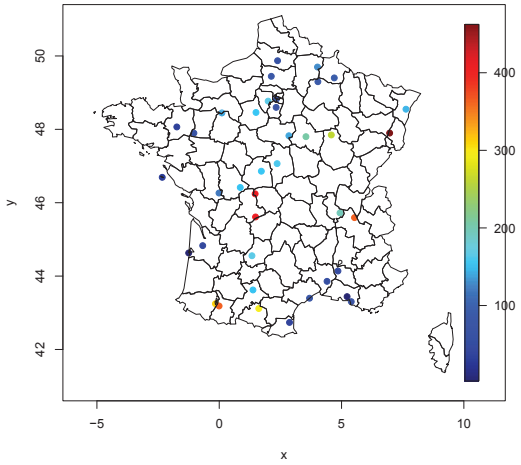


FIGURE 4.1 – Carte de la France et positions des stations météorologiques du jeu de données ECA&D avec un code couleur indiquant l'altitude par stations.

La table 4.1 recense les 41 stations choisies en indiquant son nom et :
- l'année de début et de fin d'enregistrement (données manquantes non incluses),
- le nombre total d'années disponibles,
- le pourcentage de jours pluvieux par rapport aux données disponibles,
- la valeur maximale de toute la série.

Les séries de précipitations de la base de données ECA&D présentent des dissimilarités. La longueur des données varie de 46 ans (Pamiers, dans l'Ariège) à 126 ans (Paris), mais certaines séries contiennent des trous de plusieurs années correspondant à l'une des deux guerres mondiales (voir Figures 4.4c et 4.4d).

Le climat varie également d'une position à l'autre, avec 39.85% de jours pluvieux à Oderen (Bas-Rhin) contre seulement 13.30% à Marignane (Bouches-du-Rhône). La valeur du maximum observé sur une station dépend aussi de sa situation géographique. La station ayant enregistré le plus petit maximum est celle de Ballots (Mayenne), située dans une région de climat océanique dégradé, avec un cumul journalier de 51.5mm. La valeur maximale de cumul de précipitations journalières de toutes les séries confondues est de 520mm, correspondant à la station météorologique de Mont-Aigoual (Gard), située dans le massif des Cévennes, région atypique concernant les précipitations extrêmes en France (voir section 4.4).

Au regard de ce résumé des données, il apparaît que les précipitations françaises ont un caractère très hétérogène. Par conséquent, il peut être prudent de regrouper les stations météorologiques par groupes homogènes en choisissant une région plus petite au climat semblable. Cependant, ce jeu de données est trop clairsemé spatialement pour réaliser un tel regroupement.

C'est pourquoi un second jeu de données a été analysé et est présenté dans la section 4.2.2 suivante.

4.2.2 Données Météo France

Description des données

Le second jeu de données utilisé se compose d'un total de 770 postes pluviométriques au pas de temps journalier regroupant 689 stations françaises, dont 356 postes de EDF/DTG et 333 de Météo France (MFR), 17 postes espagnols de l'INM (Instituto National de Meteorologia) et 64 stations suisses de Météo Suisse.

Nom	Début	Fin	Total	Manquantes	Jours pluvieux	Maximum
Alençon	1930	2005	76	13.01 %	30.76 %	67.2mm
Amiens	1929	2005	77	12.07 %	30.36 %	112mm
Auxerre	1956	2005	50	15.14 %	30.57 %	65.3mm
Ballots	1956	2005	50	18.56 %	30.48 %	51.5mm
Beauvais	1945	2005	61	12.72 %	30.55 %	64.7mm
Bordeaux	1920	2000	81	19.66 %	33.06 %	87.6mm
Bourges	1945	2011	67	4.13 %	30.84 %	79mm
Bretigny	1958	2005	48	15.69 %	28.98 %	92mm
Catus	1931	2005	75	12.09 %	26.81 %	112.8mm
Chartres	1946	2005	60	12.91 %	27.82 %	59.2mm
Châteauroux	1901	1999	99	12.33 %	30.79 %	66.1mm
Chatillon-Coligny	1939	2005	67	20.45 %	30.79 %	68mm
Chatillon-sur-Seine	1890	2005	116	9.25 %	33.22 %	81.2mm
Espoey	1934	2005	72	11.09 %	35.58 %	94.6mm
Faverges	1948	2005	58	15.96 %	31.03 %	120mm
Ile d'Yeu	1949	2005	57	13.51 %	30.83 %	94.1mm
Lège-Cap-Ferret	1885	2005	121	9.33 %	31.3 %	90mm
Lezay	1920	2005	86	18.17 %	30.64 %	91.5mm
Lyon	1920	2011	92	6.61 %	27.44 %	104.1mm
Marignane	1921	2011	91	6.59 %	13.3 %	161mm
Marseille	1881	2005	125	6.87 %	15.04 %	200mm
Le Massegros	1948	2005	58	14.16 %	30.02 %	173.8mm
Mont-Aigoual	1896	2011	116	5.53 %	34.11 %	520mm
Montcornet	1900	2005	106	12.67 %	34.25 %	84.3mm
Montmorillon	1937	2000	64	20.19 %	29.48 %	81.1mm
Nîmes	1920	2005	86	15.07 %	17.36 %	266.8mm
Oderen	1941	2005	65	15.93 %	39.85 %	145mm
Orange	1959	2005	47	15.97 %	18.1 %	219.2mm
Pamiers	1960	2005	46	20.35 %	27.78 %	92mm
Paris	1886	2011	126	2.21 %	28.93 %	104.2mm
Perpignan	1901	2011	111	3.23 %	14.78 %	222mm
Reims	1930	2005	76	17.51 %	29.61 %	67.3mm
Rennes	1944	2011	68	5.94 %	27.03 %	70mm
Saint Germain les Belles	1921	2005	85	13.57 %	36.47 %	106.8mm
Sète	1951	2005	55	13.94 %	14.84 %	151.2mm
La Souterraine	1942	2005	64	18.67 %	35.29 %	96.3mm
Strasbourg	1941	2011	71	11.52 %	26.73 %	141.2mm
Tarbes	1946	2005	60	12.99 %	32.71 %	79.7mm
Toulouse	1947	2011	65	4.36 %	23.44 %	83mm
Trappes	1918	2005	88	14.14 %	30.3 %	91.2mm
Vouziers	1890	2005	116	13.02 %	31.88 %	85.2mm

TABLE 4.1 – Informations sur les données des 41 stations la base ECA&D.

Ces données concernent la période 1953-2005 et ont été en particulier utilisées dans la thèse de David Penot (2014), qui construit une méthode de régionalisation de pluies journalières extrêmes pour estimer des crues et produire des cartes de niveaux de retour. Les travaux de la thèse de David Penot utilisent en particulier :

- le modèle probabiliste MEWP (*Multi-Exponential Weather Pattern*) de Garavaglia *et al.* (2010), qui décrit la distribution des cumuls journaliers de précipitations en découpant par type de temps,
- l'interpolateur de pluie SPAZM de Gottardi *et al.* (2012), méthode qui fournit une estimation des pluies en montagne en utilisant les données locales, l'information sur l'altitude et la classification par type de temps (Penot, 2014, p. 35).

La figure 4.2 représente les positions de ces stations sur la carte de la France. Un code couleur représente l'altitude (en mètres) correspondant à chaque localisation. Le jeu de données contient des stations aussi bien en basse altitude qu'en montagne, avec un maximum de 2031m dans les Alpes. Il est attendu que les valeurs extrêmes soient affectées par cette différence d'altitude, car les orages sont connus pour être plus violents en haute altitude qu'en plaine (voir conclusions de la partie II et cartes d'extrapolation du niveau de retour dans le chapitre 5).

On voit aussi que le réseau de stations est beaucoup plus dense que pour la base de données ECA&D, mais que le territoire n'est pas entièrement recouvert : ces données sont concentrées sur la moitié Sud-Est de la France.

La densité du réseau de stations de cette base de données fournit une information spatiale exhaustive sur les précipitations (voir Section 4.4).

Analyse de la qualité

Afin de supprimer les valeurs aberrantes, les décalages temporels et les ruptures chronologiques, le jeu de données a été homogénéisé au préalable, ce qui a fait l'objet de la thèse de Gottardi (2009). Trois critères de

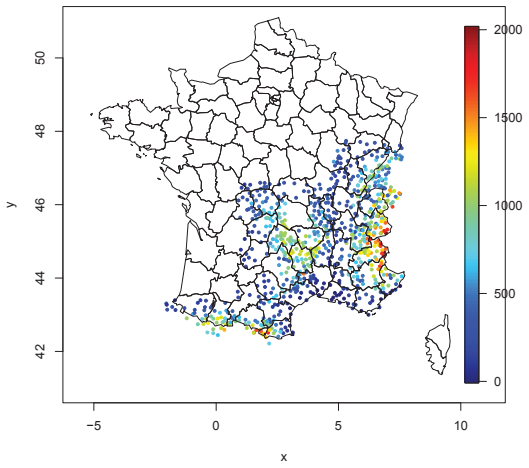
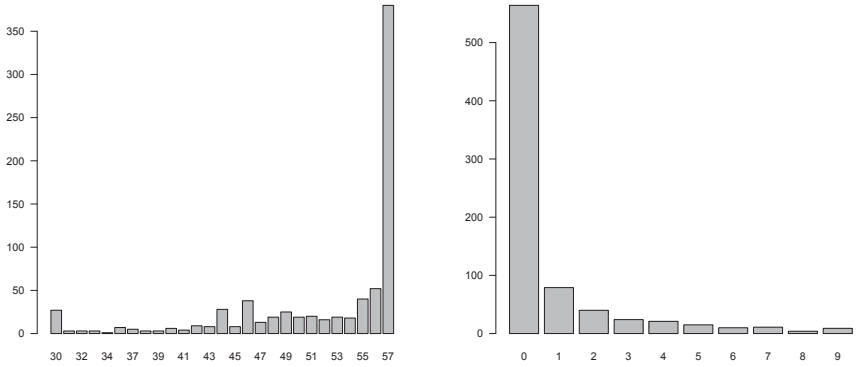


FIGURE 4.2 – Carte de la France et positions des stations météorologiques du jeu de données EDF-MFR avec un code couleur indiquant l'altitude par stations.

qualité sont vérifiés par chaque série de cumuls journaliers de cette base de données (Penot, 2014, p. 50) :

1. Elles sont toutes enregistrées sur 30 années équivalentes sans aucune donnée manquante.
2. Chaque station dispose de moins de 10% de données manquantes (située pour la plupart des séries sur la période la plus ancienne).
3. Chaque station propose des données disponibles jusqu'en 2005.

La qualité de ce jeu de données est évaluée en regardant la répartition du nombre de stations en fonction du nombre d'années disponibles (Figure 4.3a) et du pourcentage de valeurs manquantes (Figure 4.3b).



(a) Années disponibles. (b) Pourcentage de données manquantes.

FIGURE 4.3 – Nombre de stations en fonction (a) du nombre d'années disponibles et (b) du pourcentage de données manquantes.

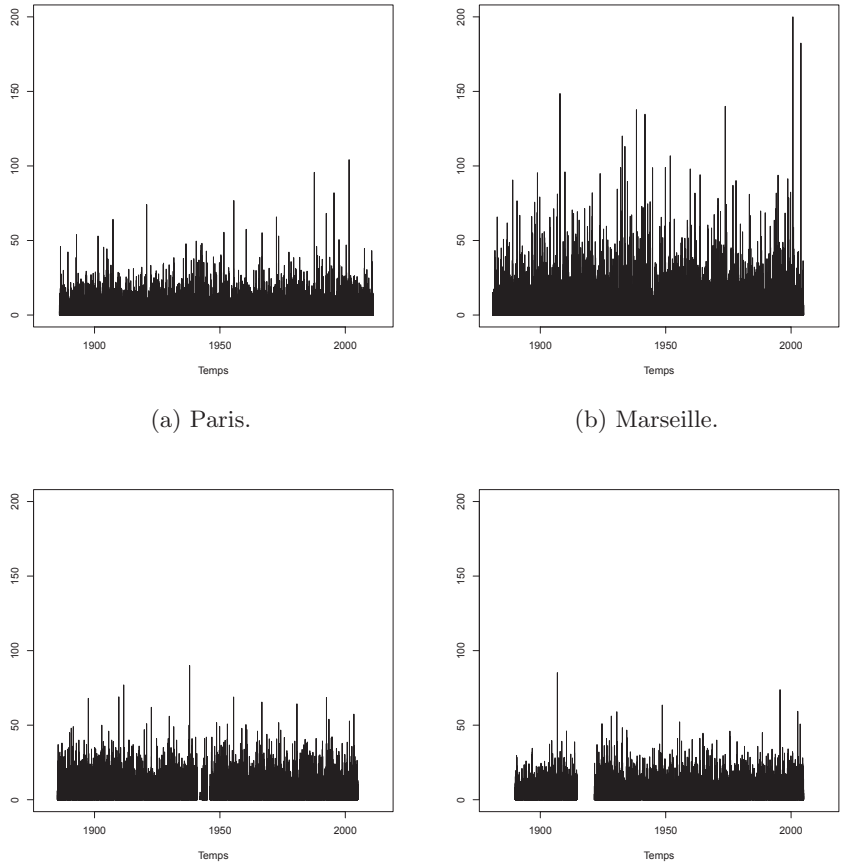
Les données affichent une bonne qualité avec au moins trente années d'observations journalières par station (équivalentes entre toutes les stations d'après Penot (2014)) et jusqu'à cinquante-sept années d'observations pour 380 d'entre elles. Les séries journalières possèdent de plus un faible taux de données manquantes (9% au maximum). Elles sont même complètes pour la grande majorité d'entre elles : 564 stations sur les 770.

4.3 Étude de stationnarité temporelle des précipitations extrêmes

Cette section s'intéresse à la stationnarité des extrêmes sur les données de précipitations issues des deux bases de données présentées dans la section 4.2. Il a été souligné dans la section 4.2.1 que certaines stations provenant du jeu de données publiques de ECA&D contiennent des enregistrements sur une importante période

de temps. C'est par exemple le cas de la station de Paris (Parc Montsouris), qui contient 128 ans de données (de 1886 à 2013 inclus).

La figure 4.4 représente les séries de précipitations pour les quatre séries contenant le plus de données journalières non manquantes : Paris, Marseille, Lège-Cap-Ferret (près d'Arcachon dans le département de la Gironde) et Vouziers (Ardennes).



(a) Paris. (b) Marseille.

(c) Lège-Cap-Ferret. (d) Vouziers.

FIGURE 4.4 – Cumuls de précipitations journalières pour quatre stations météorologiques en France.

Comme indiqué dans la section 4.2.1, on remarque une différence notable entre les pluies enregistrées à Marseille et celles des trois autres postes : les valeurs semblent généralement plus élevées, et par conséquent les valeurs extrêmes sont aussi plus importantes. Les données enregistrées à Vouziers et Lège-Cap-Ferret affichent une période de valeurs manquantes qui correspond aux années 1910 et aux années 1940 respectivement, c'est-à-dire pendant les deux guerres mondiales.

4.3.1 Tendance sur les valeurs extrêmes

Dans un contexte de changement climatique, il est attendu que le phénomène de précipitations évolue au cours du temps. Plus précisément, il semblerait (Alexander *et al.*, 2006) que les changements observés sur les précipitations aient lieu dans les événements extrêmes, que ce soit dans leur intensité ou leur fréquence, plutôt que dans les cumuls moyens.

Plusieurs études tentent d'évaluer ce changement dans les extrêmes, en regardant par exemple la présence de tendance dans les précipitations extrêmes (Klein Tank et Können, 2003; Van den Besselaar *et al.*, 2013; de Haan *et al.*, 2014). Sur des données de précipitations européennes, Van den Besselaar *et al.* (2013) constatent par exemple une augmentation en fréquence des événements extrêmes et une diminution (en médiane) de la période de retour évaluée variant entre 2% et 58%. Cependant, les résultats montrent une importante variabilité spatiale qui peuvent remettre en question la significativité des tests appliqués.

Indices ETCCDI

L'étude de la non-stationnarité des séries de précipitations journalières peut s'avérer difficile, en particulier à cause de la présence de nombreux 0 correspondant aux jours secs. Dans ce but, un ensemble d'indices liés aux extrêmes a été mis en place par l'ETCCDI (*Expert Team on Climate Change Detection and Indices*). Une liste non exhaustive d'indices liées aux précipitations extrêmes est donnée dans la table 4.2.

Indice	Description	Unité
RX1day	Maxima mensuels de précipitations cumulées sur une journée.	mm
RX5day	Maxima mensuels de précipitations cumulées sur cinq jours.	mm
SDII	Total annuel de précipitations par rapport au nombre de jours pluvieux (*Simple Daily Intensity Index*).	mm/jour
CWD	Durée maximale d'un épisode de jours pluvieux dans l'année (*Consecutive Wet Days*).	Jours
R10	Nombre annuel de jours de précipitations fortes (supérieures à 10mm).	Jours
R20	Nombre annuel de jours de précipitations très fortes (supérieures à 20mm).	Jours
R95p	Total annuel de précipitations dépassant le quantile d'ordre 95%.	mm
R99p	Total annuel de précipitations dépassant le quantile d'ordre 99%.	mm

TABLE 4.2 – Exemples d'indices ETCCDI liés aux précipitations extrêmes.

Pour évaluer l'impact du changement climatique dans les extrêmes, plusieurs tests sont disponibles en fonction du type de données dont on dispose. Renard (2006) liste plusieurs tests appliqués à des données hydrologiques :

1. le test de Mann-Kendall (Mann, 1945), construit pour la détection de tendances monotoniques,
2. le test de Pettitt (1979), similaire au premier mais adapté à la détection de points de changements abrupts dans la moyenne,
3. le test paramétrique sur une régression linéaire (les données doivent aussi vérifier la condition de normalité).

Cette section se concentre sur le test non-paramétrique de Mann-Kendall, qui est appliqué à des indices liés aux valeurs extrêmes des quatre séries de la figure 4.4.

Test sur la tendance : Mann-Kendall

Le test non-paramétrique de Mann-Kendall (Mann, 1945) porte sur la présence d'une tendance monotone dans une série temporelle. Ce test est souvent utilisé pour l'analyse de données environnementales (Hipel et McLeod, 1994; Renard, 2006)

Le test de Mann-Kendall permet de choisir entre

$$(H_0) : \ll \text{La série ne présente pas de tendance} \gg \quad \text{et}$$
$$(H_1) : \ll \text{La série a une tendance monotone} \gg \ .$$

Il est construit de la façon suivante sur l'échantillon $X = \{X_1, \ldots, X_n\}$:

1. La statistique de test de Mann-Kendall est :

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\mathbb{1}_{\{X_j > X_i\}} - \mathbb{1}_{\{X_j < X_i\}} \right) .$$

2. Sous l'hypothèse nulle (H_0) :

$$\mathbb{E}[S] = 0 \ \text{ et } \ \mathrm{Var}(S) = \frac{n(n-1)(2n+5)}{18} ,$$

3. On calcule :

$$Z = \begin{cases} \frac{S-1}{\sqrt{\mathrm{Var}(S)}} & \text{si} \quad S > 0 , \\ 0 & \text{si} \quad S = 0 , \\ \frac{S+1}{\sqrt{\mathrm{Var}(S)}} & \text{si} \quad S < 0 . \end{cases}$$

4. Sous l'hypothèse nulle (H_0), Z est asymptotiquement distribué selon la loi $\mathcal{N}(0, 1)$.

Ce test est disponible sous plusieurs formes dans R : la fonction `MannKendall` du package `Kendall` (McLeod, 2015) ou la fonction `mk.test` du package `trend` (Pohlert, 2016) en sont deux exemples. La première méthode sera préférée car elle permet de prendre en compte une série temporelle avec des valeurs manquantes.

Ce test suppose que les données $(X_1, \ldots, X_n)$ sont indépendantes. Pour vérifier cette assertion lors de son application, on utilise au préalable le test de Wald-Wolfowitz, comme indiqué par Sneyers (1984).

Test sur la corrélation : Wald-Wolfowitz

Le test non paramétrique de Wald et Wolfowitz (1943) porte sur la présence ou non de corrélation dans une série et est construit avec les hypothèses

$$(H_0) : \ll \text{La série ne présente pas de corrélation} \gg \quad \text{et}$$
$$(H_1) : \ll \text{La série présente une corrélation temporelle} \gg \ .$$

1. On recentre la série par $\tilde{X} := (\tilde{X}_1, \ldots, \tilde{X}_n)$, avec $\tilde{X}_i = X_i - \bar{X}$ où $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ est la moyenne empirique.
2. On pose $\tilde{X}_{n+1} = \tilde{X}_1$.
3. La statistique de test est :
$$R := \frac{\sum_{i=1}^n \tilde{X}_i \tilde{X}_{i+1}}{\sum_{i=1}^n \tilde{X}_i^2} \ .$$
4. Sous l'hypothèse (H_0), R est asymptotiquement distribué selon la loi normale :
$$\mathcal{N}\left(-\frac{1}{n-1}, \frac{1}{n-1}\right) \ .$$

La fonction `ww.test` du package R `trend` (Pohlert, 2016) permet de fournir une p-valeur associée au test de Wald-Wolfowitz.

Application sur des indices de valeurs extrêmes

Afin de tester l'évolution des valeurs extrêmes des précipitations, on utilise le test de Wald-Wolfowitz puis le test de Mann-Kendall sur plusieurs séries de valeurs extrêmes, parmi lesquelles des indices ETCCDI :
- la série des maxima annuels de données journalières,
- les indices RX1day et RX5day, qui correspondent à des maxima mensuels,
- les indices R10 et R20 qui regardent le nombre de valeurs extrêmes par an,
- les indices R95p et R99p qui évaluent l'amplitude des valeurs extrêmes de précipitations.

Les p-valeurs des deux tests sont données dans la table 4.3 pour les séries de cumuls de précipitations enregistrées à Paris, Marseille, Lège-Cap-Ferret et Vouziers.

Lorsque le test de Wald-Wolfowitz rejette l'hypothèse nulle (H_0) avec un risque α de 0.05, le test de Mann-Kendall devient douteux puisque la série d'indice n'est pas considérée comme indépendante. Dans ce cas, le résultat est indiqué en rouge dans la table 4.3.

	Test	Paris	Marseille	Lège-Cap-Ferret	Vouziers
Maxima annuels	W-W	0.4511	0.688	0.0422	0.5691
	M-K	0.0858	0.9731	0.8464	0.3839
RX1day	W-W	0.8292	0.0004	0.0003	0.0009
	M-K	0.7358	0.7241	0.5869	0.4934
RX5day	W-W	0.9274	< 0.0001	< 0.0001	0.0306
	M-K	0.0858	0.7859	0.198	0.0877
R10	W-W	0.8337	0.5488	0.56	0.4005
	M-K	0.2183	0.8101	0.2641	0.7442
R20	W-W	0.6529	0.2027	0.7083	0.6701
	M-K	0.3334	0.9157	0.4344	0.9064
R95p	W-W	0.5083	0.4752	0.3686	0.3914
	M-K	0.0943	0.9982	0.5237	0.8524
R99p	W-W	0.9216	0.4396	0.0598	0.8457
	M-K	0.737	0.9161	0.6950	0.4460

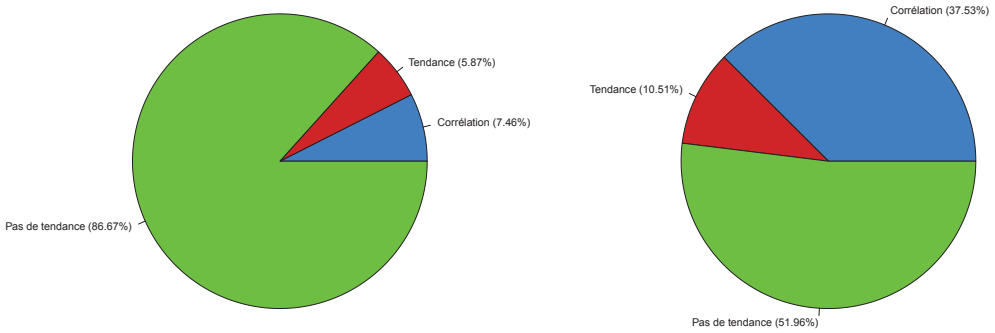
TABLE 4.3 – Résultats des tests de Wald-Wolfowitz (W-W) et de Mann-Kendall (M-K) pour quatre stations météorologiques et pour plusieurs indices de valeurs extrêmes dont certains issus de la base ETCCDI.

Sur ces quatre exemples, on voit que le test de Wald-Wolfowitz indique une présence de corrélation dans les séries de maxima mensuels (sauf pour la série de Paris) et dans les maxima annuels de Lège-Cap-Ferret, ce qui dans ces cas empêche de conclure vis-à-vis du test de Mann-Kendall. Cependant, on voit aussi que les données ne permettent pas, avec un risque $\alpha = 0.05$, de rejeter l'hypothèse nulle (H_0) du test de Mann-Kendall.

Cela signifie que pour les quatre séries testées, les précipitations journalières ne semblent ni connaître de valeurs extrêmes plus fréquentes (voir indices R10 et R20), ni voir leurs séries de maxima par blocs (annuels ou mensuels) ou d'excès de seuil (indices R95p et R99p) affectées par une tendance.

Les deux tests sont appliqués à l'ensemble des stations provenant à la fois de la base de données ECA&D et du jeu de données EDF-MFR (Penot, 2014), en se concentrant sur les maxima annuels et mensuels, qui semblent parfois présenter une corrélation d'après le test de Wald-Wolfowitz réalisé sur la table 4.3. Les résultats sont donnés sur les diagrammes de la figure 4.5. Les proportions de stations sont représentées :

- en bleu lorsque le test non-paramétrique de Wald-Wolfowitz rejette au niveau 95% l'hypothèse nulle, indiquant la présence de corrélation,
- en vert lorsque le test non-paramétrique de Mann-Kendall ne rejette pas l'hypothèse nulle au niveau 95%, indiquant l'absence d'une tendance monotone dans la série de maxima,
- en rouge lorsque le test non-paramétrique de Mann-Kendall rejette l'hypothèse nulle (en rouge), indiquant donc la présence d'une tendance monotone.



(a) Maxima annuels.

(b) Maxima mensuels.

FIGURE 4.5 – Résultats des tests de Wald-Wolfowitz et de Mann-Kendall sur les maxima annuels et mensuels pour l'ensemble des séries de précipitations journalières.

D'après ces graphes, on conclut que les maxima mensuels de cumuls de précipitations journalières sont généralement plus souvent corrélés que les maxima annuels : la proportion de stations météorologiques rejetant l'hypothèse nulle du test de Wald-Wolfowitz est plus importante. La présence de tendance est plus marquée également pour les séries de maxima mensuels, même si plus de la moitié des stations semblent accepter l'hypothèse nulle du test de Mann-Kendall.

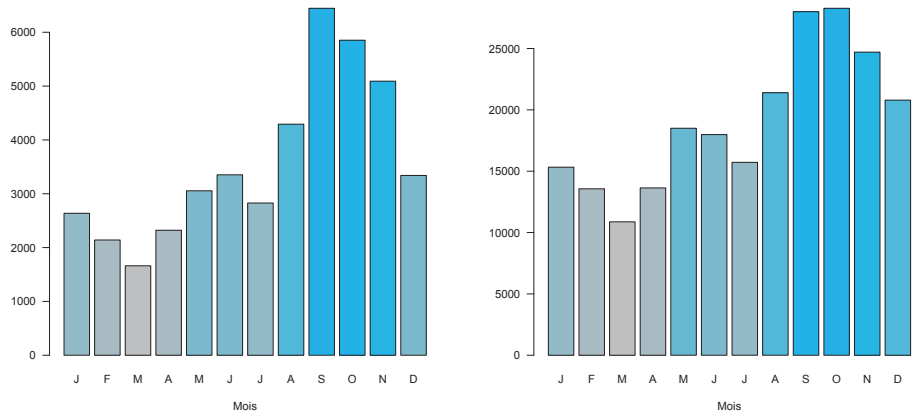
Les séries de maxima annuels sont utilisées pour modéliser les valeurs extrêmes de précipitations dans la partie II de la thèse. Si en général (environ 85 %), le test de Wald-Wolfowitz et le test de Mann-Kendall semblent confirmer que ces observations annuelles ne sont ni corrélées ni affectées par une tendance, il subsiste néanmoins des stations sur lesquelles le test n'est pas concluant.
Dans ce cas, il convient d'agir avec précaution lors de l'ajustement des modèles spatiaux.

4.3.2 Saisonnalité des occurrences d'extrêmes

Les cumuls de précipitations journalières en France sont sujets à des variations au cours de l'année : il pleut généralement moins souvent en été qu'en hiver, mais les orages y sont plus violents à cause des phénomènes convectifs discutés dans la section 4.1.

La présente section s'intéresse donc au phénomène de saisonnalité annuelle concernant les précipitations extrêmes. La figure 4.6a illustre cette saisonnalité en affichant le nombre de maxima annuels enregistrés sur les stations des deux jeux de données de précipitations par mois de l'année. Pour être comptabilisé, il est décidé qu'un maximum annuel doit être calculé sur une année contenant moins de 30 jours de données manquantes. La figure 4.6b complète cette illustration en comptant pour chaque mois de l'année le nombre de dépassements de seuils observés sur toutes les stations des jeux de données ECA&D et EDF-MFR réunies. Pour chaque série, le seuil utilisé est le quantile à 95% de la série des jours *pluvieux*.

La saisonnalité sur les valeurs extrêmes de précipitations se décrit par des occurrences d'excès et de maxima annuels plus nombreuses sur les mois d'automne (septembre, octobre et novembre), que sur les mois de printemps par exemple.



(a) Maxima annuels. (b) Excès de seuil.

FIGURE 4.6 – Occurrences du maximum annuel (a) et des excès de seuil (b) de précipitations journalières, pour toutes les stations des deux jeux de données, réparties par mois de l'année.

4.3.3 Saisonnalité des paramètres GEV

La saisonnalité des valeurs extrêmes est illustrée sur la figure 4.6 en regardant le nombre d'occurrences d'excès ainsi que la position des maxima annuels par mois de l'année. On s'intéresse ici à l'évolution annuelle des paramètres μ, σ et ξ de la loi GEV ajustée sur les maxima mensuels.

Modèle GEV périodique

Soit $\{X_t\}_{t\geqslant 1}$ une série de précipitations journalières, on se concentre sur la série associée $\{Y_{t_m}\}_{t_m\geqslant 1}$ des maxima mensuels. Rust *et al.* (2009) modélisent la série $\{Y_{t_m}\}_{t_m\geqslant 1}$ par une loi GEV non stationnaire *cyclique*, où les paramètres de localisation μ et d'échelle σ dépendent du mois de l'année avec des covariables sinusoïdales. Plus particulièrement, les paramètres de la loi GEV non stationnaire sont modélisés par :

$$\left\{\begin{array}{rcl} \mu_i & = & \beta_{\mu_0} + \beta_{\mu_1}\cos\big(2\pi i/12\big) + \beta_{\mu_2}\sin\big(2\pi i/12\big)\ , \\ \sigma_i & = & \beta_{\sigma_0} + \beta_{\sigma_1}\cos\big(2\pi i/12\big) + \beta_{\sigma_2}\sin\big(2\pi i/12\big)\ , \\ \xi_i & \equiv & \xi\ , \end{array}\right. \tag{4.1}$$

où $i \in \{1, \ldots, 12\}$ est l'indice du mois de l'année.

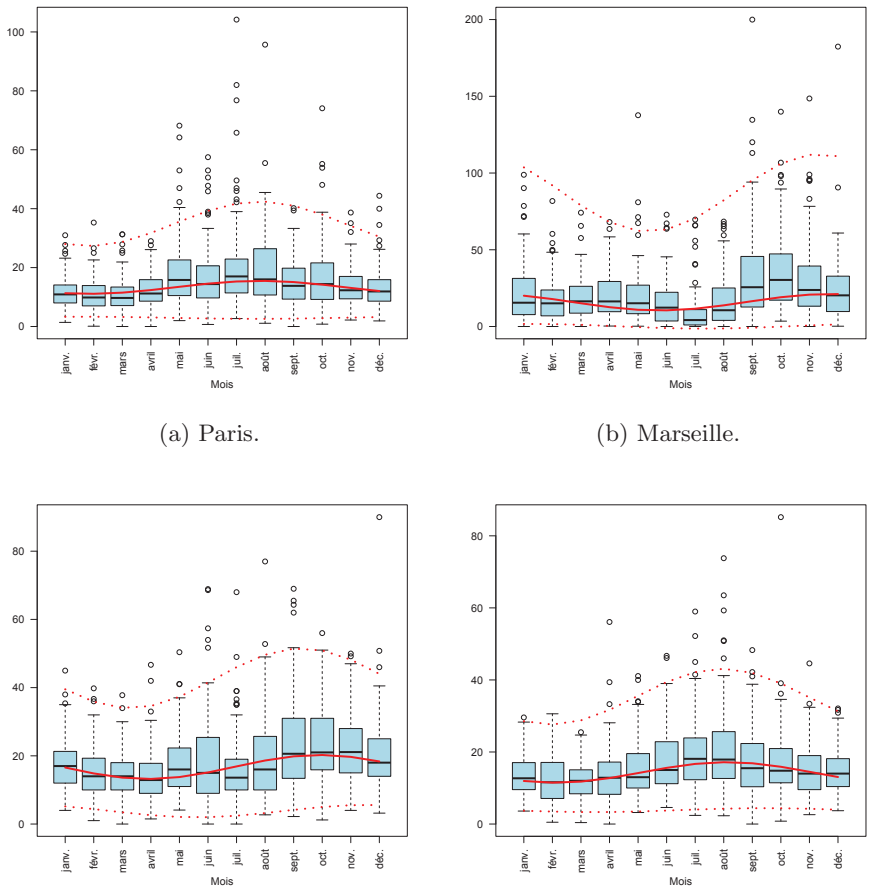
On remarque que le paramètre de forme ξ est considéré invariant en fonction du mois de l'année. Rust *et al.* (2009) comparent le modèle 4.1 avec le modèle analogue faisant varier périodiquement ce paramètre de façon mensuelle et montrent que ce dernier est accepté pour un grand nombre de stations météorologiques (situées au Royaume-Uni). Rust *et al.* (2009) décident pourtant de ne pas utiliser ce modèle en argumentant que cette saisonnalité observée pour ξ peut provenir d'un manque de données trop important sur certaines positions.

Pour ajuster le modèle GEV périodique sur des données de maxima mensuels, il est possible d'utiliser la fonction `gev.fit` du package R `ismev` (Heffernan *et al.*, 2016), qui permet d'ajouter des covariables temporelles telles que définies dans (4.1) lors de l'estimation des paramètres μ, σ et ξ.

La méthode est donc illustrée en ajustant le modèle (4.1) sur les séries de maxima mensuels observés à Paris, Marseille, Lège-Cap-Ferret et Vouziers. La figure 4.7 affiche les boîtes à moustaches de la série des maxima $\{Y_{t_m}\}_{t_m\geqslant 1}$ regroupés par mois de l'année avec une courbe rouge pleine (resp. pointillée) qui correspond à la médiane (resp. l'intervalle de confiance à 95%) de la loi GEV ajustée sur chaque mois de l'année.

Avec le modèle périodique, on peut voir que l'amplitude des valeurs extrêmes (regardées à travers les maxima mensuels) varie aussi en fonction de l'année. Pour les quatre séries représentées sur la figure 4.7, il semble que les mois d'été et d'automne soient concernés par les valeurs maximales mensuelles les plus importantes. Les comportements diffèrent cependant d'une région à l'autre, comme on peut le voir entre les stations de Paris (Figure 4.7a) et Marseille (Figure 4.7b).

Rust *et al.* (2009) utilisent cette modélisation saisonnière des maxima mensuels pour produire des cartes de niveaux de retour au Royaume-Uni, conditionnellement au mois de l'année. Un modèle spatial est donc considéré en plus de l'aspect temporel pour produire ces cartes (voir Section 6.5).



(a) Paris.

(b) Marseille.

(c) Lège-Cap-Ferret.

(d) Vouziers.

FIGURE 4.7 – Maxima mensuels de précipitations journalières regroupés par mois de l'année et ajustés par une loi GEV périodique.

Test de rapport de vraisemblance

Pour valider l'utilisation du modèle périodique, Rust *et al.* (2009) utilisent le test du rapport de vraisemblance pour deux modèles imbriqués.

Le modèle paramétrique M_0 est dit *imbriqué* dans le modèle M_1 s'il est un cas particulier de ce dernier. Autrement dit, M_0 s'obtient à partir de M_1 en fixant la valeur d'un ou plusieurs paramètres.

Le test du rapport de vraisemblance consiste alors à calculer la *déviance*, définie par :

$$D(y) = -2 \log \frac{f_0(y)}{f_1(y)} ,$$

où $f_0(y)$ et $f_1(y)$ sont respectivement les vraisemblances associées aux modèles paramétriques M_0 et M_1. Sous l'hypothèse (H_0) : « Le modèle M_0 est une simplification valide du modèle M_1 », la déviance $D(y)$ suit, sous certaines conditions, une loi de χ^2 avec $k_1 - k_0$ degrés de liberté où k_0 (resp. k_1) est le nombre de paramètres du modèle M_0 (resp. M_1), avec $k_0 < k_1$, par imbrication des modèles M_0 et M_1.

A titre d'information, le test du rapport de vraisemblance a été effectué pour les quatre stations utilisées sur la figure 4.7, et est chaque fois ressorti concluant avec une p-valeur largement inférieure à 0.05.

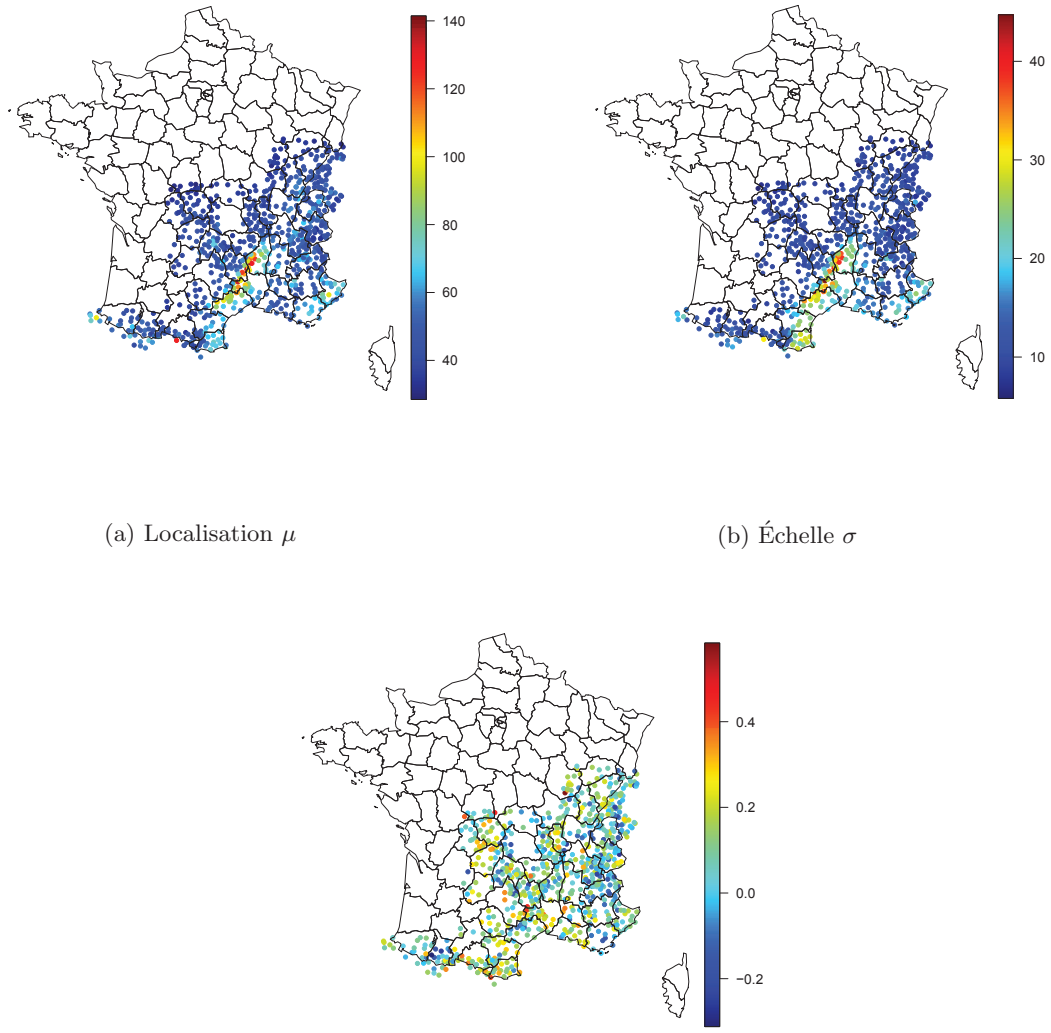
4.4 Analyse spatiale des précipitations extrêmes

Cette section évalue les différences d'événements extrêmes d'un point de vue spatial en étudiant les différences observées, entre plusieurs stations, sur les estimations des paramètres de la loi GEV associée aux maxima annuels. On discute ensuite le choix d'une région homogène qui servira à comparer les modèles spatiaux dans le chapitre 5.

Étant donné la densité du réseau de stations météorologiques et le faible nombre de données manquantes du second jeu de données, on utilise ce dernier pour regarder le comportement spatial des valeurs extrêmes.

4.4.1 Paramètres GEV site par site

On considère les maxima annuels des séries de précipitations journalières de la base de données EDF-MFR présentée dans la section 4.2.2. Les trois paramètres de la loi GEV sont estimés de façon ponctuelle, c'est-à-dire en procédant indépendamment station par station, et sont affichés sur la figure 4.8.



(a) Localisation μ

(b) Échelle σ

(c) Forme ξ

FIGURE 4.8 – Représentation spatiale des trois paramètres de la loi GEV ajustée sur les maxima annuels de précipitations journalières en procédant station par station.

Les cartes 4.8a et 4.8b montrent un motif spatial très clair pour les estimations ponctuelles des paramètres μ et σ, et a contrario une absence de forme claire pour le paramètre d'échelle ξ. La région des Cévennes est très marquée par rapport au reste de la région regardée, avec une valeur nettement plus élevée pour les paramètres d'échelle et de forme. Les épisodes de pluies extrêmes dans cette région constituent un sujet d'étude particulier (voir par exemple Lang *et al.* (2002) et Gardes et Girard (2010)). Leur formation est due à un phénomène météorologique particulier en automne : une masse nuageuse importante provenant de la mer Méditerranée vient heurter le massif des Cévennes, ce qui donne alors lieu à des orages très violents et à des inondations importantes dans la région, comme l'épisode du 9 au 13 octobre 2014 illustré sur la figure 1 en introduction de ce manuscrit.

Lors de la modélisation spatiale des paramètres GEV (voir partie II de la thèse), il convient d'après ces observations de définir un modèle pour les paramètres μ et σ et de fixer le paramètre d'échelle ξ comme identique pour toutes les stations. En effet, en plus de ne pas montrer une dépendance claire envers les covariables spatiales usuelles (longitude, latitude et altitude), ce paramètre est assez difficile à estimer en pratique. C'est pourquoi il est commun de le fixer spatialement (Davison *et al.*, 2012).

On peut aussi noter que les estimations du paramètre de forme ξ varient de façon relativement importante (de -0.307 à 0.579). Cela indique notamment que la loi des précipitations journalières, selon la position, peut varier entre une loi :

- bornée à droite lorsque $\xi < 0$ (domaine d'attraction de la loi Weibull inverse),
- à queue de distribution légère quand $\xi = 0$ (domaine d'attraction de la loi Gumbel),
- à queue lourde, dans le cas où $\xi > 0$ (domaine d'attraction de la loi Fréchet).

La boîte à moustaches de la figure 4.9 indique cependant que les estimations du paramètre ξ sont pour la plus grande partie comprises entre 0 et 0.2, indiquant une loi à queue lourde.

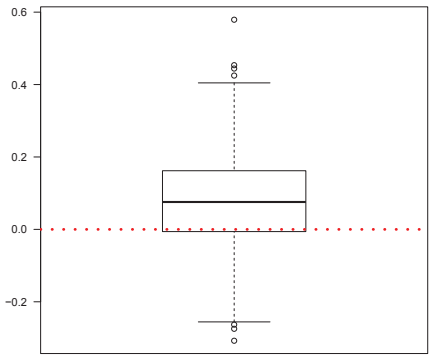


FIGURE 4.9 – Boîte à moustaches obtenue en rassemblant toutes les estimations entre les différentes stations du paramètre de forme ξ de la loi GEV ajustée sur les maxima annuels de précipitations journalières.

4.4.2 Choix d'une région homogène en vue de la modélisation spatiale

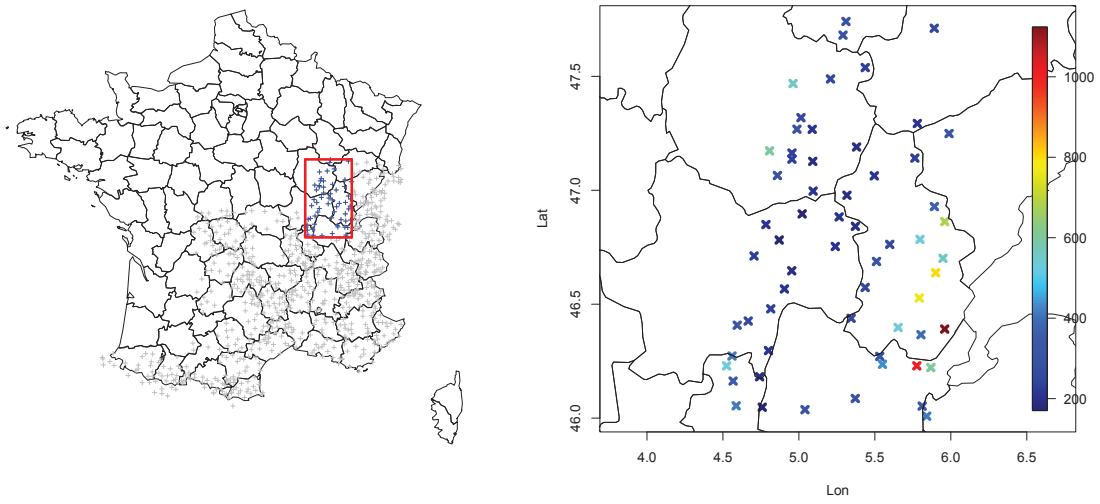
Afin d'ajuster des modèles spatiaux de valeurs extrêmes dans la partie II, on utilisera le jeu de données EDF-MFR qui possède une densité plus importante. Il a cependant aussi été décidé de se limiter à une partie de ce jeu de données et ce pour deux raisons :

1. D'abord, parce que la dépendance spatiale des valeurs extrêmes de précipitations est assez restreinte (de l'ordre de quelques dizaines de kilomètres environ, cf. Chapitre 5) et qu'il serait donc vain de relier deux stations physiquement trop éloignées.
2. Ensuite, parce que les modèles spatiaux utilisés peuvent devenir très coûteux en ressources de calcul s'ils sont ajustés sur une base de données de 770 stations.
3. Enfin, parce que l'on souhaite éviter de mettre en lien des phénomènes météorologiques qui ont des différences trop marquées, par exemple les maxima annuels de précipitations "typiques" avec des maxima enregistrés sur des épisodes cévenols.

Le choix s'est porté sur la région Centre-Est de la France, indiquée sur la figure 4.10a et composée de 61 stations. Les stations météorologiques qui y sont présentes sont placées dans une région peu élevée, relativement plane (voir Figure 4.10b) et suffisamment loin de la mer et des Cévennes pour être touchées par des précipitations extrêmes atypiques.

Les paramètres μ, σ et ξ de la loi GEV estimés à partir des observations de maxima annuels sont indiqués avec un code couleur sur les figures 4.11a, 4.11b et 4.11c respectivement.

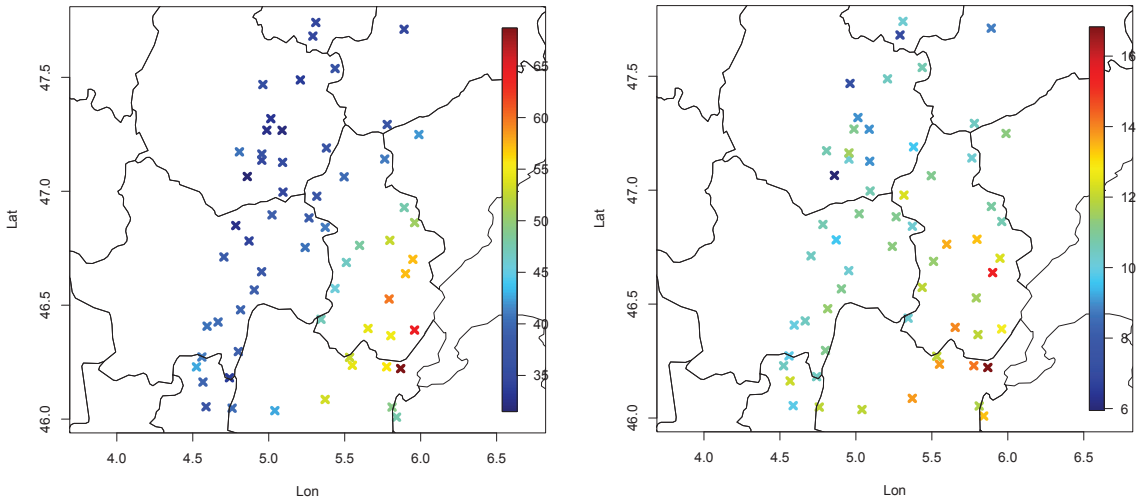
Les conclusions sur ces estimations faites station par station de manière indépendante sont similaires à ce qui a été dit en regardant l'ensemble de la base de données de la section 4.2.2. La région des Cévennes n'étant plus prise en compte, les variations d'estimation entre les positions sont plus faibles que sur les figures 4.8a et 4.8b pour les paramètres de localisation et d'échelle, mais on discerne encore un motif spatial avec une augmentation des valeurs pour ces deux paramètres dans les directions Nord-Sud et Ouest-Est.



(a) Région sélectionnée

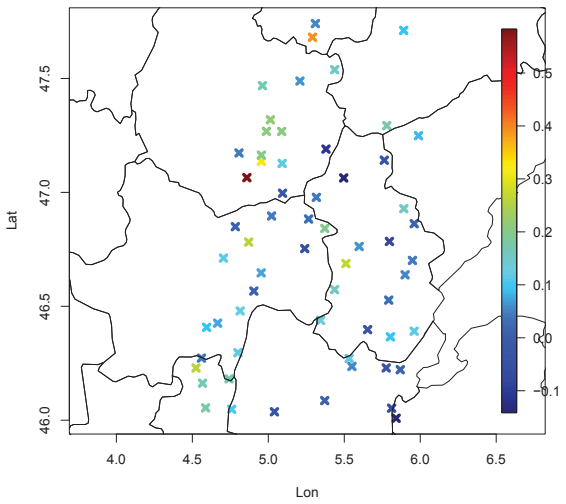
(b) Altitudes

FIGURE 4.10 – Positions des 61 stations météorologiques de la région sélectionnée parmi la base de données Météo France et altitudes correspondantes.



(a) Localisation μ

(b) Échelle σ



(c) Forme ξ

FIGURE 4.11 – Représentation spatiale des trois paramètres de la loi GEV ajustée sur les maxima annuels de précipitations journalières en procédant station par station sur la région Centre-Est sélectionnée.

Deuxième partie

Modélisation spatiale des valeurs maximales annuelles

Introduction de la partie II

Modélisation spatiale des valeurs extrêmes

Pourquoi utiliser des modèles spatiaux ?

Il a été vu dans le chapitre 4 que les données du cumul de pluies journalières possédaient une dimension spatiale. Cette observation motive le choix d'utiliser des modèles statistiques adaptés aux données spatiales pour décrire le phénomène extrême sous-jacent.

Parmi les objectifs fixés, rappelons que notre intérêt porte en partie sur deux problématiques :

1. la modélisation de la dépendance du phénomène de précipitations entre plusieurs stations météorologiques,
2. l'extrapolation d'une mesure de risque, comme par exemple le niveau de retour centennal, en une position où l'on ne dispose pas de données (*non jaugée*).

Si le premier problème fait clairement intervenir les résultats de la théorie des valeurs extrêmes multivariées (voir Section 1.2), il est important de définir la forme de la dépendance entre les stations en fonction de la distance spatiale qui les sépare. Ceci permet entre autres de mesurer un risque sur une position non jaugée sachant ce que l'on observe sur les stations météorologiques voisines.

Un modèle spatial pour décrire les valeurs extrêmes d'un phénomène consiste à faire des suppositions sur la structure qui dirige les lois marginales et les lois fini-dimensionnelles. Un ensemble de formules régit alors ces paramètres en les forçant à se conduire selon le modèle spatial choisi par le statisticien.

Cette procédure implique en général une perte de qualité du point de vue de la modélisation marginale puisque l'ajustement des lois en certains points est influencé par le choix de la structure du modèle spatial. En revanche, l'intérêt principal de se servir de cette approche est la possibilité d'utiliser ensuite les estimations des paramètres modélisés spatialement pour fournir une prédiction du phénomène observé. Extrapoler une mesure de risque devient donc possible grâce à ce type de modèle, ce qui permet de répondre aux deux points cités ci-dessus.

Modélisation spatiale des valeurs maximales

Cette partie de la thèse se concentre sur l'approche des maxima par blocs utilisée pour construire des modèles spatiaux adaptés aux valeurs extrêmes d'un processus $X = \{X(s)\}_{s\in\mathcal{S}}$, défini sur une région spatiale $\mathcal{S} \subset \mathbb{R}^p$. Dans notre cas, X est le processus de cumuls de précipitations observés à pas de temps journalier et $p = 2$ car les positions sont définies par les coordonnées géographiques : la longitude et la latitude par exemple.

Dans la littérature associée à la théorie des valeurs extrêmes, il existe une famille de modèles à caractère spatial qui permettent de décrire les maxima par blocs des processus stochastiques spatiaux : les *processus max-stables*. Ces modèles correspondent en fait à l'extension en dimension infinie des vecteurs multivariés de loi MEV.

Une classe particulière de ces approches spatiales est aussi investiguée dans cette partie : les modèles hiérarchiques. Les techniques de la géostatitstique s'appliquent difficilement sur les processus max-stables, car la forme de la structure de dépendance est plus complexe que celle d'un processus gaussien par exemple, entièrement explicité par une fonction moyenne et une fonction de covariance paramétrées. Les modèles hiérarchiques sont un moyen possible de contourner le problème : en conditionnant le phénomène spatial à caractère extrême par des processus latents, il devient possible d'extrapoler les valeurs extrêmes en un site non jaugé.
En se servant du cadre d'inférence bayésienne (cf. Chapitre 3), on peut alors utiliser les outils classiques de la géostatistique tels que le krigeage (cf. Chapitre 2) pour réaliser une prédiction des valeurs extrêmes en une position non jaugée.

Plusieurs modélisations des extrêmes d'un processus spatial sont possibles. Une étude comparative est donc requise pour désigner la méthode la plus adaptée à la question posée. Une telle comparaison, portant sur cinq modèles spatiaux de valeurs extrêmes, a fait l'objet d'un article en cours d'évaluation et est présenté dans le chapitre 5.

Aucun modèle n'est parfait en tout point, mais selon l'objectif que le statisticien se fixe, telle ou telle méthode apparaîtra la plus adaptée au problème. C'est dans ce sens que l'étude comparative a été menée, en discutant les avantages et inconvénients de chacun des modèles tout en les évaluant sur des critères précis utilisés dans les application.

Plan de la Partie II

Compétition de modèles spatiaux

Une des contributions scientifiques apportées par cette thèse est une étude comparative menée sur des modèles spatiaux de valeurs extrêmes pour la plupart fréquemment employés dans la littérature. L'objectif de cette mise en compétition est de pouvoir proposer des solutions adaptées aux besoins des utilisateurs.

Dans un premier temps, cinq modèles de processus spatiaux sont présentés au début du chapitre 5. Chacun permet d'approcher le processus limite des valeurs maximales observées pour un processus spatial. Les maxima par blocs pour un processus $X = \{X(s)\}_{s\in\mathcal{S}}$ sont définis, de façon analogue au cas d'un vecteur aléatoire multivarié, en considérant le maximum des réalisations "point par point" du processus :

$$M_n(s) := \bigvee_{i=1}^{n} X_i(s) \ , \quad \text{pour tout } s \in \mathcal{S} \ .$$

Dans un second temps, la compétition est menée à travers des simulations provenant de chacun des cinq modèles spatiaux ajustés au préalable sur les données de précipitations présentées dans le chapitre 4. La raison de ce choix est double : en premier lieu parce que l'on souhaite travailler avec des simulations pour comparer les modèles sur des critères objectifs, et ensuite parce que l'on utilise des données générées par chaque modèle pour étudier la robustesse des estimations sans utiliser un cadre qui favoriserait une des cinq méthodes.

Les critères de comparaison utilisés sont directement liés aux deux objectifs de la thèse discutés dans la page précédente. Les conclusions tirées par cette étude sont rediscutées dans le chapitre de conclusions à la fin de cette partie du manuscrit.

Accent porté sur un modèle en particulier

Parmi les six modèles spatiaux définis, l'un d'entre eux a fait l'objet d'une étude plus poussée pendant la durée de la thèse. Il s'agit du processus hiérarchique de Reich et Shaby (2012), qui modélise à la fois les paramètres des lois marginales par des processus gaussiens afin d'optimiser la prédiction spatiale et prend en compte une structure de dépendance max-stable autre que l'indépendance.

Un second apport scientifique associé à cette thèse est l'analyse approfondie de ce modèle, ainsi que l'implémentation d'un package R : `hkevp` (Sebille, 2016), qui lui est entièrement dédié.

Le chapitre 6 présente les différentes caractéristiques de ce modèle en explicitant les résultats de Reich et Shaby (2012) pour démontrer certaines propriétés mathématiques relatives à ce processus particulier. Dans le même temps, on présente aussi la procédure d'inférence bayésienne de ce modèle, faisant intervenir un algorithme MCMC assez complexe. Les différents avantages et inconvénients liés à ce modèle hiérarchique sont également abordés.

Chapitre 5

A comparative study of spatial extreme value models

Modeling extreme rainfall
A comparative study of spatial extreme value models

Quentin Sebille, Anne-Laure Fougères, Cécile Mercadier

Université de Lyon, CNRS UMR 5208, Université Lyon 1, Institut Camille Jordan,
43 blvd du 11 novembre 1918, 69622 Villeurbanne, France.

Abstract

In this paper, focus is done on spatial models for extreme events and on their respective efficiency regarding the estimation of two risk measures: one extrapolating marginal distributions and one summarizing the spatial bivariate dependence of extremes. A wide comparison is performed on a simulation plan that has been specifically designed from a daily precipitation data set. The objective of this paper is twofold: firstly, pointing out the inherent properties of each model, and secondly, advising users on how to choose the model depending on the specific type of risk.

Keywords: Spatial modeling of extreme events, extreme value theory, max-stable processes, hierarchical models, spatial prediction, precipitation data.
2010 MSC: 60G70, 62G32, 62H11, 62P12

1. Introduction

Analyses of extreme values of environmental variables such as precipitation are of great importance since it involves human lives as well as considerable losses of money when a catastrophic event occurs. Recently for instance, between May 28th and June 7th 2016, extreme precipitation events affected a part of Europe including France. This huge and sudden amount of rainfall caused nineteen deaths and, for France only, one billion Euro in damages. Accurate risk measures for such extreme phenomena are therefore needed to prevent from this type of scenario.

The risk estimation of these events is challenging because they involve values that are beyond the range of the observations. For this purpose, adapted tools come from extreme value theory. See for instance de Haan and Ferreira (2006), Beirlant et al. (2004), Finkenstädt and Rootzén (2004) and Coles (2001). Since precipitation phenomena have a spatial feature and data are generally observed at several stations, the dedicated setting to handle this question is that of max-stable spatial models. Detailed and helpful reviews on these models are Cooley et al. (2012), Davison et al. (2012) or Ribatet (2013).

Five max-stable spatial models have been selected among the most popular and the most recent in the literature. This choice includes Bayesian and frequentist concepts, and goes from simple to more complex spatial dependence structure. Within this paper, the main goal is to answer: *Which of the five competing models yields the best spatial prediction for extreme behaviors of simulated processes mimicking precipitation data?* Addressing this question supposes in particular a careful consideration of the data sets involved, as well as a relevant choice of performance criteria.

Two comparative criteria adapted to extreme events prediction are evaluated. The estimation of rare events at a location where no data is available is handled first; this induces the capacity of spatial extrapolation of the extreme behavior when looking at marginal information only. A clear and well known way to summarize this marginal information into a concrete risk measure is the return level. Then a second and complementary criterion is the measure of extreme sets for the bivariate distribution at a pair of locations. This aims at capturing the spatial dependence structure of extremes. One could

also look at indices involving more than two dimensions, but most of the dependent models for extremes have explicit formulae only in the bivariate case, so it would induce heavier numerical calculations.

Several options can be chosen to define the terms of comparison. One possible option could be to start from an expert point of view and compare each model with an a priori value of the previous criteria. To depart from a subjective choice, an intensive simulation study has been preferred. A parametric bootstrap procedure is used to produce simulations as close as possible to a true phenomenon. More precisely, a real precipitation data set is considered over a central-east region of France on which each of the five max-stable spatial models is fitted. These fitted models are then fixed to play the role of extreme rainfall generators. The two comparative criteria are finally evaluated on each generated sample and compared to the corresponding (known) true value.

The remainder of this article is organized as follows. The theoretical background of extreme value theory is addressed in the following section, with an emphasis on the so-called block maxima approach and on the five max-stable spatial models that we consider. Section 3 describes the simulation study, the two criteria used to evaluate each study and the results. Conclusions are drawn in Section 4, where some recommendations are provided with respect to different purposes.

2. Notations and models

2.1. Definition of max-stable processes

Let $\mathcal{S}$ be a compact subset of $\mathbb{R}^d$ that represents the spatial region of interest, d being a positive integer. Consider a random process $Y(\cdot) = \{Y(s)\}_{s\in\mathcal{S}}$ defined over $\mathcal{S}$, with continuous sample paths. Write $Y_1(\cdot), \ldots, Y_T(\cdot)$ for independent copies of $Y(\cdot)$. The process Y is called *max-stable* if for each $T > 1$, there exist continuous functions $a_T(\cdot) > 0$ and $b_T(\cdot) \in \mathbb{R}$ such that:

$$\bigvee_{t=1}^{T} \frac{Y_t(\cdot) - b_T(\cdot)}{a_T(\cdot)} \stackrel{d}{=} Y(\cdot) \ ,$$

where $\stackrel{d}{=}$ denotes equality in distribution. Such max-stable processes arise as non degenerate limits for pointwise maxima of stochastic processes on $\mathcal{S}$, and have been introduced by de Haan (1984). In particular, for each $s \in \mathcal{S}$, the random variable $Y(s)$ follows a generalized extreme value distribution $\mathrm{GEV}(\mu(s), \sigma(s), \xi(s))$, where the location $\mu(s) \in \mathbb{R}$, scale $\sigma(s) > 0$ and shape $\xi(s) \in \mathbb{R}$ parameters are indexed over space $\mathcal{S}$. Recall that the cumulative distribution function (cdf) of a $\mathrm{GEV}(\mu, \sigma, \xi)$ random variable Y is:

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

where $z_+ = \max(0, z)$ for all $z \in \mathbb{R}$. Thanks to the one-to-one mapping:

$$Z(s) = \left[1 + \xi(s)\frac{Y(s) - \mu(s)}{\sigma(s)}\right]^{1/\xi(s)} \ , \tag{1}$$

one transforms $Y(\cdot)$ into a *simple max-stable process* $Z(\cdot)$, that is with unit Fréchet margins, corresponding to a GEV(1,1,1) process at each spatial location in $\mathcal{S}$.

The joint cdf of $Z(\cdot)$ at a set of sites $\{s_1, \ldots, s_n\} \subset \mathcal{S}$ is given by:

$$\Pr\left(Z(s_1) \leqslant z_1, \ldots, Z(s_n) \leqslant z_n\right) = \exp\left[-V_{s_1,\ldots,s_n}(z_1, \ldots, z_n)\right] \ , \tag{2}$$

in terms of an *exponent measure* $V_{s_1,\ldots,s_n}$, also denoted V when there is no ambiguity. This exponent measure contains the information about the spatial dependence of extremes (see e.g de Haan and Ferreira, 2006, Chapter 9). The two limit cases are the independence case, with $V(z_1, \ldots, z_n) = \sum_{i=1}^{n} z_i^{-1}$ and the total positive dependence case, with $V(z_1, \ldots, z_n) = \bigvee_{i=1}^{n} z_i^{-1}$. A common measure

of spatial dependence between the set of sites $\{s_1, \ldots, s_n\}$ is the extremal coefficient $\theta(s_1, \ldots, s_n)$ of Schlather and Tawn (2003) defined by:

$$\Pr\left(Z(s_1) \leqslant z, \ldots, Z(s_n) \leqslant z\right) = \Pr\left(Z(s_1) \leqslant z\right)^{\theta(s_1,\ldots,s_n)} \ .$$

One can interpret $\theta(s_1, \ldots, s_n) \in [1, n]$ as the equivalent number of components of $\{Z(s_1), \ldots, Z(s_n)\}$ that are independent. The extremal coefficient is linked to the exponent measure via $\theta(s_1, \ldots, s_n) = V_{s_1,\ldots,s_n}(1, \ldots, 1)$.

In practice, the description of a max-stable process is done in two parts. First, the marginal distribution is captured by the processes that represent the GEV parameters. Classically, inference on the marginals is carried out under a linear model involving covariates. Such details are postponed until later (as in Equation (4) for instance). Second, the spatial dependence of extremes is measured by $V_{s_1,\ldots,s_n}(\cdot)$, or summarized through $\theta(s_1, \ldots, s_n)$. Parametric models for V can be helpful and some examples are presented in the next section.

2.2. Spectral representation and parametric models

Max-stable processes can be described thanks to the following spectral representation, due to de Haan (1984). Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be a Poisson point process on $(0, \infty)$ with intensity measure $\zeta^{-2}d\zeta$ and consider independent copies $\{W_j(s), s \in \mathcal{S}\}_{j\in\mathbb{N}}$ of a stationary process $W(\cdot)$ assuming $\sup_{s\in\mathcal{S}} W(s) < \infty$ with $\mathbb{E}\left[W(s)_+\right] = 1$. Then the process $Z(\cdot)$ defined for each $s \in \mathcal{S}$ by

$$Z(s) = \max_{j\geqslant 1} \zeta_j W_j(s) \ , \tag{3}$$

is max-stable with unit Fréchet margins. Its joint cdf is expressed as:

$$\Pr\left(Z(s) \leqslant z(s),\ s \in \mathcal{S}\right) = \exp\left(-\mathbb{E}\left[\sup_{s\in\mathcal{S}} \frac{W(s)}{z(s)}\right]\right) \ .$$

Different choices for the so-called spectral processes $W_j(\cdot)$ in (3) lead to different max-stable models. Three such max-stable models are now presented. Along this subsection, the dependence structure between two given sites s_1 and s_2 will be expressed skipping the notation in terms of the s_i's and h will denote the Euclidean distance between s_1 and s_2.

2.2.1. The Schlather model: EGP

A first possible choice is to take $W(\cdot) = \sqrt{2\pi}\varepsilon(\cdot)_+$, in terms of a standard stationary Gaussian process $\varepsilon(\cdot)$ with correlation function $\rho(\cdot)$. The resulting max-stable process, defined by Schlather (2002), is called the *Extremal Gaussian Process* (EGP). The associated bivariate exponent measure is:

$$V_{\text{EGP}}(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2\frac{[\rho(h)+1]z_1 z_2}{(z_1+z_2)^2}}\right) \ .$$

If the $\varepsilon_j(\cdot)$'s are assumed isotropic, then so is the max-stable process $Z(\cdot)$. The bivariate extremal coefficient of the EGP can be shown to be

$$\theta_{\text{EGP}}(h) = 1 + \sqrt{\frac{1-\rho(h)}{2}} \ .$$

A shortcoming of the EGP is that it cannot account for asymptotic independence, even when the distance h between two sites increases to infinity. Due to $\rho(\cdot)$ being positive definite, the bivariate extremal coefficient, in dimension $d = 2$, does not span over the interval $[1, 2]$ but instead spans the interval of $[1, 1.838]$ (see Appendix A for details).

Different correlation functions $\rho(\cdot)$ can be chosen. We work with the powered exponential form $\rho(h) = (1-\eta)\exp\left[-(h/\lambda)^\nu\right]$, where $\eta \in [0, 1)$, $\nu \in (0, 2]$ and $\lambda > 0$ are respectively the nugget, the smoothness and the range parameters. These parameters need to be estimated when fitting the EGP, and we denote them by $\eta_{\text{EGP}}, \nu_{\text{EGP}}$ and λ_{EGP}, respectively.

2.2.2. The Brown-Resnick model: BRP

A second possibility is to take $W(\cdot) = \exp\left[\varepsilon(\cdot) - \gamma(\cdot)\right]$, where here $\varepsilon(\cdot)$ is a Gaussian process with stationary increments and semivariogram $\gamma(\cdot)$. Then, representation (3) leads to a process $Z(\cdot)$ called the *geometric Gaussian process*. Kabluchko et al. (2009) showed that choosing $\varepsilon(\cdot)$ as a fractional Brownian motion yields the *Brown-Resnick process* (BRP) of Brown and Resnick (1977). Note that the term Brown-Resnick process is sometimes abusively referring to the geometric Gaussian process. The associated bivariate exponent measure is

$$V_{\mathrm{BRP}}(z_1, z_2) = \frac{1}{z_1}\Phi\left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log(z_2/z_1)}{\sqrt{2\gamma(h)}}\right) + \frac{1}{z_2}\Phi\left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log(z_1/z_2)}{\sqrt{2\gamma(h)}}\right) ,$$

and its extremal coefficient is given by

$$\theta_{\mathrm{BRP}}(h) = 2\Phi\left(\sqrt{\frac{\gamma(h)}{2}}\right) .$$

When $\varepsilon(\cdot)$ is a fractional Brownian motion, the semivariogram $\gamma(\cdot)$ is given by $\gamma(h) = \left(\frac{h}{\lambda}\right)^{\nu}$, where $\lambda > 0$ and $\nu \in (0, 2]$ are respectively the range and smoothness parameters. These parameters must be estimated to describe the spatial structure, and they are denoted by λ_{BRP} and ν_{BRP}, respectively.

2.2.3. The extremal-t model: ETP

Opitz (2013) obtained the only possible max-stable limit for asymptotically dependent elliptical processes, namely *the extremal-t process*, denoted ETP here. Let $\delta > 0$, $m_\delta = \sqrt{\pi}2^{1-\delta/2}\Gamma\left(\frac{\delta+1}{2}\right)^{-1}$ and $\varepsilon(\cdot)$ be a standard stationary Gaussian process with correlation function $\rho(\cdot)$. As for the EGP, we work with the powered exponential form for $\rho(\cdot)$.
The spectral process of the extremal-t model is $W(\cdot) = m_\delta \varepsilon(\cdot)_+^\delta$ and the corresponding exponent measure V for a pair of sites $\{s_1, s_2\}$ is given by

$$V_{\mathrm{ETP}}(z_1, z_2) = \frac{1}{z_1}T_{\delta+1}\left(-\frac{\rho(h)}{\varphi_\delta(h)} + \frac{1}{\varphi_\delta(h)}\left(\frac{z_2}{z_1}\right)^{1/\delta}\right) + \frac{1}{z_2}T_{\delta+1}\left(-\frac{\rho(h)}{\varphi_\delta(h)} + \frac{1}{\varphi_\delta(h)}\left(\frac{z_1}{z_2}\right)^{1/\delta}\right) ,$$

where $T_{\delta+1}(\cdot)$ is the cdf of the Student distribution with $(\delta + 1)$ degrees of freedom and $\varphi_\delta(h) = \sqrt{(1 - \rho(h)^2)/(\delta + 1)}$. The extremal coefficient of the ETP is given by

$$\theta_{\mathrm{ETP}}(h) = 2T_{\delta+1}\left(\frac{1 - \rho(h)}{1 - \rho(h)^2}(\delta + 1)\right) .$$

Note that both the EGP of Schlather (2002) and the BRP of Kabluchko et al. (2009) are special cases of the ETP model. Indeed, the EGP is obtained straightforwardly when $\delta = 1$, while the ETP model converges towards the BRP when $\delta \to \infty$. The dependence parameters that need to be estimated are then: δ_{ETP}, the degree of freedom, and the parameters of the correlation function, namely $\eta_{\mathrm{ETP}}, \lambda_{\mathrm{ETP}}$ and ν_{ETP}, respectively.

2.3. Hierarchical modeling

Hierarchical models typically assume that the observed process $Y(\cdot)$ is spatially independent conditionally on unobserved latent processes or variables. The interested reader may find examples of such hierarchical spatial models in Banerjee et al. (2004) for instance. Models based on hierarchical approaches are usually defined within the Bayesian paradigm. The estimation is generally performed through a Markov Chain Monte Carlo (MCMC) algorithm which allows one to usually sample straightforwardly from the posterior distribution of the parameters given the data. Recent studies exploiting hierarchical modeling for extreme precipitation are for instance Apputhurai and Stephenson (2013) and Dyrrdal et al. (2015).

2.3.1. The latent variable model: LVM

Davison et al. (2012) introduced a simple hierarchical structure for spatial extremes, called *the latent variable model* (LVM). Note that the usual lack of clear spatial pattern for the shape $\xi(\cdot)$ when dealing with precipitation data, jointly with the difficulty of estimating this parameter lead in this paper to consider $\xi(\cdot) \equiv \xi_0$. The random variables $\{Y(s)\}_{s\in\mathcal{S}}$ are assumed to be independent conditionally on latent processes that describe the GEV parameters. More precisely,

$$Y(s)|\{\mu,\sigma,\xi\} \overset{\text{indep}}{\sim} \text{GEV}(\mu(s),\sigma(s),\xi(s)) \ ,$$

where

$$\left\{ \begin{array}{rcl} \mu(s) & = & \beta_\mu^T c(s) + \varepsilon_\mu(s) \ , \\ \sigma(s) & = & \beta_\sigma^T c(s) + \varepsilon_\sigma(s) \ , \\ \xi(s) & \equiv & \xi_0 \ , \end{array} \right. \tag{4}$$

where the random parts $\varepsilon_\cdot(\cdot)$ are assumed to be independent stationary zero-mean Gaussian processes and where $c(s)$ denotes covariates associated to each position $s \in \mathcal{S}$. As a consequence, the mean function of the latent location parameter process $\mu(\cdot)$ (resp. $\sigma(\cdot)$) is written as a linear combination of covariates with unknown vector of coefficients β_μ (resp. β_σ). The covariates we use in this paper are the constant, the longitude, latitude and altitude. Note that Davison et al. (2012) consider the exponential form $\rho_{\varepsilon_\cdot}(h) = \delta_\cdot \exp(-h/\lambda_\cdot)$ to model the correlation of the latent Gaussian processes. We follow this choice to be consistent with the choices made in Sections 2.2.1 and 2.2.3 when defining the EGP and ETP.

The assumption of conditional independence leads to consider a multivariate exponent measure $V_{\text{LVM}}(z_1,\ldots,z_n) = \sum_{i=1}^n {z_i}^{-1}$ and the bivariate extremal coefficient $\theta_{\text{LVM}} \equiv 2$ that does not depend on the positions of $s_1,\ldots,s_n$.

2.3.2. The Reich and Shaby model: HKEVP

From Davison et al. (2012), one knows that the LVM is particularly appealing when the estimation of the marginal distributions is of interest, as it focuses on modeling the GEV parameters. The main drawback is that the dependence structure of extremes is not considered since conditional independence is assumed.

The aim of this subsection is to present a model that describes both the marginal effect and the dependence structure within a Bayesian framework. The *Hierarchical Kernel Extreme Value Process* (HKEVP) has been introduced by Reich and Shaby (2012) and further developed in Shaby and Reich (2012) and Reich et al. (2014). It is defined as follows. Suppose that $Y(s) \sim \text{GEV}(\mu(s),\sigma(s),\xi(s))$ and model the margins as the LVM as described by (4). Again, exponential correlations are used for the latent Gaussian processes $\varepsilon_\mu(\cdot)$ and $\varepsilon_\sigma(\cdot)$ in order to be consistent with previous choices.

Consider now $Z(\cdot)$, the associated simple max-stable process. The HKEVP allows for spatial dependence combined with a nugget effect by assuming $Z(s) = U_\alpha(s)\vartheta_\alpha(s)$ for all $s \in \mathcal{S}$, where $\alpha \in (0,1]$ is a parameter that controls the magnitude of the nugget effect.

- $U_\alpha(\cdot)$ is a spatially-independent process with common marginal distribution $\text{GEV}(1,\alpha,\alpha)$ representing the nugget effect.
- $\vartheta_\alpha(s) = \left(\sum_{\ell=1}^L A_\ell \omega_\ell^{1/\alpha}(s)\right)^\alpha$ describes the spatial dependence structure, constructed with:
  - deterministic positive kernel functions $\{\omega_1(\cdot),\ldots,\omega_L(\cdot)\}$ satisfying $\sum_{\ell=1}^L \omega_\ell(s) = 1$, $\forall s \in \mathcal{S}$,
  - and associated iid random variables $\{A_1,\ldots,A_L\}$ following the positive stable distribution with characteristic exponent α, denoted $\text{PS}(\alpha)$.

The kernels $\omega(\cdot)$ used in Reich and Shaby (2012) are rescaled Gaussian densities. More precisely, let $\mathcal{V} := \{v_1,\ldots,v_L\} \subset \mathcal{S}$ be a set of knots, and

$$\omega_\ell(s) = \frac{K(s|v_\ell,\tau)}{\sum_{j=1}^L K(s|v_j,\tau)} \ , \quad \forall \ell \in \{1,\ldots,L\} \ ,$$

where τ is the bandwidth parameter and

$$K(s|v_\ell, \tau) = \frac{1}{2\pi\tau^2} \exp\left[-\frac{1}{2\tau^2}(s - v_\ell)^T(s - v_\ell)\right] .$$

The product of a fixed $\vartheta_\alpha(s)$ by a $\mathrm{GEV}(1, \alpha, \alpha)$ distributed random variable $U_\alpha(s)$ gives a GEV distributed random variable $Z(s)$ whose parameters are known. More explicitly, by conditioning on $\vartheta_\alpha(s)$ (accordingly, over the random variables $A_1, \ldots, A_L$), one gets:

$$Z(s) \mid \vartheta_\alpha(s) \sim \mathrm{GEV}\big(\vartheta_\alpha(s), \alpha\vartheta_\alpha(s), \alpha\big) ,$$

and using relation (1) between processes $Y(\cdot)$ and $Z(\cdot)$ leads to the following hierarchical formulation:

$$\begin{aligned}
Y(s)|\mu, \sigma, \xi, \alpha, \vartheta_\alpha &\overset{\text{indep}}{\sim} \mathrm{GEV}(\mu^*(\mathrm{s}), \sigma^*(\mathrm{s}), \xi^*(\mathrm{s})) , \\
\mu^*(s) &= \mu(s) + \frac{\sigma(s)}{\xi(s)}\left(\vartheta_\alpha(s)^{\xi(s)} - 1\right) , \\
\sigma^*(s) &= \alpha\sigma(s)\vartheta_\alpha(s)^{\xi(s)} , \\
\xi^*(s) &= \alpha\xi(s) , \\
\vartheta_\alpha(s) &= \left(\sum_{\ell=1}^{L} A_\ell \omega_\ell^{1/\alpha}(s)\right)^\alpha , \\
A_1, \ldots, A_L &\overset{iid}{\sim} \mathrm{PS}(\alpha) .
\end{aligned}$$

If the conditioning random variable A follows the positive stable distribution as stated, then this model has the nice feature that the multivariate distribution obtained is both conditionally *and* unconditionally max-stable, and has an explicit exponent measure for any set of sites $\{s_1, \ldots, s_n\}$ with logistic form (Stephenson, 2009):

$$V_{\mathrm{HKEVP}}(z_1, \ldots, z_n) = \sum_{\ell=1}^{L}\left[\sum_{i=1}^{n}\left(\frac{\omega_\ell(s_i)}{z_i}\right)^{1/\alpha}\right]^\alpha .$$

Therefore, the bivariate extremal coefficient of the HKEVP is:

$$\theta_{\mathrm{HKEVP}}(s_1, s_2) = \sum_{\ell=1}^{L}\left(\omega_\ell(s_1)^{1/\alpha} + \omega_\ell(s_2)^{1/\alpha}\right)^\alpha .$$

The parameter α plays a central role in the HKEVP: it controls the degree of spatial dependence in the extremes by tuning the magnitude of both processes $U_\alpha(\cdot)$ and $\vartheta_\alpha(\cdot)$. If α tends to 0, the nugget is low, with $U_\alpha(\cdot)$ approximately equal to 1 everywhere, so that $Z(\cdot)$ is very smooth. On the opposite, if $\alpha = 1$, the random variables $\{A_1, \ldots, A_L\}$ are degenerated to 1, and by using the normalized condition on the kernel functions, one gets therefore $\vartheta_\alpha(\cdot) \equiv 1$. The process $Z(\cdot)$ is then a full-nugget spatial process. Refer to (Reich and Shaby, 2012, Section 2.1) for additional comments.
When jointly α tends to 0 and L tends to infinity, the HKEVP converges to a pioneer max-stable model known as the Smith's model (Smith, 1990). The Smith's model has not been considered in the present paper because of its unrealistic oversmooth feature and a resulting lack of fit on real data. See for instance Davison et al. (2012) or Davison et al. (2013), where the Smith's model is compared with the EGP and the BRP, among others.

2.4. Inference procedures

2.4.1. Inference for max-stable processes

Let $y_t(s_i)$ be the t-th observation of $Y(s_i)$, for $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, n\}$, where T stands for the number of years of study, and let y be the set of all observations. If observations are assumed

to be temporally independent, the likelihood of y under a max-stable spatial model is

$$L(y; \psi_{\text{GEV}}, \psi_V) = \prod_{t=1}^{T} g_n\big(y_t(s_1), \dots, y_t(s_n); \psi_{\text{GEV}}, \psi_V\big) \ , \tag{5}$$

where g_n is the n-multivariate max-stable density function. The latter depends on the n-multivariate marginal parameters μ, σ, and ξ that are completely determined through a set of spatial parameters ψ_{GEV} using for instance formulation (6), and whose exponent function V is driven by parameters ψ_V.

The computation of $g_n\big(y_t(s_1), \dots, y_t(s_n); \psi_{\text{GEV}}, \psi_V\big)$, for $t \in \{1, \dots, T\}$ is obtained by the transformation of the margins via (1) and by differentiating (2) with respect to $(z_1, \dots, z_n)$. One gets:

$$\frac{\partial^n}{\partial z_1 \dots \partial z_n} V_{s_1,\dots,s_n}(z_1, \dots, z_n) = \sum_{\Pi \subset \mathcal{P}} \prod_{\pi \in \Pi} -V_\pi(z_1, \dots, z_n) \exp\left\{-V_{s_1,\dots,s_n}(z_1, \dots, z_n)\right\} \ ,$$

where $\mathcal{P}$ denotes the set of all partitions of indices $\{1, \dots, n\}$ and V_π stands for the partial derivatives of the exponent measure $V_{s_1,\dots,s_n}$ with respect to the elements of $\pi \in \Pi \subset \mathcal{P}$. Unfortunately, the cardinal of $\mathcal{P}$ equals the Bell's number and is greater than 10^5 when $n \geqslant 10$, making the full likelihood quickly untractable even when dealing with a reasonable amount of positions.

For this reason, inference on max-stable processes has been done by maximizing the *composite likelihood* (Lindsay, 1988), and more specifically the *pairwise likelihood*, which is the likelihood evaluated at all pairs of positions $\{s_i, s_j\}_{i,j \in \{1,\dots,n\}}$:

$$L_{\text{pairwise}}(y; \psi_{\text{GEV}}, \psi_V) := \prod_{t=1}^{T} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} g_2\big(y_t(s_i), y_t(s_j); \psi_{\text{GEV}}, \psi_V\big) \ .$$

In the latter formulation, $g_2\big(y_t(s_i), y_t(s_j); \psi_{\text{GEV}}, \psi_V\big)$ is the bivariate max-stable density function with spatial marginal parameters ψ_{GEV} and dependence parameters ψ_V.

Estimation procedures using maximization of the pairwise likelihood for max-stable processes (Padoan et al., 2010) have been implemented in the following two R packages: `SpatialExtremes` (Ribatet, 2015) and `RandomFields` (Schlather et al., 2016). In Section 3, we use the former one which allows fit and simulation at non-gridded locations.

2.4.2. Inference for hierarchical models

Hierarchical models such as the LVM and the HKEVP are generally well handled in a Bayesian setting, and inference is thus performed through a MCMC Metropolis-within-Gibbs procedure.

For the LVM, the posterior distribution of the GEV parameters is given by:

$$\pi(\mu, \sigma, \xi | y) \propto L(y; \mu, \sigma, \xi, \psi_{\text{GEV}}) \pi(\mu, \sigma, \xi | \psi_{\text{GEV}}) \pi(\psi_{\text{GEV}} | \dots) \ ,$$

where $\propto$ stands for *proportional to* and:

- $\psi_{\text{GEV}} = \{\beta_\mu, \beta_\sigma, \delta_\mu, \delta_\sigma, \lambda_\mu, \lambda_\sigma, \xi_0\}$ is the set of parameters related to the latent Gaussian processes,
- $L(y; \mu, \sigma, \xi, \psi_{\text{GEV}})$ is the likelihood (5) with independent spatial structure,
- $\pi(\mu, \sigma, \xi | \psi_{\text{GEV}})$ is the product of the Gaussian densities of the latent processes μ, σ and ξ,
- $\pi(\psi_{\text{GEV}} | \dots)$ denotes the prior densities associated to the parameters ψ_{GEV}. Hyperparameters of these prior distributions are symbolized by the dots.

For the HKEVP, the posterior distribution of parameters $(\mu, \sigma, \xi, \alpha, \tau)$ is given by:

$$\pi(\mu, \sigma, \xi, \alpha, \tau | y) \propto L(y | \mu, \sigma, \xi, A, \alpha, \tau, \psi_{\text{GEV}}) \pi(\mu, \sigma, \xi | \psi_{\text{GEV}}) \pi(A | \alpha) \pi(\psi_{\text{GEV}}, \alpha, \tau | \dots) \ ,$$

where :

- $L(y|\mu, \sigma, \xi, A, \alpha, \tau, \psi_{\text{GEV}})$ is the likelihood (5) with *independent* spatial structure, since conditioning over the process ϑ_α (according to A, α and τ) gives spatial independence,
- $\pi(\mu, \sigma, \xi|\psi_{\text{GEV}})$ is the product of the Gaussian densities of the latent processes μ, σ and ξ,
- $\pi(A|\alpha)$ is the positive stable density with characteristic exponent α of the conditioning random variable A,
- $\pi(\psi_{\text{GEV}}, \alpha, \tau| \dots)$ is the product of independent prior densities for the spatial marginal parameters ψ_{GEV} and for the two parameters α and τ of the exponent function V_{HKEVP}.

As far as we know, there was no package including the fitting procedure for the HKEVP. The only connected function is the routine `abba` included in the package `extRemes` of Gilleland and Katz (2011) and related to the recent paper of Stephenson et al. (2015). Therein, the model uses a CAR prior over a network of thousands of gridded locations and is therefore not designed to make prediction outside the observed set of sites. The authors of the HKEVP propose nonetheless an open code available on Reich's website[1]. Sebille (2016) recently published on CRAN the R package `hkevp` that contains in particular a routine called `hkevp.fit` which fits the HKEVP model. This function is widely inspired by Reich & Shaby's code, and the main changes are listed in the reference manual of `hkevp`.

The MCMC algorithm associated to the LVM is available in the R package `SpatialExtremes` under the routine `latent`, and in the `hkevp` library under the function `latent.fit`. The latter allows to set the shape parameter $\xi(s)$ as spatially-invariant. To keep a maximum of coherences between the five models that are compared in this paper, we work with the `hkevp::latent.fit` routine.
The two MCMC algorithms are applied for this paper with 30.000 iterations after a burn-in period of size 15.000, and with a thinning procedure of length 15 so that the resulting Markov chains are of length 1.000. Whenever pointwise estimation is needed, the median of the corresponding chain is taken.

3. Comparison of the spatial models

This section is threefold. Firstly, the design of the simulation study is outlined. Secondly, the spatial models are fitted on the resulting artificial simulated data, and the results are analyzed. Finally, additional characteristics of these models are discussed.

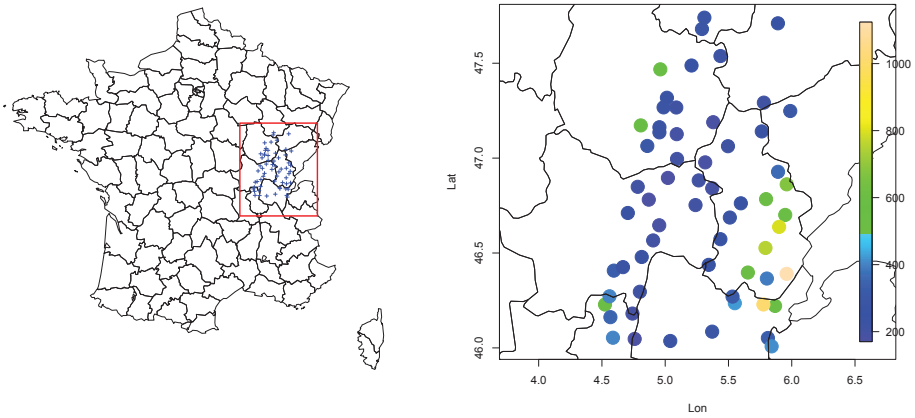
3.1. Designing the simulation as a precipitation data set

In this section, the five models described in Section 2, namely the LVM, HKEVP, EGP, BRP, and ETP are applied to a set of precipitation data recorded in France.
Along the paper, the models are sorted this way to respect an increasing "smoothness within dependence modeling", going from conditional independence to a spatial dependence structure with finite conditioning and then to three continuous max-stable dependence structures.

3.1.1. The precipitation data

The data set used in this article is extracted from the one analyzed by Penot (2014) and composed of $n = 61$ rain gauges in central-east of France, including 10 stations from EDF/DTG and 51 stations from Météo France. A map of France is given on Figure 1a with the positions of each meteorological station in blue and the region $\mathcal{S}$ in a red rectangle. Figure 1b indicates the elevation associated to each station with a color code. The latter shows that the considered region is relatively flat, with most of the positions located below 600m.

When available, annual maxima are taken at each station in the period between 1961 and 2005, which leads to $T = 45$ observations per site. For a given meteorological station, each annual maximum



(a) Region of interest $\mathcal{S}$. (b) Elevation of the meteorological stations.

Figure 1: Map of France with positions of the meteorological stations and the region $\mathcal{S}$ (a) and elevation of each station in $\mathcal{S}$ represented with a color code (b).

has been computed if there was less than 30 days missing in the corresponding year. Half the stations have complete recorded time series, and all of them have at least 30 non-missing annual maxima.

Pointwise estimations of the GEV distribution are performed at each site of the region $\mathcal{S}$. Figure 2 shows the estimated values of μ, σ and ξ on the map with a color code. From these maps, we can visually assess for spatial trends in the GEV parameters. A brief analysis shows that parameters μ and σ depend on longitude, latitude and altitude. The shape parameter ξ at each location has no clear visual pattern, and a regression analysis failed to significantly link the pointwise estimates with any of the three covariates available. This parameter is then set as spatially-invariant for all models.

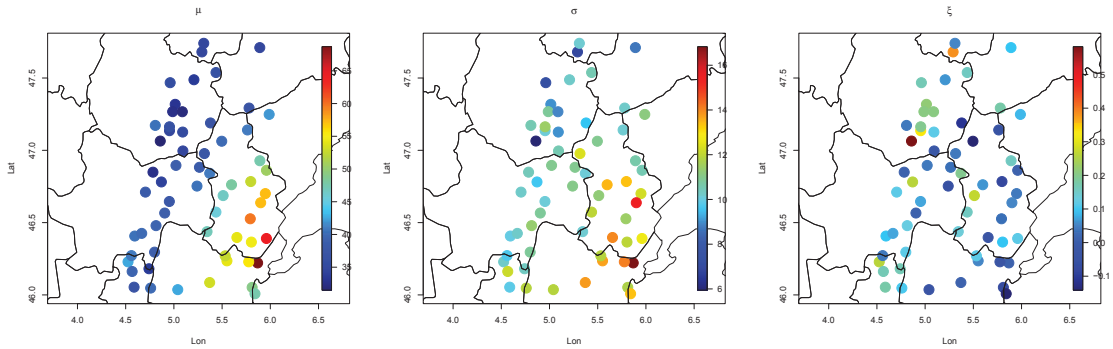


Figure 2: Pointwise estimates of GEV location (right), scale (middle) and shape (left) parameters.

3.1.2. Fitting the models on the data

The five spatial models defined in Section 2 are fitted on the precipitation dataset described in Section 3.1.1. Motivated by the spatial trend analysis of the GEV parameters, we use the following

[1]Homepage `http://www4.stat.ncsu.edu/~reich/`

marginal spatial model for the EGP, the BRP and the ETP:

$$\left\{ \begin{array}{rcl} \mu(s) & = & \beta_{\mu,0} + \mathrm{lon}(s)\beta_{\mu,1} + \mathrm{lat}(s)\beta_{\mu,2} + \mathrm{alt}(s)\beta_{\mu,3} \ , \\ \sigma(s) & = & \beta_{\sigma,0} + \mathrm{lon}(s)\beta_{\sigma,1} + \mathrm{lat}(s)\beta_{\sigma,2} + \mathrm{alt}(s)\beta_{\sigma,3} \ , \\ \xi(s) & = & \beta_{\xi,0} \ . \end{array} \right. \tag{6}$$

For the LVM and the HKEVP, the same linear combination is taken to which are added the latent zero-mean Gaussian processes $\varepsilon_\mu(\cdot)$ and $\varepsilon_\sigma(\cdot)$, as previously described in (4).

Figure 3 shows the GEV Quantile-Quantile plots (QQ-plots) obtained from the five spatial models at each of the 61 positions. A color has been assigned to each station in order to distinguish roughly the results between one place to another. First, it can be pointed out that the pointwise estimation

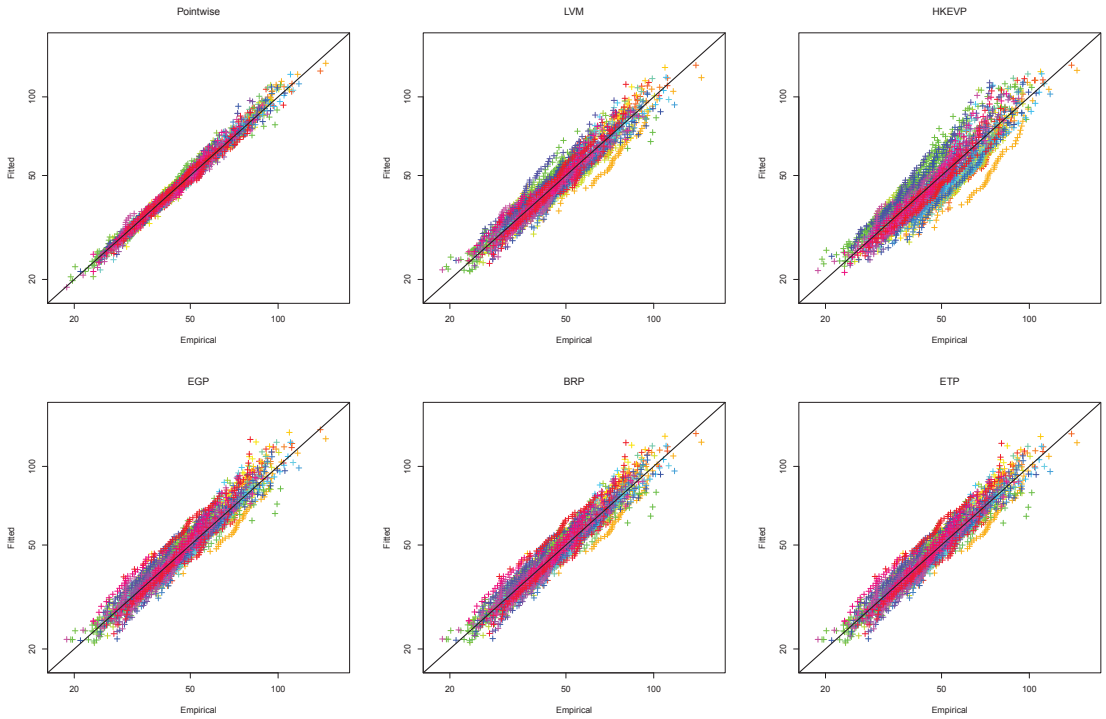


Figure 3: GEV Quantile-Quantile plots of observed data against fitted with pointwise GEV estimation (no spatial structure, top left figure) and each of the five spatial models. Colors are used to distinguish the meteorological stations.

(upper-left figure) is relevant for every meteorological station, that accredits the GEV assumption. The addition of a spatial structure in the GEV parameters through a spatial model leads to a minor spoiling of the fitting quality, but this is necessary if the goal is to perform spatial extrapolation. The QQ-plots obtained by the three max-stable models are roughly equivalent: this is not surprising since they share the same marginal spatial modeling (6). The use of a hierarchical formulation increases the flexibility of the marginal estimation. The QQ-plots obtained by the LVM and the HKEVP (top middle and top right) show however that the corresponding fits are less accurate, overall on the right part of the distribution. The estimated value of ξ_0 might be an explanation for the second hierarchichal model: it equals 0.20 for the HKEVP, while it is approximately equal to 0.09 for the four other ones. See discussions in Section 3.3.2.

The second step of the analysis consists in exploring the spatial bivariate dependence structure. This is done through the estimations of the extremal coefficient between two sites s_1 and s_2. Recall that if ν is the F-madogram of Z, that is:

$$\nu(s_1, s_2) = \mathbb{E}\big|F(Z(s_1)) - F(Z(s_2))\big| \ ,$$

then the extremal coefficient θ is given for all $s_1, s_2 \in \mathcal{S}$ by:

$$\theta(s_1, s_2) = \frac{1 + 2\nu(s_1, s_2)}{1 - 2\nu(s_1, s_2)} \ .$$

The non parametric estimator of $\theta(s_1, s_2)$ is thus naturally derived from the madogram-based estimator of Cooley et al. (2006).
Figure 4 plots $\{\hat{\theta}(s_i, s_j)\}_{1 \leqslant i < j \leqslant n}$, the empirical estimates of the bivariate extremal coefficient for all pairs of sites. The corresponding extremal coefficients of each spatial model fitted on the annual maxima of precipitation data are also given in the same figure, with different colors to dissociate them. For EGP, BRP and EGP, the isotropic assumption allows to depict the curves of $\hat{\theta}$ in terms of h, while the estimation of θ is drawn for all pairs of sites when fitted by the anisotropic HKEVP model (see several red points at each distance value abscissa). Note that the estimated posterior medians of α and τ were used as point estimates to compute the bivariate extremal coefficient for the HKEVP. The LVM has been omitted since the conditional independence assumption implies $\theta \equiv 2$.

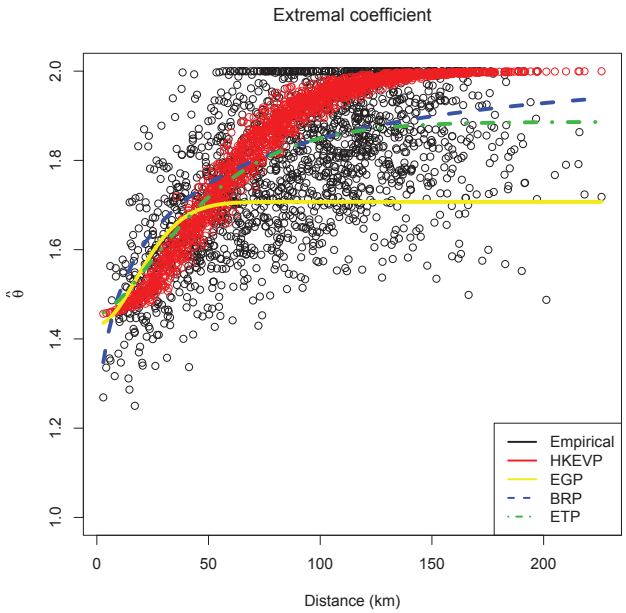


Figure 4: Empirical extremal coefficient $\hat{\theta}$ (in black) and fitted ones for each model $\hat{\theta}_{\text{HKEVP}}$, $\hat{\theta}_{\text{EGP}}$, $\hat{\theta}_{\text{BRP}}$ and $\hat{\theta}_{\text{ETP}}$ (in color).

One can deduce from Figure 4 that the four models with spatial dependence structure fit the data reasonably well. Two particular features can be commented on this figure. First, the EGP does not reach the limit value $\theta = 2$ (independence), even when the distance h becomes large. This model's drawback has been highlighted in Section 2.2.1. The limit value here is not 1.838 as stated previously but 1.707, since we consider a powered exponential correlation function $\rho(\cdot)$ that is always positive. Secondly, the impact of the nugget parameter can be seen on several models when h is close to zero: the bivariate extremal coefficient does not go to 1 as it should be but is lower-bounded. For example, as stated in Reich and Shaby (2012), the bivariate extremal coefficient of the HKEVP between a location s and itself is $\theta(s, s) = 2^{\alpha}$, so that it equals one only when the nugget α is fading to zero. The nugget parameter (α in HKEVP, η_{EGP} in EGP and η_{ETP} in ETP) models generally the error measurements as well as microscale variability at a given location, which motivates its use.
Extreme rainfall data are known to exhibit a low range of spatial dependence. For instance, Fawcett and Walshaw (2014) found out that the dependence between extreme precipitation in Great Britain is

genuine for a distance $h < 100$ km. In the present dataset, half of the inter-distances are lower than 100 km, and there is still some trace of extreme dependence between 100 km and 200 km.

3.2. Comparative study of the spatial models

Recall that the five models are fitted so that they can now be used as data generators. The goal of this technique, close to the parametric bootstrap, is to work with simulated data so that we can compare estimated criteria with the known true value. Since we do not want to favor one particular model by setting an arbitrary simulation design, we use the fitted parameters of the five spatial models (obtained on the same dataset) to generate artificial data.
Despite the loss of preciseness in the marginal and the bivariate modelling observed respectively in Figure 3 and 4, we keep all five models and their fitted versions to create artificial data generator. Indeed, the goal is to work with simulations that are close enough to precipitation to be relevant, but we do not want to perfectly mimic the phenomena, otherwise the generators we use in the competition will be equivalent. The five spatial models will compete through the simulation design presented below.

3.2.1. Simulated data sets

The aim of this section is to present the way to produce simulations of annual maxima of daily precipitation. Samples are drawn from the five models with parameters fixed from Section 3.1.2. Routines `rmaxstab` and `hkevp.rand` are used from packages `SpatialExtremes` and `hkevp` respectively. We denote by:

$$\mathbb{Y}_{\mathcal{G}}^{(k)} := \begin{bmatrix} Y_1^{(k)}(s_1|\mathcal{G}) & \dots & Y_1^{(k)}(s_n|\mathcal{G}) \\ \vdots & \ddots & \vdots \\ Y_{n_y}^{(k)}(s_1|\mathcal{G}) & \dots & Y_{n_y}^{(k)}(s_n|\mathcal{G}) \end{bmatrix}_{n_y \times n},$$

the k-th replicate, for $k \in \{1, \dots, K = 50\}$, coming from the generator model $\mathcal{G}$. Artificial data are generated on the $n = 61$ positions $\{s_1, \dots, s_n\}$ coinciding with the sites of Figure 1. The number of simulations $n_y = 45$ is the maximal length observed of yearly maxima series per station on this dataset. Equivalent datasets $\mathbb{Z}_{\mathcal{G}}^{(k)}$ are also used: they differ from $\mathbb{Y}_{\mathcal{G}}^{(k)}$ by being unit Fréchet distributed.

3.2.2. Two comparative criteria

The performances of the competing spatial models are measured on the two criteria that are described below.

The first criterion focuses on the quality of spatial extrapolation of the GEV marginal distribution. It consists in the estimation of the T-year return level at a "target" site s^* where no data is available. For any period of time T, this value is a quantile (denoted by $y_T(s^*)$) of the marginal distribution that we expect to be exceeded once over T years. It is therefore defined by $\Pr(Y(s^*) \leqslant y_T(s^*)) = 1 - 1/T$. Knowing the marginal parameters μ, σ and ξ evaluated at s^*, it is possible to explicitly compute its value:

$$y_T(s^*) = \mu(s^*) + \frac{\sigma(s^*)}{\xi(s^*)} \left[\log\left(\frac{T}{T-1}\right)^{-\xi(s^*)} - 1 \right] . \tag{7}$$

The linear model (6) allows direct extrapolation of marginal parameters at an ungauged site s^*, provided that all spatial covariates associated to s^* are known. Thanks to (7), the computation of the first criterion is straightforward for the EGP, the BRP and the ETP models.
For the hierarchical models LVM and HKEVP, the first criterion can be estimated in different ways. We choose the median of the predictive posterior distribution of $y_T(s^*)$, by using a kriging estimator to obtain $\mu(s^*), \sigma(s^*)$ and $\xi(s^*)$ at each MCMC step.

The second criterion we study provides a measure of the spatial bivariate dependence structure of a specific model. For this purpose, we use the bivariate extremal coefficient θ evaluated at a specific pair of sites (s_1, s_2). This fixed pair is chosen so that the associated distance h is approximately 50

km, a range where all spatial models display some spatial dependence when fitted on the data (see Figure 4).
Apart from summarizing the dependence structure, the bivariate extremal coefficient may be used to compute a joint probability. For example, the probability that one of the two annual maxima $Y(s_1)$ and $Y(s_2)$ be greater than their respective 100-years return levels $y_{100}(s_1)$ and $y_{100}(s_2)$ is given by:

$$\Pr\left\{Y(s_1) > y_{100}(s_1) \text{ or } Y(s_2) > y_{100}(s_2)\right\} = 1 - \exp\left(-\frac{\theta(s_1, s_2)}{z_{100}}\right) ,$$

where z_{100} is the 100-years return level of a unit Fréchet random variable.

3.2.3. Estimating the comparative criteria with the five models

The five spatial models are all refitted to each set of artificial data $\mathbb{Y}_{\mathcal{G}}$, and the two criteria defined in Section 3.2.2 are estimated and compared to the true value associated to $\mathcal{G}$. Let $\mathcal{F}$ denote a fitting model chosen among the five spatial models LVM, HKEVP, EGP, BRP, and ETP.

- As the first criterion involves a spatial extrapolation at an ungauged position, we use a κ-fold cross validation procedure. The 61 sites are randomly split into $\kappa = 6$ classes, represented in Figure 5 with different colors. The models $\mathcal{F}$ are fitted on five classes of sites and the sixth plays the role of the ungauged positions. We obtain estimations of $y_{100}(\mathbf{s}_\kappa; \mathcal{G})$, the 100-years return value of $Y(\cdot)$ generated by $\mathcal{G}$ at ungauged positions $\mathbf{s}_\kappa$ of the κ-th class of sites.
  We repeat this procedure for each class κ, and compute, for each generator $\mathcal{G}$ and each fitting model $\mathcal{F}$, the MSE:

$$\mathrm{MSE}(\mathcal{F}; \mathcal{G}) = \frac{1}{6}\frac{1}{|\mathbf{s}_\kappa|}\sum_{\kappa=1}^{6}\sum_{\mathbf{s}_\kappa}\left(y_{100}(\mathbf{s}_\kappa; \mathcal{G}) - \widehat{y_{100}}(\mathbf{s}_\kappa; \mathcal{F})\right)^2 .$$

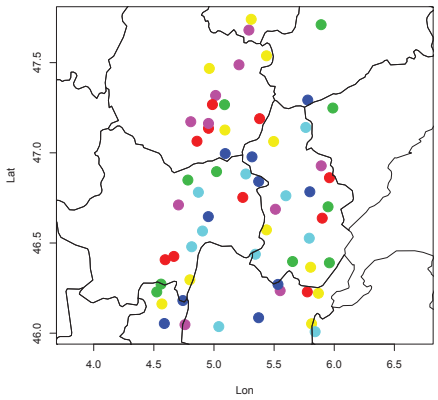
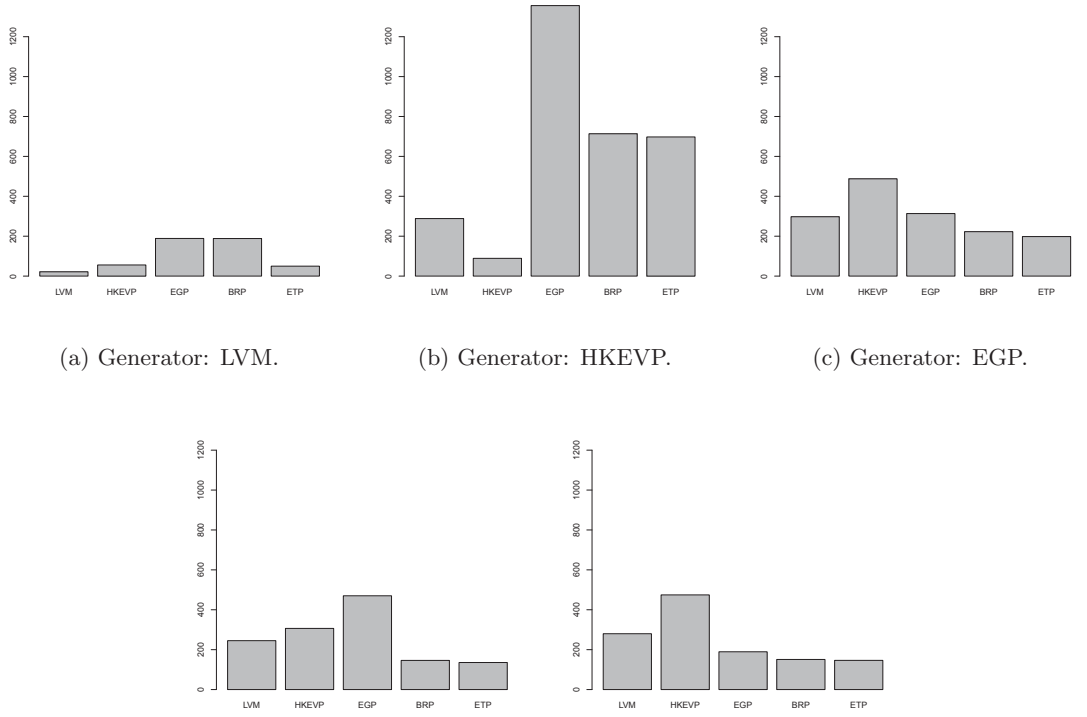


Figure 5: Classes of sites chosen randomly to perform the κ-fold cross validation on first criterion. Each color defines a different class.

- The second criterion, i.e. the bivariate extremal coefficient $\theta(s_1, s_2)$, is estimated by $\mathcal{F}$ on the artificial set $\mathbb{Z}_{\mathcal{G}}$, with unit Fréchet margins. The results obtained are samples of $K = 50$ estimations of $\hat{\theta}_{(s_1,s_2)}(\mathcal{F}; \mathcal{G})$.

3.2.4. Results

The comparative studies are summarized in Figures 6 and 7, which respectively show MSE($\mathcal{F};\mathcal{G}$) and the boxplots of $\hat{\theta}_{(s_1,s_2)}(\mathcal{F};\mathcal{G})$. In both figures, $\mathcal{G}$ is specified in the legend below each panel (from (a) to (e)) and $\mathcal{F}$ is given below each bar or boxplot, keeping always the same ordering from left to right, namely LVM, HKEVP, EGP, BRP, and ETP. In Figure 7, the dashed red lines correspond to the true value $\hat{\theta}_{(s_1,s_2)}(\mathcal{G})$.



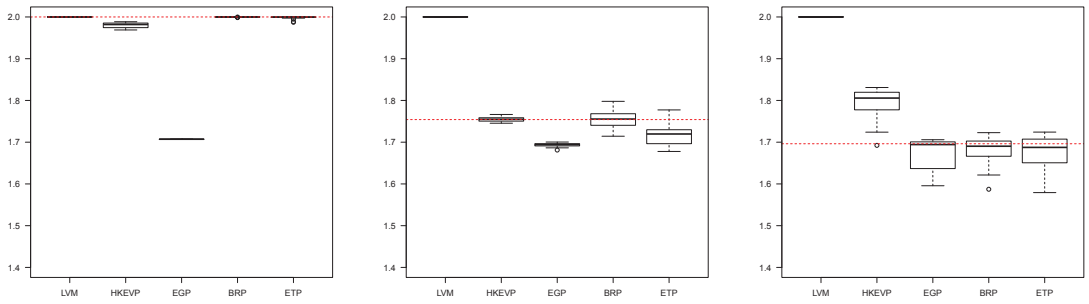
(a) Generator: LVM. (b) Generator: HKEVP. (c) Generator: EGP.

(d) Generator: BRP. (e) Generator: ETP.

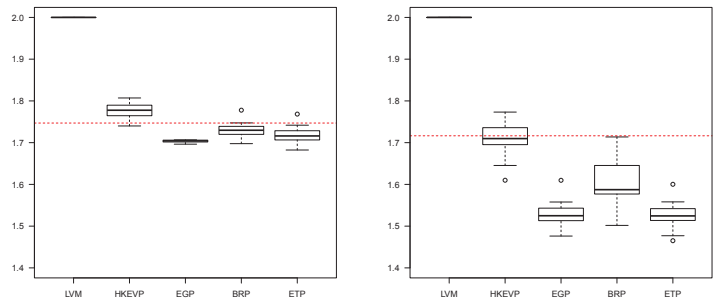
Figure 6: MSE($\mathcal{F};\mathcal{G}$) obtained on criterion 1 for all generators $\mathcal{G}$ and fitting models $\mathcal{F}$.

From these two figures, it is possible to roughly order the five spatial extreme value models by their performances over the two criteria defined in Section 3.2.2.

The two "best" competitors may be the BRP of Brown and Resnick (1977); Kabluchko et al. (2009) and the ETP of Opitz (2013). Both methods show similar satisfying results in terms of the two criteria of comparison, with an exception on Figure 6b when the generator is the HKEVP. In this case, the MSE obtained on the first criterion is quite large. On the one hand, the ETP outperforms slightly the BRP in terms of marginal extrapolation (see Figures 6a and 6c), while the BRP seems more accurate regarding the second criterion (see Figure 7e). In particular (and rather surprisingly), the ETP shows the biggest bias on the extremal coefficient when artificial data come from the ETP itself. Based on these few observed differences, the ETP could be recommended for the spatial extrapolation of the 100-years return level, while the BRP should be preferred for the estimation of the extremal coefficient. The HKEVP of Reich and Shaby (2012) is a good model if the goal is the estimation of the dependence structure, but it has some difficulties regarding the extrapolation of the return level. Indeed, in terms of the first criterion, the MSE obtained by fitting this model is generally higher than for the other models. The two cases where it performs well are when $\mathcal{G}$ is either the LVM or the HKEVP. In other words, this model is good when true marginal distributions come from Gaussian processes. Recall that



(a) Generator: LVM. (b) Generator: HKEVP. (c) Generator: EGP.



(d) Generator: BRP. (e) Generator: ETP.

Figure 7: Estimations $\hat{\theta}_{(s_1,s_2)}(\mathcal{F};\mathcal{G})$ of criterion 2 for all generators $\mathcal{G}$ and fitting models $\mathcal{F}$.

when fitted on real data, the HKEVP is also the most atypic, with a shape parameter ξ that is twice the value of the other generators. However, despite the poor robustness in marginal estimation, this model is the best one in terms of the second criterion. It has a very good accuracy and robustness, with the only exception being when $\mathcal{G}$ is the EGP. When data are generated with the ETP, it is the only method that performs very well.

The LVM of Davison et al. (2012) is a model designed for the margins, so it is expected that the performances of this method will be illustrated on the extrapolation of the return level. The competition performed in this article is in adequation with this expectation (see Figure 6). On the one hand, when data are generated with a max-stable model constructed with the spectral representation, the MSE associated to the LVM fits are greater than those obtained by the ETP and the BRP (see Figures 6c, 6d and 6e). On the other hand, if the data come from the atypic simulations of the HKEVP, the LVM has a stable MSE (see Figure 6b). It can be concluded then that despite being outperformed by the ETP and the BRP, the LVM shows the best robustness and thus appears more reliable. However, this model assumes conditional independence, so it cannot detect any trace of spatial dependence. The estimations of the extremal coefficient then always equal 2 under this method.

The model that comes in last position of the competition is the EGP of Schlather (2002). The MSE obtained by this model is the highest one for three generators over five (Figures 6a, 6b and 6d) and is the second highest for the two other generators. In particular, when data are generated with the HKEVP, the MSE is huge. As for the second criterion, the drawback of the EGP being unable to model independence (mentioned in Section 2.2.1) is illustrated in this comparative study. On the one hand, as long as the true bivariate extremal coefficient is lower than the possible limit 1.707, the results are rather similar to those of the ETP and the BRP. In particular in Figure 7e, it shows the same biased

estimations than the ETP. On the other hand, if the true value of θ is higher than 1.707, (Figures 7a, 7b and 7d), the EGP is degenerated toward its limit value.

3.3. Beyond the scope of this study: some additional comparisons

In the previous section, a comparative study has been driven between five spatial models based on two risk measures. The first one assesses the marginal behavior through the extrapolation of a 100-years return level, while the second evaluates the dependence structure with the bivariate extremal coefficient. While these two criteria summarize the efficiency of the models for spatial extreme value analyses, it may seem insufficient to just look at these values for a complete comparison. In this section, we discuss other properties of the five spatial models that have not been fully taken into consideration yet. The aim is to guide the practitioner for a better choice of model, depending on the pursued goal.

3.3.1. Bayesian models

Two Bayesian models have been used: LVM and HKEVP. The results obtained when fitting these two models are Markov chains that, if convergence is assumed, represent a posterior distribution for each parameter or for a function of parameters like the 100-years return level. To be feasible, the competition needed to focus on point estimates of the two criteria and it has been chosen to take the median of the posterior distributions. This choice may therefore seem rather restrictive for such models that can provide much more information on a given parameter, as an estimation of the uncertainty associated to the point estimates, for example. Bayesian models present however two main drawbacks at the inference step:

- First, they have generally more input parameters than the non-Bayesian ones. For instance, the user of the function `latent` in `SpatialExtremes` must provide values that control initial steps, prior distributions, random walk jumps for candidate generation and, above all, the number of iterations to use in the algorithm for assessing convergence (burn-in period) and to provide a satisfactory sample of the posterior distribution. For functions `hkevp.fit` and `latent.fit`, default values are available for these arguments, though they remain to be considered with care.

- Secondly, these models are more time consuming to fit, especially when the desired length of the posterior samples is high. In our case, one estimation with this model on the real or artificial dataset takes around 30 minutes and one hour respectively with LVM and HKEVP. The model of Reich and Shaby (2012) is thus the heaviest in this case and the computational time increases with the number of sites n and the number of knots L.

Because of the time needed to get estimations from the LVM and the HKEVP, parallel computing is strongly advised to perform the comparative study of Section 3. We worked with R package `parallel` (Ripley et al., 2015), which uses the random number generator of L'Ecuyer et al. (2002) to assess for independence between worker processors.

3.3.2. More about the HKEVP

A few drawbacks can be pointed out for the HKEVP. They are listed here:

- As discussed by Castruccio et al. (2015), realizations from the HKEVP are dependent on the choice of spatial knots that define the kernels in the dependence structure. When fitting data with the HKEVP, the choice of knots has to be done carefully and has to be seen as a trade-off between efficient estimation and computational burden. Indeed, all values of the random effect are updated at each MCMC step, which represents $L \times T$ parameters. The impact of a misspecification of the knots set has been studied in Reich and Shaby (2012). The conclusions can be summarized saying that too few knots may lead to a larger bias in the estimation of the GEV parameters, while considering too many knots than necessary does not improve significantly these estimations but increases the computational time drastically.

- A second drawback is the non-mixing property of the model. Indeed, even with a great number of iterations and with a large burn-in period, a huge thinning procedure is required to obtain stationary, non-correlated resulting Markov chains. Otherwise, the chains (in particular the ones of α and τ) show traces of dependence. In the comparison study, this aspect has been ignored because only point estimates were needed.

- Another drawback appears when the exact value of α is near 0, case where the spatial process is very smooth. In this case, the Markov chains are evolving very slowly due to the fact that the values of the random effect A are uniformly distributed over $\mathbb{R}_+\backslash\{0\}$. However, it has to be noted that this feature appears when the observed process is perfectly smooth (as for the Smith (1990)'s model). There is no reason that annual maxima of a natural phenomenon such as precipitation should show this type of realization.

Despite these shortcomings, the HKEVP model has two main advantages. The first one is that conditioning on the random effect A, we obtain independent responses in the hierarchical formulation. This allows us to use the full likelihood of the process rather than the composite likelihood. The second advantage is that its exponent function V is explicit for any set of sites. Comparatively, the exponent functions for the EGP, the BRP and the ETP are computed with multivariate Gaussian or Student cdf and thus are only explicit when evaluated at pairs of sites. If the goal is the evaluation of a multivariate probability on the spatial process $Y(\cdot)$, approximations must be made for this model (Genz and Bretz, 2009), which increases the computational cost. This is particularly the case for the ETP (see Thibaud and Opitz (2015) in a peaks-over-threshold approach).

3.3.3. Summary of the comparison

As a summary, Table 1 provides a visual assessment of the five models over the two criteria and the points discussed above.

	LVM	HKEVP	EGP	BRP	ETP
Marginal extrapolation	✓	✗	✗	≈	✓
Joint probability	✗	✓	✗	✓	≈
Bayesian approach	✓	✓	✗	✗	✗
Fast program	≈	✗	✓	✓	≈
Explicit multivariate cdf	✓	✓	✗	✗	✗

Table 1: Sketch of the characteristics of the five spatial max-stable models. A check ✓(resp. a cross ✗) means that the model is performant (resp. not advised) for the corresponding criterion or satisfies (resp. not) the given characteristic. In some cases where the decision is difficult to make, a "≈" symbol is given.

4. Discussion and conclusion

4.1. General conclusions

In this article, five models for spatial extreme values are competing over two risk measures that represent usual interest in application: the extrapolation of a 100-years return level at an ungauged site and the estimation of the bivariate extremal coefficient.

Results from Section 3.2.4 show dissimilarities between models and tend to discard some of them, depending on what is the main objective. On the one hand, if the interest lies in the estimation of the marginal effect, one should prefer the LVM of Davison et al. (2012) or the ETP of Opitz (2013). On the other hand, if the goal involves the modeling of the joint dependence structure (e.g. a joint probability), the HKEVP of Reich and Shaby (2012) seems the best choice but the BRP of Brown and Resnick (1977) or the ETP may also provide reliable estimates.

It is important to note that this comparative study has been made under circumstances that may influence the general conclusions. Namely, the whole simulation design was driven from the set of precipitation data using a parametric bootstrap. This procedure has been chosen for seek of objectivity. If this competition were realized with different data, the conclusions may have been slightly different. For instance, a previous version of the present article (Sebille et al., 2016) was based on precipitation data that were sparsely located over France, thus exhibiting spatial asymptotic independence. In this case, unsurprisingly, the LVM of Davison et al. (2012) was always the best model in terms of the extrapolation of the 100-years return level, while the HKEVP was described as a good compromise towards both the marginal extrapolation and the estimation of the dependence structure.

As discussed in Section 3.3, the comparison between these spatial models may also be generally more complex because of several features that characterize each of them. For instance, the HKEVP is the only one which can give an explicit formulation of the dependence structure for an arbitrary set of sites. This allows in particular conditional sampling of the yearly maxima process, though conditional simulation on max-stable processes can also be performed, see Dombry et al. (2013) for instance. However, its inference via MCMC is less easy to handle, it involves arbitrary choices like the positions of the knots, and it demands more computational resources than the other four models to be properly fitted.

This paper can be regarded as a practical guide when fitting annual maxima of precipitation data. Depending on the question of interest, the user has to choose between max-stable (EGP, BRP and ETP) or hierarchical max-stable (LVM and HKEVP) models. Moreover, the simulation plan is based on real precipitation data so that the comparison does not suffer from subjectivity.

4.2. Return level maps with the best models

Using the conclusions of this comparison study, we now provide an interpolation map of the 100-years return level of precipitation over the studied region. To this end, we use a regular grid of positions that covers the region of interest $\mathcal{S}$. Elevation for each point is represented in Figure 8 with a color code.

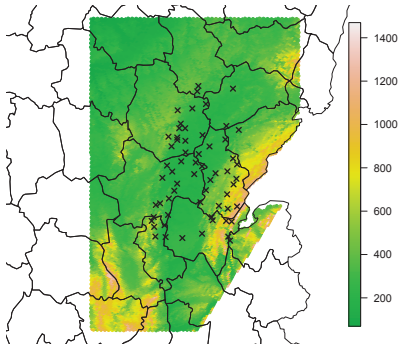
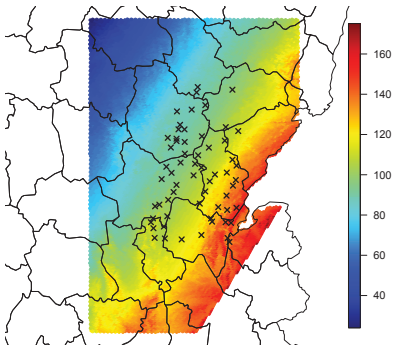


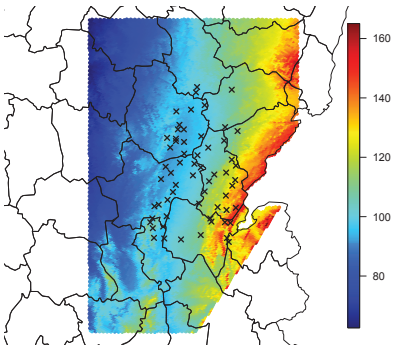
Figure 8: Elevation map of the central-east of France that covers the studied region $\mathcal{S}$ where data are located.

A part of the grid that corresponds to the Northern French Alps has been deliberately truncated because altitudes are much higher, making the prediction of the 100-years return level highly uncertain in this area. The effect on the interpolation map with this mountainous region was that only differences between plains and mountains were visible and variations in the whole regions were not illustrated well. The ETP and the LVM are the two spatial models that extrapolate better the marginal parameters, so we use them to produce the map of 100-years return level. Figure 9a (resp. 9b) displays the

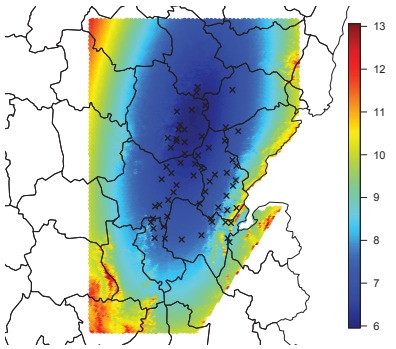
extrapolated map with the ETP (resp. LVM). Standard errors are shown on Figures 9c and 9d for the ETP and the LVM respectively. For LVM, it corresponds to the posterior standard deviation. How to compute these errors for ETP is explained in Appendix B.



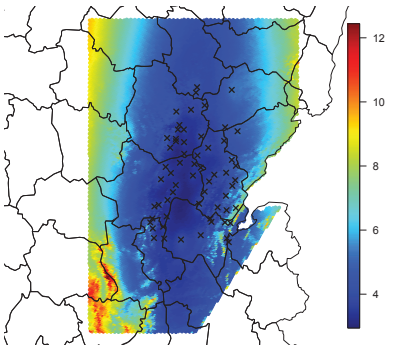
(a) Extrapolation map with ETP.



(b) Extrapolation map with LVM.



(c) Standard errors with ETP.



(d) Standard errors with LVM.

Figure 9: Maps of the extrapolated 100-years return level of precipitation and associated standard errors obtained with the ETP and the LVM.

The maps of Figures 9a and 9b show some dissimilarities between the extrapolation from the two spatial models. With ETP, the 100-years return level depends only on the spatial covariates which are the longitude, the latitude and the altitude. As a result, it increases with altitude and along the South-East direction. In the case of the LVM, the 100-years return level is affected by the variability in the latent Gaussian processes. The map of return levels show an increase towards the East direction and seems more correlated to the altitude covariate.
As expected, the more the prediction is made outside the convex hull of the stations, the greater the prediction error. Another feature that increases uncertainty in marginal extrapolation is the altitude: it can be seen in regions like the frontier with Switzerland or the upper-right part of the Auvergne massif (lower-left part of the region) where altitude is higher than 1.000 meters. Finally, we can see that the error obtained over the region $\mathcal{S}$ is slightly lower for the LVM (between 4mm and 6mm over most of the region) than for the ETP (between 6mm and 10mm over the same domain).

Acknowledgements

We would like to thank many contributors without whom this paper would never exist. First of all, we thank the authors of R packages `SpatialExtremes` and `RandomFields`, in particular Mathieu Ribatet for his very explicit functions around the EGP, the ETP and the BRP and the great help he provided us.
Secondly, we thank Aurélien Ribes from Météo France and Benjamin Renard from IRSTEA Lyon, who provided us useful tools to extract covariates information from the precipitation data set.
We also thank a lot Brian Reich and Benjamin Shaby for their model (HKEVP) and the indications they have given, allowing moreover the implementation of the R package `hkevp`.

This paper has been written during the PhD thesis of the first author. His thesis has been financed by EDF. We would like to thank several researchers from EDF R&D, namely Anne Dutfoy, Marie Gallois, Thi Thu Huong Hoang, Sylvie Parey and Nicolas Bousquet for numerous fruitful discussions that improved substantially this work.

Finally, we are grateful to Peter Craigmile who made numerous suggestions for improving the text, to the R project which provides free material, and to Météo France and EDF/DTG for allowing us to work freely on their data.

Appendix A. Upper bound of θ for the EGP

By following (Matérn, 1986, p. 16), the correlation function $\rho(\cdot)$ of a Gaussian process satisfies:

$$\rho(h) \geqslant \inf_h \Lambda_{\frac{n-2}{2}}(h) \ ,$$

with $\Lambda_k(h) = \Gamma(k+1)(2/h)^k J_k(h)$ and $J_k(\cdot)$ being the Bessel function of the first kind (Abramowitz and Stegun, 1964).

In our case $h \in \mathbb{R}^2$, so the correlation function satisfies $\rho(h) > -0.403$. By expression

$$\theta_{\mathrm{EGP}}(h) = 1 + \sqrt{\frac{1-\rho(h)}{2}}$$

of the EGP extremal coefficient, the inequality $\theta_{\mathrm{EGP}}(h) < 1.838$ is directly obtained. Recall that since we choose a powered exponential form for $\rho(\cdot)$ in the comparison, which is strictly positive, then the upper bound of θ_{EGP} becomes 1.707 in this case.

Appendix B. Computation of standard errors

In Section 4.2, return levels have been computed for a set of ungauged locations s^*, with associated standard errors. This appendix describes how to obtain them from the ETP.

The estimator $\hat{\Psi}$ of the ETP parameters $\Psi = (\psi_{\mathrm{GEV}}, \psi_V)$ is obtaiend by maximizing the pairwise likelihood (see Section 2.4) and is asymptotically normal (Ribatet, 2013):

$$\hat{\Psi} \sim \mathcal{N}\big(\Psi_0, H^{-1}(\hat{\Psi})J(\hat{\Psi})H^{-1}(\hat{\Psi})\big) \ ,$$

with $H(\Psi) = \mathbb{E}\big[\nabla^2 L_{\mathrm{pairwise}}(y;\Psi)\big]$ the Hessian matrix and $J = \mathrm{Var}\big(\nabla L_{\mathrm{pairwise}}(y;\Psi)\big)$ the variance score.
To compute standard errors for the 100-years return level extrapolated at s^* with ETP, we first need to get standard errors for $\hat{\Psi}$ by estimating $H(\hat{\Psi})$ and $J(\hat{\Psi})$:

- $\hat{H}(\hat{\Psi}) = \nabla^2 L_{\mathrm{pairwise}}(y;\hat{\Psi})$ is obtained straightforwardly by evluation of the Hessian matrix at $\hat{\Psi}$,
- $\hat{J}(\hat{\Psi}) = \sum_{t=1}^{T} \nabla L_{\mathrm{pairwise}}(y_t;\hat{\Psi}) \nabla L_{\mathrm{pairwise}}(y_t;\hat{\Psi})'$ (Varin and Vidoni, 2005).

The 100-years return level at any position s^* can be estimated by (7), which in terms of $\hat{\Psi}$ and a function h can be written $\hat{y}_{100}(s^*) := h(\hat{\Psi}, s^*)$. Standard errors associated to $\hat{y}_{100}(s^*)$ are then obtained using the Delta method:

$$\mathrm{Var}\big(\hat{y}_{100}(s^*)\big) = \nabla h(\hat{\Psi};s^*)\mathrm{Var}(\hat{\Psi})\nabla h(\hat{\Psi};s^*)' \ .$$

References

Abramowitz, M., Stegun, I. A., 1964. Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Vol. 55. Courier Corporation.

Apputhurai, P., Stephenson, A. G., 2013. Spatiotemporal hierarchical modelling of extreme precipitation in western australia using anisotropic gaussian random fields. Environmental and ecological statistics 20 (4), 667–677.

Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical modeling and analysis for spatial data. Monographs on statistics and applied probability. Chapman & Hall.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Vol. 558. John Wiley & Sons.

Brown, B. M., Resnick, S. I., 1977. Extreme values of independent stochastic processes. J. Appl. Probability 14 (4), 732–739.

Castruccio, S., Huser, R., Genton, M. G., 2015. High-order composite likelihood inference for max-stable distributions and processes. Journal of Computational and Graphical Statistics (just-accepted), 1–32.

Coles, S. G., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer-Verlag London Ltd., London.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., Sun, Y., 2012. A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. Revstat 10 (1), 135–165.

Cooley, D., Naveau, P., Poncet, P., 2006. Variograms for spatial max-stable random fields. In: Dependence in probability and statistics. Springer, pp. 373–390.

Davison, A. C., Huser, R., Thibaud, E., 2013. Geostatistics of dependent and asymptotically independent extremes. Mathematical Geosciences 45 (5), 511–529.

Davison, A. C., Padoan, S. A., Ribatet, M., May 2012. Statistical Modeling of Spatial Extremes. Statistical Science 27 (2), 161–186.

de Haan, L., 1984. A spectral representation for max-stable processes. The annals of probability, 1194–1204.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer.

Dombry, C., Ribatet, M., et al., 2013. Conditional simulation of max-stable processes. Biometrika 100 (1), 111–124.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., Stordal, F., 2015. Bayesian hierarchical modeling of extreme hourly precipitation in norway. Environmetrics 26 (2), 89–106.

Fawcett, L., Walshaw, D., 2014. Estimating the probability of simultaneous rainfall extremes within a region: a spatial approach. Journal of Applied Statistics 41 (5), 959–976.

Finkenstädt, B., Rootzén, H., 2004. Extreme values in finance, telecommunications and the environment. Chapman & Hall/CRC, Boca Raton.

Genz, A., Bretz, F., 2009. Computation of multivariate normal and t probabilities. Vol. 195. Springer Science & Business Media.

Gilleland, E., Katz, R. W., 2011. `extRemes`: New software to analyze how extremes change over time. R package.

Kabluchko, Z., Schlather, M., De Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. The Annals of Probability, 2042–2065.

L'Ecuyer, P., Simard, R., Chen, E. J., Kelton, W. D., 2002. An object-oriented random-number package with many long streams and substreams. Operations research 50 (6), 1073–1075.

Lindsay, B. G., 1988. Composite likelihood methods. Contemporary mathematics 80 (1), 221–39.

Matérn, B., 1986. Spatial variation, vol. 36 of. Lecture Notes in Statistics 2.

Opitz, T., 2013. Extremal t processes: Elliptical domain of attraction and a spectral representation. Journal of Multivariate Analysis 122, 409–413.

Padoan, S. A., Ribatet, M., Sisson, S. A., 2010. Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105 (489), 263–277.

Penot, D., 2014. Cartographie des événements hydrologiques extrêmes et estimation schadex en sites non jaugés. Ph.D. thesis, Hydrologie. Université Grenoble Alpes, 2014. <NNT:2014GRENU022>. <tel-01233267>. `https://tel.archives-ouvertes.fr/tel-01233267/file/41960_PENOT_2014_archivage.pdf`.

Reich, B. J., Shaby, B. A., 2012. A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics 6 (4), 1430.

Reich, B. J., Shaby, B. A., Cooley, D., 2014. A hierarchical model for serially-dependent extremes: A study of heat waves in the western us. Journal of Agricultural, Biological, and Environmental Statistics 19 (1), 119–135.

Ribatet, M., 2013. Spatial extremes: Max-stable processes at work.

Ribatet, M., 2015. `SpatialExtremes`: Modelling Spatial Extremes. R package version 2.0-2.

Ripley, B., Tierney, L., Urbanek, S., 2015. `parallel`. R package version 3.3.1.

Schlather, M., 2002. Models for stationary max-stable random fields. Extremes 5 (1), 33–44.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, P. J., Gross, S., Ober, U., Christoph Berreth, Burmeister, K., Manitz, J., Morena, O., Ribeiro, P., Singleton, R., Pfaff, B., R Core Team, 2016. RandomFields: Simulation and Analysis of Random Fields. R package version 3.1.12.
URL `http://CRAN.R-project.org/package=RandomFields`

Schlather, M., Tawn, J. A., 2003. A dependence measure for multivariate and spatial extreme values: Properties and inference. Biometrika 90 (1), 139–156.

Sebille, Q., 2016. `hkevp`: A hierarchical model for Spatial Extremes. R package version 1.0.

Sebille, Q., Fougères, A.-L., Mercadier, C., 2016. A comparison of spatial extreme value models: application to precipitation data. Unpublished: `https://hal.archives-ouvertes.fr/hal-01300751`.

Shaby, B. A., Reich, B. J., 2012. Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. Environmetrics 23 (8), 638–648.

Smith, R. L., 1990. Max-stable processes and spatial extremes. Preprint, Department of Mathematics, University of Surrey, Guilford.

Stephenson, A. G., 2009. High-dimensional parametric modelling of multivariate extreme events. Australian & New Zealand Journal of Statistics 51 (1), 77–88.

Stephenson, A. G., Shaby, B. A., Reich, B. J., Sullivan, A. L., 2015. Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. Journal of Applied Meteorology and Climatology 54 (2), 395–407.

Thibaud, E., Opitz, T., 2015. Efficient inference and simulation for elliptical pareto processes. Biometrika 102 (4), 855–870.

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. Biometrika 92 (3), 519–528.

Chapitre 6

Le modèle de Reich et Shaby (HKEVP)

Ce chapitre est entièrement consacré au modèle de Reich et Shaby (2012) désigné dans le chapitre précédent par l'acronyme HKEVP (*Hierarchical Kernel Extreme Value Process*).

6.1 Propriétés du HKEVP

6.1.1 Rappel de la construction du modèle

Rappelons la définition de ce modèle à la fois hiérarchique et max-stable. Soit Z le processus spatial défini pour tout $s \in \mathcal{S}$ par le produit

$$Z(s) = U_\alpha(s) \times \vartheta_\alpha(s) \ ,$$

où U_α est un processus indépendant de lois marginales $\mathrm{GEV}(1, \alpha, \alpha)$ et ϑ_α est le processus construit par :

- $\vartheta_\alpha(s) = \left(\sum_{\ell=1}^{L} A_\ell \omega_\ell(s)^{1/\alpha}\right)^\alpha$,
- $A_1, \ldots, A_L$ sont des variables indépendantes de loi stable positive $\mathrm{PS}(\alpha)$, d'exposant caractéristique α,
- $\omega_1, \ldots, \omega_L$ sont des fonctions déterministes positives vérifiant la condition $\sum_{\ell=1}^{L} \omega_\ell(s) = 1$ pour tout $s \in \mathcal{S}$.

Par cette construction avec les hypothèses données, Z est un processus max-stable simple et sa fonction de répartition évaluée sur l'ensemble de sites $\{s_1, \ldots, s_d\}$ s'écrit :

$$\Pr\big(Z(s_1) \leqslant z_1, \ldots, Z(s_d) \leqslant z_d\big) = \exp\left(-\sum_{\ell=1}^{L}\left[\sum_{j=1}^{d}\left\{\frac{\omega_\ell(s_j)}{z_j}\right\}^{1/\alpha}\right]^\alpha\right) \ . \tag{6.1}$$

Les marges de Z sont de loi Fréchet unitaire et il est possible de se ramener à un processus Y dont la loi marginale en un site s est une GEV de paramètres $\mu(s), \sigma(s)$ et $\xi(s)$ grâce à la transformation :

$$Z(s) = \left[1 + \xi(s)\frac{Y(s) - \mu(s)}{\sigma(s)}\right]^{1/\xi(s)} \ .$$

Ainsi, on obtient la formulation hiérarchique suivante pour le modèle de Reich et Shaby (2012) :

$$\begin{aligned}
Y(s)|\mu, \sigma, \xi, \alpha, \vartheta_\alpha &\overset{\text{indep}}{\sim} \mathrm{GEV}(\mu^*(\mathrm{s}), \sigma^*(\mathrm{s}), \xi^*(\mathrm{s})) \ , \\
\mu^*(s) &= \mu(s) + \frac{\sigma(s)}{\xi(s)}\left(\vartheta_\alpha(s)^{\xi(s)} - 1\right) \ , \\
\sigma^*(s) &= \alpha\sigma(s)\vartheta_\alpha(s)^{\xi(s)} \ , \\
\xi^*(s) &= \alpha\xi(s) \ , \\
\vartheta_\alpha(s) &= \left(\sum_{\ell=1}^{L} A_\ell \omega_\ell^{1/\alpha}(s)\right)^\alpha \ , \\
A_1, \ldots, A_L &\overset{iid}{\sim} \mathrm{PS}(\alpha) \ .
\end{aligned}$$

L'indépendance des réalisations du processus Y conditionnellement à l'effet aléatoire latent A permet de calculer la vraisemblance des maxima annuels observés, en prenant le produit des fonctions de densité des lois GEV pour chaque site. L'inférence de ce modèle est faite par un algorithme MCMC de type *Metropolis-within-Gibbs* (cf. Section 3.2), implémenté dans le package R `hkevp` (Sebille, 2016) et présenté en détails dans la section 6.2.

6.1.2 Preuves liées à la construction du processus

Fonction exposante du HKEVP

Cette section montre que la fonction de répartition du processus Z est (6.1). Cela implique entre autres que la fonction exposante V du modèle de Reich et Shaby (2012) évaluée aux positions $\{s_1, \ldots, s_d\}$ est de la forme :

$$V(z_1, \ldots, z_d) = \sum_{\ell=1}^{L} \left[\sum_{j=1}^{d} \left\{ \frac{\omega_\ell(s_j)}{z_j} \right\}^{1/\alpha} \right]^{\alpha} .$$

Démonstration. En premier lieu, on rappelle que la transformée de Laplace d'une variable aléatoire A de loi stable positive d'exposant caractéristique α s'écrit :

$$\mathbb{E}\left[e^{-tA}\right] = \int_0^\infty e^{-tA} f_{\mathrm{PS}(\alpha)}(A) dA = e^{-t^\alpha} \ , \quad t \in \mathbb{R} \ ,$$

où $f_{\mathrm{PS}(\alpha)}$ est la fonction de densité de la loi stable positive $\mathrm{PS}(\alpha)$. Cette densité n'est pas explicite pour $\alpha \in (0,1)$. Par la suite, on note par $A = (A_1, \ldots, A_L)$ le vecteur aléatoire dont les composantes sont iid de lois stable positive $\mathrm{PS}(\alpha)$ et par $a = (a_1, \ldots, a_L)$ une réalisation particulière de ce vecteur. La fonction de répartition multidimensionnelle $f_{\mathrm{PS}(\alpha)}(a)$ est le produit, par indépendance des a_ℓ, des fonctions marginales $f_{\mathrm{PS}(\alpha)}(a_\ell)$. On note aussi $\theta_\alpha(s) = \left(\sum_{\ell=1}^{L} a_\ell \omega_\ell(s)^{1/\alpha}\right)^{\alpha}$ pour désigner $\vartheta_\alpha(s) \mid A = a$.

La fonction de répartition jointe G du vecteur $\big(Z(s_1), \ldots, Z(s_d)\big)$ s'écrit :

$$\begin{aligned}
G(z_1, \ldots, z_d) &= \Pr(Z(s_1) \leqslant z_1, \ldots, Z(s_d) \leqslant z_d) \\
&= \int_{\mathbb{R}_+^L} \Pr(Z(s_1) \leqslant z_1, \ldots, Z(s_d) \leqslant z_d \mid A = a) f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \Pr\left(U_\alpha(s_1) \leqslant \frac{z_1}{\vartheta_\alpha(s_1)}, \ldots, U_\alpha(s_d) \leqslant \frac{z_d}{\vartheta_\alpha(s_d)} \;\middle|\; A = a \right) f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \prod_{j=1}^{d} \Pr\left(U_\alpha(s_j) \leqslant \frac{z_j}{\vartheta_\alpha(s_j)} \;\middle|\; A = a \right) f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \prod_{j=1}^{d} \exp\left[-\left(\frac{\theta_\alpha(s_j)}{z_j} \right)^{1/\alpha} \right] f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \prod_{j=1}^{d} \exp\left[-\left(\frac{\left(\sum_{\ell=1}^{L} a_\ell \omega_\ell(s_j)^{1/\alpha}\right)^{\alpha}}{z_j} \right)^{1/\alpha} \right] f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \prod_{j=1}^{d} \exp\left[-\sum_{\ell=1}^{L} a_\ell \left(\frac{\omega_\ell(s_j)}{z_j} \right)^{1/\alpha} \right] f_{\mathrm{PS}(\alpha)}(a) da \\
&= \int_{\mathbb{R}_+^L} \exp\left(-\sum_{\ell=1}^{L} a_\ell \underbrace{\left[\sum_{j=1}^{d} \left(\frac{\omega_\ell(s_j)}{z_j} \right)^{1/\alpha} \right]}_{t_\ell} \right) f_{\mathrm{PS}(\alpha)}(a) da \\
&= \mathbb{E}\left[e^{-\sum_{\ell=1}^{L} t_\ell A_\ell} \right] \\
&= \prod_{\ell=1}^{L} \mathbb{E}\left[e^{-t_\ell A_\ell} \right] \\
&= \prod_{\ell=1}^{L} e^{-t_\ell^\alpha} \\
&= \exp\left(-\sum_{\ell=1}^{L} t_\ell^\alpha \right) \\
&= \exp\left(-\sum_{\ell=1}^{L} \left[\sum_{j=1}^{d} \left(\frac{\omega_\ell(s_j)}{z_j} \right)^{1/\alpha} \right]^{\alpha} \right) .
\end{aligned}$$

□

A partir de ce résultat, il est facile de montrer que le processus Z est de marges GEV$(1,1,1)$ en utilisant l'hypothèse de normalisation des noyaux :

$$\begin{aligned}\Pr\big(Z(s)\leqslant z\big) &= \exp\left(-\sum_{\ell=1}^{L}\frac{\omega_\ell(s)}{z}\right)\\ &= \exp\big(-1/z\big)\ .\end{aligned}$$

Formulation hiérarchique

La formulation hiérarchique du modèle max-stable de Reich et Shaby (2012) utilise d'une part la forme du produit $Z(s)=U_\alpha(s)\vartheta_\alpha(s)$ et d'autre part la relation entre Z et Y :

$$Z(s)=\left[1+\xi(s)\frac{Y(s)-\mu(s)}{\sigma(s)}\right]^{1/\xi(s)}\ .$$

Démonstration. Comme le processus U_α est spatialement indépendant et de lois marginales GEV$(1,\alpha,\alpha)$, on a par le produit $U_\alpha\vartheta_\alpha$:

$$Z(s)\mid\vartheta_\alpha \overset{\text{indep}}{\sim} \text{GEV}\big(\vartheta_\alpha(s),\alpha\vartheta_\alpha(s),\alpha\big)\ ,$$

et comme Z est de marges Fréchet unitaires, on a :

$$\begin{aligned}\Pr\big(Z(s)\leqslant z\mid\vartheta_\alpha(s)\big) &= \exp\left(-\left[1+\alpha\frac{z-\vartheta_\alpha(s)}{\alpha\vartheta_\alpha(s)}\right]^{-1/\alpha}\right)\\ &= \exp\left(-\vartheta_\alpha(s)^{1/\alpha}z^{-1/\alpha}\right)\ .\end{aligned}$$

Les lois marginales du processus Y conditionné par ϑ_α sont données pour tout s par :

$$\begin{aligned}\Pr\big(Y(s)\leqslant y\mid\vartheta_\alpha(s)\big) &= \Pr\left(\mu(s)+\frac{\sigma(s)}{\xi(s)}\left[Z(s)^{\xi(s)}-1\right]\leqslant y\ \middle|\ \vartheta_\alpha(s)\right)\\ &= \Pr\left(Z(s)\leqslant\left[1+\xi(s)\frac{y-\mu(s)}{\sigma(s)}\right]^{1/\xi(s)}\ \middle|\ \vartheta_\alpha(s)\right)\\ &= \exp\left(-\vartheta_\alpha(s)^{1/\alpha}\left[1+\xi(s)\frac{y-\mu(s)}{\sigma(s)}\right]^{-1/\alpha\xi(s)}\right)\ .\end{aligned}$$

En posant pour tout $s\in\mathcal{S}$

$$\left\{\begin{array}{lcl}\mu^*(s) &=& \mu(s)+\frac{\sigma(s)}{\xi(s)}\left(\vartheta_\alpha(s)^{\xi(s)}-1\right)\\ \sigma^*(s) &=& \alpha\sigma(s)\vartheta_\alpha(s)^{\xi(s)}\\ \xi^*(s) &=& \alpha\xi(s)\end{array}\right.\ ,$$

on obtient par identifications successives de $\xi^*(s),\sigma^*(s)$ et $\mu^*(s)$:

$$\begin{aligned}\Pr\big(Y(s)\leqslant y\mid\vartheta_\alpha(s)\big) &= \exp\left(-\left[\vartheta_\alpha(s)^{-\xi(s)}+\xi^*(s)\frac{y-\mu(s)}{\alpha\vartheta_\alpha(s)^{\xi(s)}\sigma(s)}\right]^{-1/\xi^*(s)}\right)\\ &= \exp\left(-\left[\vartheta_\alpha(s)^{-\xi(s)}+\xi^*(s)\frac{y-\mu(s)}{\sigma^*(s)}\right]^{-1/\xi^*(s)}\right)\\ &= \exp\left(-\left[1+\vartheta_\alpha(s)^{-\xi(s)}-1+\xi^*(s)\frac{y-\mu(s)}{\sigma^*(s)}\right]^{-1/\xi^*(s)}\right)\\ &= \exp\left(-\left[1+\frac{\xi^*(s)}{\sigma^*(s)}\frac{\sigma^*(s)}{\xi^*(s)}\left(\vartheta_\alpha(s)^{-\xi(s)}-1\right)+\xi^*(s)\frac{y-\mu(s)}{\sigma^*(s)}\right]^{-1/\xi^*(s)}\right)\\ &= \exp\left(-\left[1+\frac{\xi^*(s)}{\sigma^*(s)}\left(y-\left[\mu(s)-\frac{\sigma^*(s)}{\xi^*(s)}\left(\vartheta_\alpha(s)^{-\xi(s)}-1\right)\right]\right)\right]^{-1/\xi^*(s)}\right)\\ &= \exp\left(-\left[1+\xi^*(s)\frac{y-\mu^*(s)}{\sigma^*(s)}\right]^{-1/\xi^*(s)}\right)\ .\end{aligned}$$

□

Max-stabilité

La propriété de max-stabilité du processus Z est démontrée dans l'annexe de Reich et Shaby (2012). Il faut montrer que pour tout ensemble de sites $\{s_1, \ldots, s_d\}$ et $t > 0$, on a l'égalité :

$$\Pr\big(Z(s_1) \leqslant tz_1, \ldots, Z(s_d) \leqslant tz_d\big)^t = \Pr(Z(s_1) \leqslant z_1, \ldots, X(s_d) \leqslant z_d) \ .$$

Démonstration. D'après la formule (6.1), on a :

$$\begin{aligned}
\Pr\big(Z(s_1) \leqslant tz_1, \ldots, Z(s_d) \leqslant tz_d\big)^t &= \exp\left(-\sum_{\ell=1}^{L}\left[\sum_{j=1}^{d}\left(\frac{\omega_\ell(s_j)}{tz_j}\right)^{1/\alpha}\right]^{\alpha}\right)^t \\
&= \exp\left(-\frac{1}{t}\sum_{\ell=1}^{L}\left[\sum_{j=1}^{d}\left(\frac{\omega_\ell(s_j)}{z_j}\right)^{1/\alpha}\right]^{\alpha}\right)^t \\
&= \Pr\big(Z(s_1) \leqslant z_1, \ldots, Z(s_d) \leqslant z_d\big) \ .
\end{aligned}$$

□

6.1.3 Remarques générales

Modélisation des marges

Les paramètres des lois marginales du processus Y sont modélisés par des processus gaussiens latents de la même manière que pour le modèle à variables latentes (LVM) de Davison *et al.* (2012). En effet, pour tout $s \in \mathcal{S}$:

$$\begin{aligned}
\mu(s) &= f_\mu(s) + \varepsilon_\mu(s) \ , \\
\sigma(s) &= f_\sigma(s) + \varepsilon_\sigma(s) \ , \\
\xi(s) &= f_\xi(s) + \varepsilon_\xi(s) \ ,
\end{aligned}$$

où f_μ, f_σ et f_ξ sont des fonctions déterministes décrivant la moyenne des processus μ, σ et ξ, et $\varepsilon_\mu, \varepsilon_\sigma$ et ε_ξ des processus gaussiens centrés de fonctions de corrélation respectives ρ_μ, ρ_σ et ρ_ξ.

Afin de s'assurer de la condition $\sigma(s) > 0$ pour tout $s \in \mathcal{S}$, Reich et Shaby (2012) modélisent le logarithme du paramètre d'échelle $\gamma(s) = \log \sigma(s)$ par un processus latent gaussien :

$$\gamma(s) = f_\gamma(s) + \varepsilon_\gamma(s) \ .$$

Définition des noyaux

L'ensemble des fonctions noyaux $\{\omega_1, \ldots, \omega_L\}$ doit vérifier la condition $\sum_{\ell=1}^{L} \omega_\ell(s) = 1$ pour tout $s \in \mathcal{S}$ pour que le modèle soit valide (cf. Section 6.1.2). Une façon simple de définir de tels noyaux est de considérer un ensemble de *nœuds* associés $\{v_1, \ldots, v_L\}$ répartis uniformément sur le domaine $\mathcal{S}$ et de poser pour tout $s \in \mathcal{S}$:

$$\omega_\ell(s) = \frac{K(s|v_\ell, \tau)}{\sum_{j=1}^{L} K(s|v_j, \tau)} \quad , \ \forall \ell = 1, \ldots, L \ ,$$

où les fonctions $K(\cdot|v_\ell, \tau)$ sont des noyaux centrés en v_ℓ et de paramètre *d'étendue* τ. Reich et Shaby (2012) choisissent des noyaux gaussiens :

$$K(s|v_\ell, \tau) = \frac{1}{2\pi\tau^2} \exp\left[-\frac{1}{2\tau^2}(s - v_\ell)^T(s - v_\ell)\right] \ .$$

Dans cette configuration, τ est le *paramètre des noyaux* du modèle : il contrôle l'amplitude de ces derniers et joue donc un rôle important dans la description de la structure de dépendance (voir section suivante).

En pratique, pour approcher le modèle de Reich et Shaby sur des données de maxima annuels, il est nécessaire de faire un choix sur le nombre de nœuds L. Ce choix arbitraire a un impact sur la qualité d'estimation des paramètres du modèle mais aussi sur le temps de calcul nécessaire pour estimer ces paramètres dans la procédure MCMC. Comme indiqué par Reich et Shaby (2012), ne pas considérer assez de nœuds augmente de l'erreur d'estimation, mais en choisir plus que nécessaire n'améliore pas cette erreur. Dans ce dernier cas par contre, le temps de calcul peut augmenter considérablement (voir Section 6.2 pour plus de détails).

Paramètres de dépendance

La structure de dépendance spatiale du HKEVP est décrite par les deux paramètres $\tau > 0$ et $\alpha \in (0, 1]$. Le rôle de ces paramètres sur la forme de la dépendance du processus max-stable Z est discuté dans cette section.

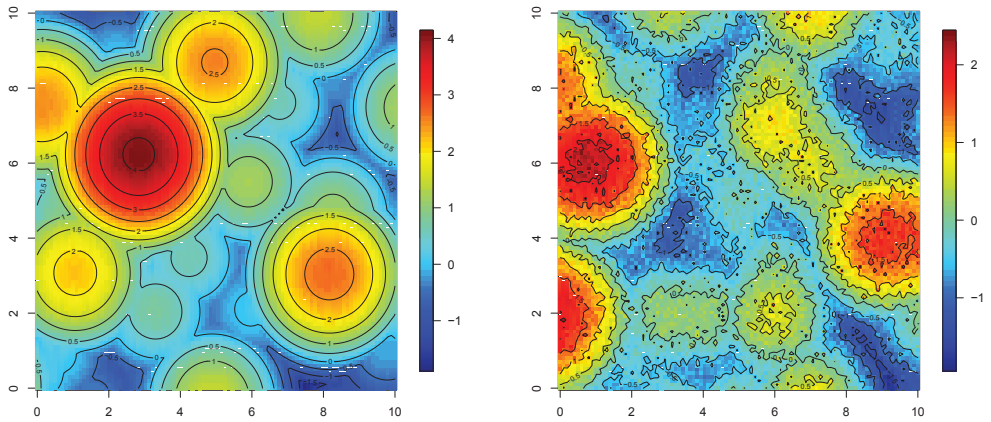
Le paramètre α intervient à plusieurs reprises dans la construction du modèle :
- comme paramètre d'échelle et de forme du processus indépendant U_α,
- comme paramètre de la forme logistique (Gumbel, 1960) appliquée lors de la construction du processus ϑ_α,
- en tant qu'exposant caractéristique de la loi stable positive associée aux variables latentes $A_1, \ldots, A_L$.

Ce paramètre sert de médiateur entre le processus indépendant U_α et le processus lisse ϑ_α.

D'un côté, si $\alpha = 1$, alors $\sum_{\ell=1}^{L} \omega_\ell(s) = 1$ par la condition sur les noyaux et en notant que la loi stable positive d'exposant caractéristique 1 est égale à la loi Dirac(1), on obtient $\vartheta_\alpha(s) = 1$ pour tout $s \in \mathcal{S}$. Le processus spatial Z est alors complètement indépendant : dans ce cas, le modèle de Reich et Shaby (2012) est similaire au modèle simple à variables latentes de Davison *et al.* (2012).

D'un autre côté, si $\alpha \approx 0$, le processus indépendant U_α joue un rôle moins important dans le produit : $U_\alpha(s) \approx 1$ pour tout $s \in \mathcal{S}$. Dans ce cas, le processus Z est entièrement déterminé par le processus lisse ϑ_α. En particulier, si le nombre de nœuds L tend vers $+\infty$ et que ceux-ci sont les réalisations d'un processus de Poisson homogène sur $\mathbb{R}^p$, alors pour $\alpha \to 0$, le modèle de Reich et Shaby (2012) converge vers le modèle max-stable de Smith (1990). Le HKEVP peut donc être considéré comme une généralisation du modèle de Smith, avec un paramètre α qui ajoute un degré de bruitage pour rendre le processus max-stable moins lisse.

Cette remarque est illustrée sur les figures 6.1a et 6.1b qui affichent des simulations ressemblantes entre les processus de Smith et du HKEVP avec $\alpha = 0.1$ respectivement. L'amplitude des noyaux (et des fonctions spectrales pour le modèle de Smith) a été fixée à 1, et 121 nœuds répartis sur la grille régulière $\{0, \ldots, 10\}^2$ ont été utilisés pour générer le processus de Reich et Shaby.



(a) Simulation selon le modèle de Smith. (b) Simulation selon le HKEVP pour $\alpha = 0.1$

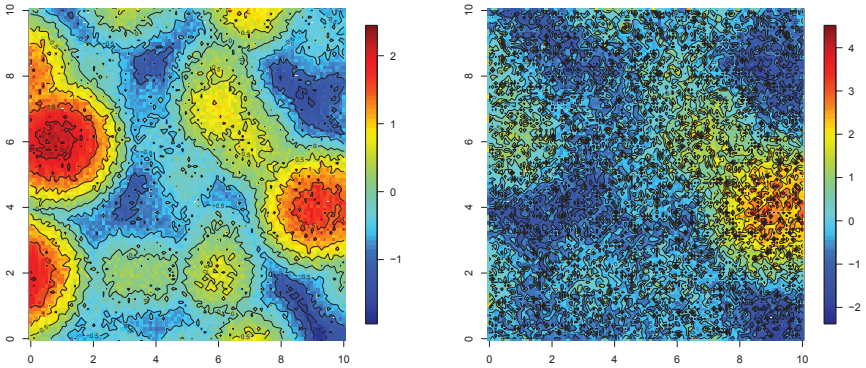
FIGURE 6.1 – Simulations de processus spatiaux Y selon les modèle de Smith (a) et HKEVP (b).

Dans le chapitre 5, il est mis en évidence que le paramètre α joue le rôle de *pépite* sur le processus Z. En géostatistique, la pépite correspond à l'erreur de mesure en un point $s \in \mathcal{S}$ donné. Dans le cas du modèle de Reich et Shaby (2012), cette erreur s'illustre sur le coefficient extrémal bivarié θ évalué à la même position :

$$\begin{aligned} \theta(s,s) &= \sum_{\ell=1}^{L} \left(\omega_\ell(s)^{1/\alpha} + \omega_\ell(s)^{1/\alpha} \right)^\alpha \\ &= 2^\alpha \underbrace{\sum_{\ell=1}^{L} \omega_\ell(s)}_{=1} = 2^\alpha \ . \end{aligned}$$

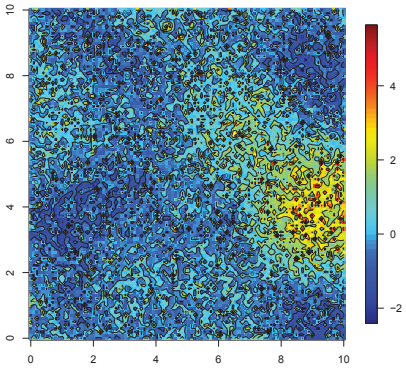
Pour $\alpha > 0$, $\theta(s,s) \neq 1$, et donc la dépendance complète n'est jamais prise en compte par ce modèle. Les processus max-stables de Schlather (2002) et Opitz (2013) peuvent vérifier la même propriété si un paramètre de pépite est ajouté à la fonction de corrélation utilisée dans la construction spectrale.

La figure 6.2 montre plusieurs réalisations du processus de Reich et Shaby (2012) pour différentes valeurs du paramètre α. On peut voir sur ces figures que quand α est petit, le processus spatial est relativement lisse et ressemble bien au modèle de Smith (1990) avec la formation de bulles correspondant aux densités gaussiennes bivariées. Plus la valeur de α augmente, plus le bruit s'intensifie, jusqu'à ce que le processus Z devienne totalement indépendant.

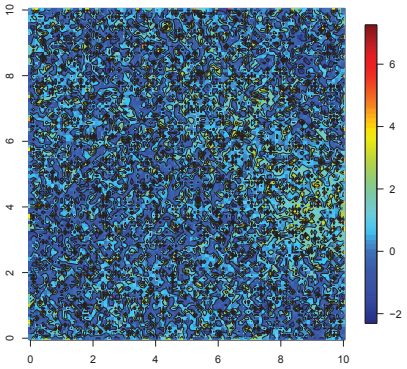


(a) $\alpha = 0.1$.

(b) $\alpha = 0.4$.



(c) $\alpha = 0.6$.



(d) $\alpha = 0.9$.

FIGURE 6.2 – Simulations de processus spatiaux Y selon le modèle max-stable HKEVP pour plusieurs valeurs du paramètre α.

Le rôle du paramètre des noyaux τ est illustré pour deux cas limites sur les simulations produites sur la figure 6.3. L'ensemble de nœuds choisis pour générer ces processus correspond à la grille de points régulière $\{0, \ldots, 10\}^2$. La valeur du paramètre α est fixée à 0.1 pour que le rôle de τ puisse être illustré à travers des réalisations du processus $Z \approx \vartheta_\alpha$.

Sur la partie de gauche, ce paramètre est très petit par rapport à la distance minimale de 1 entre les nœuds. Dans ce cas, le processus affiche un comportement dégénéré avec des zones spatiales fortement marquées. Ce phénomène s'explique par le fait que pour tout $s \in \mathcal{S}$, seul le noyau ω_{ℓ^*} centré sur le nœud v_{ℓ^*} le plus proche de s sera déterminant. Autrement dit, pour tout $\ell \in \{1, \ldots, L\}$:

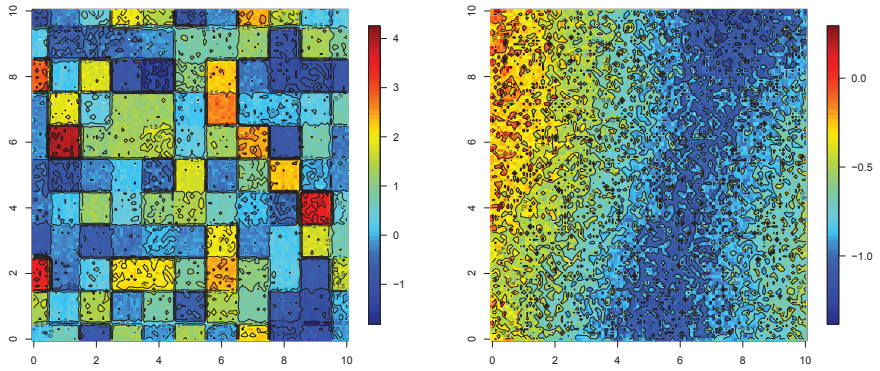
$$\omega_\ell(s) \approx \begin{cases} 0 & \text{si} \quad \ell \neq \ell^* \\ 1 & \text{si} \quad \ell = \ell^* \end{cases} ,$$

et donc :

$$\vartheta_\alpha(s) = \left(\sum_{\ell=1}^{L} A_\ell \omega_\ell^{1/\alpha}(s) \right)^\alpha = A_{\ell^*}^\alpha \ .$$

Les carreaux visibles sur la partie gauche de la figure 6.3 correspondent donc au pavage de Voronoï formé à partir de l'ensemble de nœuds $\{v_1, \ldots, v_L\}$.

Dans le cas contraire où la valeur de τ est grande par rapport au domaine (partie de droite de la figure 6.3),



(a) $\tau = 0.1$. (b) $\tau = 5$.

FIGURE 6.3 – Simulations de processus spatiaux Y selon le modèle max-stable HKEVP pour deux valeurs critiques du paramètre τ.

le processus obtenu ϑ_α est très plat. En effet, dans ce cas on a $\omega_1(s) \approx \cdots \approx \omega_L(s) \approx \omega^*$ constante et donc :

$$\begin{aligned} \vartheta_\alpha(s) &= \left(\sum_{\ell=1}^{L} A_\ell \omega_\ell^{1/\alpha}(s) \right)^\alpha , \\ &= \omega^* \left(\sum_{\ell=1}^{L} A_\ell \right)^\alpha , \end{aligned}$$

qui est invariant en fonction de la position $s \in \mathcal{S}$.

Le modèle dans la littérature

Brian Reich et Ben Shaby ont publié en complément de leur article de 2012 une série d'articles utilisant le modèle HKEVP, explorant à chaque fois un nouvel aspect dans la modélisation des valeurs extrêmes :

- Shaby et Reich (2012) utilisent le HKEVP en ajoutant une covariable temporelle dans les processus latents μ, σ et ξ pour prendre en compte l'effet d'une tendance dans les maxima annuels de températures en Europe.
  Le modèle est aussi utilisé pour réaliser une prédiction spatiale avec en particulier la dissociation entre la prédiction de type *krigeage* et la prédiction de type *climatique*. Cet aspect est discuté dans la partie 6.4.
- Reich *et al.* (2014) étudient la dépendance extrémale sur des séries temporelles de températures. Le modèle de Reich et Shaby (2012) est donc utilisé en dimension $p = 1$ sur des observations journalières extrêmes en utilisant l'approche des dépassements de seuil et la modélisation des excès par une loi Pareto généralisée (cf. Chapitre 1).
- Stephenson *et al.* (2015) appliquent la méthode d'inférence du modèle décrite dans la section 6.2 sur un réseau très dense de sites (environ 17.000). Pour alléger le calcul, le modèle fait appel à une loi a priori conditionnelle auto-régressive (CAR), comme pour les modèles hiérarchiques de Schliep *et al.* (2009) et Cooley et Sain (2010).

Castruccio *et al.* (2015) mènent une étude comparative sur la performance numérique de plusieurs modèles dont le HKEVP en utilisant le maximum de vraisemblance composite. Les auteurs se servent de l'estimation par maximum de vraisemblance composite et en faisant varier la dimension de la vraisemblance composite, c'est-à-dire en prenant les paires, puis les triplets, et ainsi de suite. Le gain en efficacité est ainsi comparé avec le temps de calcul nécessaire pour calculer la vraisemblance.

6.1.4 Prédiction spatiale

Pour une année donnée t, une question d'intérêt est de prédire la valeur du processus Y en un point $s^* \in \mathcal{S}$ en conditionnant par les observations en $\{s_1, \ldots, s_d\}$. Autrement dit, on veut estimer :

$$Y_t(s^*) \mid \{Y_t(s_1), \ldots, Y_t(s_d)\} , \tag{6.2}$$

La question de la prédiction spatiale pour les modèles max-stables est un axe de recherche actif dans la théorie des valeurs extrêmes. En effet, si les techniques usuelles de krigeage sont efficaces pour des observations gaussiennes (cf. Chapitre 2), ce n'est pas le cas pour les modèles spatiaux de valeurs extrêmes. Deux méthodes de prédiction peuvent être citées :

1. La méthode de Wang et Stoev (2011) échantillonne la valeur à prédire conditionnellement à ce qui est observé. Le modèle max-stable doit cependant avoir une représentation spectrale discrète. Cooley *et al.* (2012) étendent cette approche au cas où les observations correspondent à des dépassements de seuil et fournit une forme analytique de l'approximation de la loi conditionnelle au lieu d'un échantillon issu de la loi a posteriori.
2. Dombry *et al.* (2013) utilisent un modèle bayésien pour échantillonner la valeur à prédire spatialement pour un processus max-stable arbitraire. La méthode de cet article est basée sur les *hitting scenarii* des fonctions spectrales sur les points d'observations. Cette méthode est applicable à n'importe quel processus max-stable mais elle a l'inconvénient d'être très coûteuse en temps dès lors que l'on dispose d'un nombre important de stations. Oesting *et al.* (2014) proposent une méthode permettant d'améliorer cet aspect, tandis que Bechler *et al.* (2015) appliquent la procédure avec le modèle Extrémal-t.

Le modèle HKEVP permet de fournir une loi a posteriori prédictive plus simplement. En effet, conditionnellement aux variables latentes A, les réalisations de Y sont indépendantes et les paramètres marginaux μ, σ, ξ peuvent être extrapolés par krigeage. La procédure de prédiction est décrite dans la section 2.2 de l'article de Shaby et Reich (2012) pour une observation t fixée :

1. Calculer $\vartheta_{\alpha,t}(s^*)$, la valeur de $\vartheta_\alpha(s^*)$ sur l'année t, sachant $\{A_{1t}, \ldots, A_{Lt}, \alpha, \tau\}$.
2. Estimer les paramètres marginaux $\mu(s^*), \sigma(s^*), \xi(s^*)$ en utilisant le krigeage sur les processus gaussiens latents μ, σ et ξ.
3. Calculer les paramètres de la loi conditionnelle $\mu^*(s^*), \sigma^*(s^*)$ et $\xi^*(s^*)$.
4. Échantillonner $Y_t(s^*)$ selon la loi $\mathrm{GEV}\big(\mu_t^*(s^*), \sigma_t^*(s^*), \xi^*(s^*)\big)$.

L'algorithme précédent est appliqué pour chaque état de la chaîne de Markov résultant de la procédure d'estimation (voir Section 6.2). On obtient ainsi des réalisations de la loi a posteriori de (6.2) en échantillonnant $Y_t(s^*)$ selon la loi GEV conditionnelle à chaque état.

Shaby et Reich (2012) différencient deux types de prédiction qui diffèrent dans le choix des variables stables positives $A_t := \{A_{1t}, \ldots, A_{Lt}\}$ utilisées pour calculer $\vartheta_{\alpha,t}(s^*)$. La première méthode, dite de *type krigeage* a pour but d'estimer ce qui s'est réellement passé sur le site s^*, en se servant de la valeur observée de A_t (en pratique, en se servant de l'état de la chaîne de Markov associée à A_t). La seconde méthode, dite de *type climatologique* prédit ce qui aurait pu se passer avec les paramètres du modèle, en échantillonnant A_t de façon indépendante selon la loi stable positive $\mathrm{PS}(\alpha)$.

6.2 Inférence avec le modèle de Reich-Shaby

Soit $\{y_t(s_i)\}_{t\in 1,\ldots,T}$ les données observées de T maxima annuels sur l'ensemble de sites $\{s_1, \ldots, s_d\}$. Les paramètres à estimer pour le HKEVP sont :

$$\psi := \{\mu(s_1), \sigma(s_1), \xi(s_1), \ldots, \mu(s_d), \sigma(s_d), \xi(s_d), \alpha, \tau\} .$$

La procédure d'estimation des paramètres du modèle HKEVP correspond à l'algorithme MCMC de *Metropolis-within-Gibbs* (cf. Chapitre 3) dont les détails sont fournis en annexe dans l'article de Reich et Shaby (2012).

6.2.1 Algorithme général

Le principe général de l'algorithme, rappelé ici, est de produire une chaîne de Markov dont la loi stationnaire est la loi a posteriori du vecteur de paramètres inféré. Pour cela, les paramètres du modèle sont mis à jour successivement. A chaque étape de l'algorithme, une proposition est faite pour l'état suivant de la chaîne de Markov, qui est acceptée selon une probabilité calculée à partir de la vraisemblance du modèle et de la loi a priori associée au paramètre concerné. Si la proposition est rejetée, le nouvel état de la chaîne est égal à l'ancien.

La chaîne de Markov obtenue à la fin de l'algorithme représente un échantillon empirique de la loi a posteriori, sur laquelle il est possible de calculer des statistiques élémentaires comme la moyenne (pour une estimation ponctuelle) ou la variance (pour évaluer l'incertitude d'estimation).

La convergence en loi de la chaîne de Markov obtenue par MCMC peut être vérifiée en utilisant une période de chauffe et une procédure de thinning (voir Section 3.2.3 de ce manuscrit).

6.2.2 Mise à jour des paramètres

L'algorithme de Metropolis-within-Gibbs permettant d'estimer la loi a posteriori des paramètres ψ du modèle de Reich et Shaby est décrit pas à pas dans cette section, en se plaçant à une étape r arbitraire de la chaîne de Markov. Pour alléger les notations, le processus ϑ_α s'écrit ϑ.

Processus latents : paramètres GEV

D'après les hypothèses formulées, $y_t(s)$ suit une loi GEV de paramètres μ, σ et ξ. Ces paramètres sont mis à jour successivement en procédant site par site. La procédure est donnée pour le paramètre de localisation μ et pour un site s_j particulier, mais elle est identique pour la mise à jour de σ et ξ et pour n'importe quel site de $\{s_1, \ldots, s_d\}$ où les données sont observées.

1. Un candidat $\mu_{(c)}(s_j)$ est généré en utilisant une marche aléatoire centrée en la valeur actuelle $\mu_{r-1}(s_j)$.
2. La probabilité d'acceptation du candidat est calculée par :

$$p_{\mu_{(c)}(s_j)} = \min\left\{1, \prod_{t=1}^{T}\prod_{j=1}^{d} \frac{g\big(y_t(s_j) \mid \mu_{(c)}(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta_t(s_j)\big)}{g\big(y_t(s_j) \mid \mu_{r-1}(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta_t(s_j)\big)} \times \frac{\pi\big(\mu_{(c)}(s_j) \mid \mu(s_{-j})\big)}{\pi\big(\mu_{r-1}(s_j) \mid \mu(s_{-j})\big)}\right\},$$

où $g\big(\cdot \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta_\alpha(s_j)\big)$ est la fonction de densité de la loi GEV de paramètres $\mu^*(s_j), \sigma^*(s_j)$ et $\xi^*(s_j)$, où $\pi\big(\mu(s_j) \mid \mu(s_{-j})\big)$ correspond à la loi a priori gaussienne de $\mu(s_j)$ sachant les valeurs actuelles de $\mu(s_{-j}) = \{\mu(s_i)\}_{i \neq j}$, et où $\vartheta_t(s_j)$ correspond à la valeur du processus ϑ en s_j à l'observation t.

Il peut arriver que l'un des paramètres GEV soit considéré comme constant spatialement. C'est souvent le cas pour le paramètre de forme ξ : $\xi(s) \equiv \xi_0$. Dans ce cas, la mise à jour se fait pour tous les sites en même temps : un candidat unique $\xi_{(c)}$ est généré et devient le nouvel état de la chaîne ξ_r avec probabilité

$$p_{\xi_{(c)}} = \min\left\{1, \prod_{t=1}^{T}\prod_{j=1}^{d} \frac{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi_{(c)}, \alpha, \vartheta_t(s_j)\big)}{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi_{r-1}, \alpha, \vartheta_t(s_j)\big)} \times \frac{\pi\big(\xi_{(c)}\big)}{\pi\big(\xi_{r-1}\big)}\right\},$$

où cette fois, $\pi\big(\xi\big)$ est la loi a priori normale associée au paramètre GEV constant.

Paramètres spatiaux des processus gaussiens latents

Les processus latents μ, σ et ξ sont des processus gaussiens décrits par un modèle linéaire (voir (4) dans le chapitre précédent) d'une part et par une fonction de covariance d'autre part. La forme exponentielle

$$c_\mu(h) = \delta_\mu \exp\left(-\frac{h}{\lambda_\mu}\right)$$

est choisie par Reich et Shaby (2012) pour le HKEVP et par Davison *et al.* (2012) de façon analogue pour le LVM, avec δ_μ le paramètre de palier et λ_μ le paramètre de portée (cf. Chapitre 2). Des formes similaires sont considérées pour les paramètres σ et ξ s'ils sont définis comme spatialement variables.

1. Pour le vecteur de coefficients β_μ de la régression linéaire, une loi a priori normale conjuguée est utilisée pour simuler directement le nouvel état de la chaîne de Markov.
2. Le palier δ_μ utilise également une loi a priori conjuguée de type Inverse-Gamma qui permet une simulation directe du nouvel état.
3. Pour le paramètre de portée, il est nécessaire de générer un candidat $\lambda_{\mu,(c)}$ et de calculer sa probabilité d'acceptation :

$$p_{\lambda_{\mu,(c)}} := \min\left\{1, \frac{\pi\big(\mu \mid \beta_\mu, \delta_\mu, \lambda_{\mu,(c)}\big)}{\pi\big(\mu \mid \beta_\mu, \delta_\mu, \lambda_{\mu,r-1}\big)} \times \frac{\pi\big(\lambda_{\mu,(c)}\big)}{\pi\big(\lambda_{\mu,r-1}\big)}\right\},$$

où $\pi\big(\mu \mid \beta_\mu, \delta_\mu, \lambda_\mu\big)$ est la fonction de densité de la loi gaussienne multivariée associée au processus latent μ et $\pi(\lambda_\mu)$ est la fonction de densité de la loi a priori associée au paramètre de portée λ_μ.

Paramètres de dépendance

Les deux paramètres de dépendance α et τ définissent en particulier le processus ϑ_α, qui doit donc être calculé dès qu'un candidat est généré pour α ou pour τ. Par souci de lisibilité, on utilise les notations :
- $\vartheta_{(c)}$ est le processus ϑ_α candidat,
- $\vartheta_{t,(c)}$ est l'observation au temps t de $\vartheta_{(c)}$,
- ϑ_t est l'observation au temps t de ϑ_α.

Les paramètres α et τ du HKEVP sont mis à jour successivement de la façon suivante :

1. Générer un candidat $\alpha_{(c)}$ et calculer le processus candidat $\vartheta_{(c)}$.
2. Calculer la probabilité d'acceptation :
$$p_{\alpha_{(c)}} = \min\left\{1, \prod_{t=1}^{T}\prod_{j=1}^{d} \frac{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha_{(c)}, \vartheta_{t,(c)}(s_j)\big)}{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha_{r-1}, \vartheta_t(s_j)\big)} \times \frac{\pi\big(\alpha_{(c)}\big)}{\pi\big(\alpha_{r-1}\big)}\right\} ,$$
où $g\big(\cdot \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta(s_j)\big\}$ est la fonction de densité de la loi GEV de paramètres $\mu^*(s_j), \sigma^*(s_j)$ et $\xi^*(s_j)$, et où $\pi\big(\alpha\big)$ correspond à la loi a priori associée au paramètre α.
3. Si le candidat est accepté, $\alpha_r = \alpha_{(c)}$ et le nouveau processus ϑ est le candidat $\vartheta_{(c)}$. Sinon, $\alpha_r = \alpha_{r-1}$ et la valeur du processus ϑ ne change pas.
4. Générer un candidat $\tau_{(c)}$ pour le paramètre des noyaux τ et calculer le processus candidat associé $\vartheta_{(c)}$.
5. Calculer la probabilité d'acceptation :
$$p_{\tau_{(c)}} = \min\left\{1, \prod_{t=1}^{T}\prod_{j=1}^{d} \frac{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta_{t,(c)}(s_j)\big)}{g\big(y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha, \vartheta_t(s_j)\big)} \times \frac{\pi\big(\tau_{(c)}\big)}{\pi\big(\tau_{r-1}\big)}\right\} ,$$
où $\pi\big(\tau\big)$ est la fonction de densité de la loi a priori associée au paramètre τ.

Effet latent stable positif

Les variables aléatoires indépendantes $\{A_{\ell t}\}$, $\ell \in \{1, \ldots, L\}$ et $t \in \{1, \ldots, T\}$ sont mises à jour élément par élément en simulant à chaque fois un candidat. La difficulté rencontrée pour une variable aléatoire de loi stable positive est que la fonction de densité $f_{\mathrm{PS}(\alpha)}$ n'a pas de forme explicite pour $\alpha \in (0, 1)$, ce qui rend impossible le calcul direct de la probabilité d'acceptation.
Pour résoudre ce problème, Stephenson (2009) suggère d'utiliser une variable auxiliaire, notée $B \in (0, 1)$, telle que le couple (A, B) suit une loi f_{AB} qui s'écrit :
$$f_{\mathrm{AB}}(A, B \mid \alpha) = \frac{\alpha A^{-1/(1-\alpha)}}{1-\alpha} c(B) \exp\left[-c(B) A^{-\alpha/(1-\alpha)}\right] ,$$
avec
$$c(B) = \left[\frac{\sin(\alpha\pi B)}{\sin(\pi B)}\right]^{1/(1-\alpha)} \frac{\sin\big((1-\alpha)\pi B\big)}{\sin(\alpha\pi B)} .$$
Sous ces hypothèses, $A \mid B \sim \mathrm{PS}(\alpha)$. La variable auxiliaire B doit alors être mise à jour à chaque étape de l'algorithme MCMC. La probabilité d'acceptation pour B est calculée en se servant de la loi $f(A, B|\alpha)$. La méthode s'écrit donc pour tout $\ell \in \{1, \ldots, L\}$ et pour tout $t \in \{1, \ldots, T\}$:

1. Générer un candidat $A_{\ell t,(c)}$ selon une marche aléatoire log-normale centrée sur la valeur actuelle $A_{\ell t,r-1}$ et calculer le processus $\vartheta_{t,(c)}$ associé à cette réalisation.
2. Calculer la probabilité d'acceptation :
$$\begin{aligned} p_{A_{\ell t,(c)}} \quad = \quad & \min\left\{1, \prod_{j=1}^{d} \frac{g\big(Y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha_{(c)}, \vartheta_{t,(c)}(s_j)\big)}{g\big(Y_t(s_j) \mid \mu(s_j), \sigma(s_j), \xi(s_j), \alpha_{r-1}, \vartheta_t(s_j)\big)}\right. \\ & \left. \times \frac{f_{\mathrm{AB}}\big(A_{\ell t,(c)}, B_{\ell t} \mid \alpha\big)}{f_{\mathrm{AB}}\big(A_{\ell t,r-1} B_{\ell t} \mid \alpha\big)} \times \frac{q_{\mathrm{LN}}\big(A_{\ell t,r-1} \mid A_{\ell t,(c)}\big)}{q_{\mathrm{LN}}\big(A_{\ell t,(c)} \mid A_{\ell t,r-1}\big)}\right\} , \end{aligned}$$
où $q_{\mathrm{LN}}(x \mid y)$ est la fonction de densité de la loi log-normale centrée en y utilisée pour générer le candidat $A_{t,(c)}$.
3. Générer un candidat $B_{\ell t,(c)}$ selon la loi log-normale centrée en $B_{\ell t,r-1}$.
4. Calculer la probabilité d'acceptation :
$$p_{B_{\ell t,(c)}} = \min\left\{1, \frac{f_{\mathrm{AB}}\big(A_{\ell t}, B_{\ell t,(c)} \mid \alpha\big)}{f_{\mathrm{AB}}\big(A_{\ell t} B_{\ell t,r-1} \mid \alpha\big)} \times \frac{q_{\mathrm{LN}}\big(B_{\ell t,r-1} \mid B_{\ell t,(c)}\big)}{q_{\mathrm{LN}}\big(B_{\ell t,(c)} \mid B_{\ell t,r-1}\big)} \mathbb{1}_{\left\{B_{\ell t,(c)} \in (0,1)\right\}}\right\} .$$

6.2.3 Lois a priori

Reich et Shaby (2012) indiquent les lois a priori utilisées dans la procédure d'inférence de leur modèle pour tous les paramètres. Basée sur leurs suggestions, la liste des lois a priori choisies dans la comparaison du chapitre 5 est donnée ici :

1. Les éléments de β_μ, β_σ et β_ξ suivent des lois a priori indépendantes $\mathcal{N}(0, 100^2)$.
2. Les paramètres de palier $\delta_\mu, \delta_\sigma$ et δ_ξ ont pour loi a priori la loi InvGamma$(0.1, 0.1)$.
3. Les lois a priori définies pour les paramètres de portée $\lambda_\mu, \lambda_\sigma$ et λ_ξ ainsi que pour le paramètre des noyaux τ sont des lois InvGamma$(0.1, 0.1)$ dans Reich et Shaby (2012). Une légère différence est apportée ici. En effet, on suppose
$$\frac{\psi}{2D_{\max}} \sim \text{Beta}(2, 5) \ ,$$
où $\psi \in \{\delta_\mu, \delta_\sigma, \delta_\xi, \tau\}$ et où $D_{\max}$ est la distance maximale observée entre les stations $\{s_1, \ldots, s_d\}$ et représente donc le diamètre de la région étudiée. Pour argumenter ce choix, on fait remarquer que ces paramètres n'ont aucune raison d'être plus grands que deux fois le diamètre de la région $\mathcal{S}$ et qu'ils ont plus de chances, en général, d'avoir des valeurs comprises dans l'intervalle $(0, D_{\max}]$, d'où le choix de la loi Beta$(2, 5)$.
4. Le paramètre α est associé à une loi a priori non informative Unif$(0, 1)$.

6.2.4 Modèle non mélangeant

Les chaînes de Markov obtenues en ajustant le modèle de Reich et Shaby (2012) sur les données de précipitations lors de l'étude comparative du chapitre 5 sont affichées sur la figure 6.4.

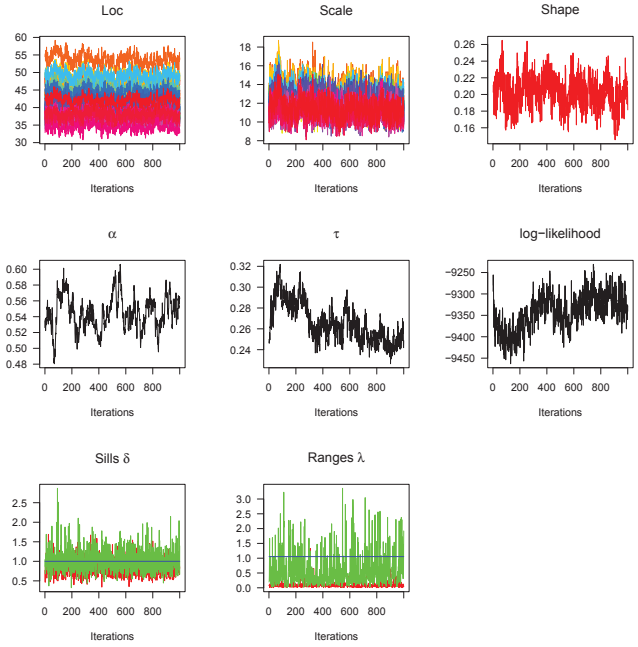


FIGURE 6.4 – Chaînes de Markov obtenues par le HKEVP de Reich et Shaby (2012) lorsqu'ajusté sur les données de précipitations dans l'étude comparative des modèles spatiaux.

Cette figure illustre l'un des aspects négatifs du modèle : celui d'être *non mélangeant*. Le résultat attendu est une chaîne de Markov stationnaire et non corrélée pour chaque paramètre du modèle. Toutefois, si ce constat peut être visuellement acceptable pour les valeurs du paramètre de localisation $\mu(s_j)$ et d'échelle $\sigma(s_j)$ pour $j \in \{1, \ldots, d\}$, ce n'est pas le cas pour le paramètre de forme spatialement constant ξ ou pour les paramètres α et τ décrivant la structure de dépendance.

Ces derniers paramètres affichent à la fois une corrélation et une non-stationnarité. Pour information, le test de Wald-Wolfowitz (Wald et Wolfowitz, 1943) décrit dans le chapitre 4 rejette l'hypothèse d'indépendance et de stationnarité pour toutes les sous-chaînes de Markov montrées sur la figure 6.4.
Pourtant, une importante procédure de thinning (cf. Chapitre 3) a été nécessaire pour ce modèle (de taille $n_{\text{th}} = 40$), mais qui n'a pas suffi à effacer la corrélation et la non-stationnarité dans les sous-chaînes de Markov. Les sous-chaînes obtenues avec le modèle à variables latentes de Davison *et al.* (2012) sont montrées sur la figure 6.5 à titre de comparaison.

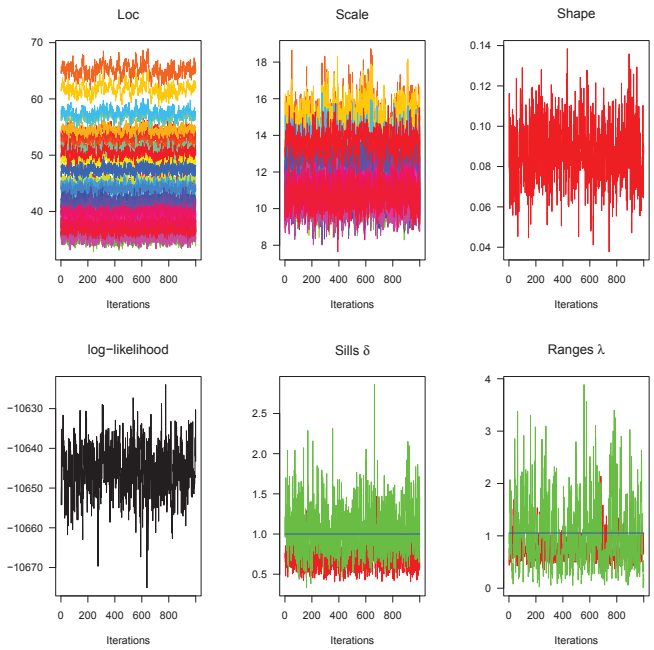


FIGURE 6.5 – Chaînes de Markov obtenues par le LVM de Davison *et al.* (2012) lorsqu'ajusté sur les données de précipitations dans l'étude comparative des modèles spatiaux.

Si le test de corrélation de Wald-Wolfowitz rejette également l'hypothèse nulle de stationnarité et d'indépendance des états des sous-chaînes de Markov associées au LVM, les graphes de la figure 6.5 permettent visuellement d'accepter le résultat comme les réalisations d'une loi stationnaire.

6.3 Avantages et inconvénients du modèle

Cette section résume les remarques qui ont été faites dans ce chapitre sur le HKEVP ainsi que ce qui a été dit dans les conclusions du chapitre 5 sur la compétition entre les modèles.

Le principal atout du modèle de Reich et Shaby (2012) est la forme explicite de sa fonction exposante V, et ce pour tout ensemble fini de sites $\{s_1, \ldots, s_d\}$. Le conditionnement par l'ensemble de variables latentes A permet de retrouver la fonction de répartition (6.1) et de donner une formulation de la dépendance spatiale des extrêmes observés sur un ensemble arbitraire de stations météorologiques.

Grâce à cette forme, la prédiction spatiale du processus max-stable observé peut se faire directement à partir des états de la chaîne de Markov résultant de la procédure d'inférence du HKEVP (Shaby et Reich, 2012).

Enfin, il a été montré dans l'étude comparative du chapitre 5 que cette approche est la plus robuste de la comparaison du chapitre 5 pour l'estimation du coefficient extrémal θ. Les résultats de ce modèle affichent même une performance supérieure à celle des modèles classiques utilisés dans la littérature des extrêmes comme les modèles de Brown-Resnick et Extrémal-t.

En revanche, cette méthode montre aussi certains défauts. Le premier est illustré dans l'étude comparative lorsque le but est l'extrapolation d'un niveau de retour centennal en une position non jaugé. Le processus tend à afficher un biais sur la prédiction, qui semble s'expliquer par une estimation du paramètre de forme ξ supérieure à celle des autres modèles. Le chapitre de conclusions de cette partie IImontre la carte d'interpolation du niveau de retour estimé par le HKEVP, où l'on voit des valeurs supérieures à ce qui est prédit par le LVM ou le modèle Extrémal-t.

Le HKEVP possède d'autres inconvénients dont il est important d'avoir conscience avant de l'utiliser pour l'analyse de valeurs extrêmes. D'abord, c'est une méthode très demandeuse en temps de calcul, en particulier lorsque les nombres de sites et de nœuds sont élevés. Ensuite, il s'agit d'un modèle non-mélangeant, comme le montre la section 6.2.4 : il est donc impératif d'utiliser l'algorithme MCMC pendant une longue période de temps et d'appliquer une taille n_{th} de thinning importante pour obtenir un échantillon stationnaire de la loi a posteriori. Enfin, en pratique, le choix de la position des nœuds $\{v_1, \ldots, v_L\}$ doit être fait par l'utilisateur. Par défaut, il est suggéré de faire correspondre cet ensemble avec celui des stations météorologiques $\{s_1, \ldots, s_d\}$.

6.4 Implémentation du package `hkevp`

Une des contributions de la thèse est l'implémentation du package `hkevp` (Sebille, 2016), qui contient les procédures d'estimation et de simulation du modèle de Reich et Shaby (2012) et du modèle à variables latentes de Davison *et al.* (2012), vu comme un cas particulier du HKEVP. Les principales fonctionnalités de ce package sont données dans cette partie.

6.4.1 Motivations

Le développement du package `hkevp` sur R (R Core Team, 2013) a été motivé par le manque de fonctions associées au modèle de Reich et Shaby (2012). Deux alternatives étaient jusqu'alors proposées :

- Le code R mis en ligne par Brian Reich sur son site internet, à l'adresse
  `http://www4.stat.ncsu.edu/` $\sim$ `reich/code/Bayes_GEV.R` .
  Cette fonction contient cependant plusieurs aspects non expliqués par rapport aux explications données dans Reich et Shaby (2012). De plus, le code étant entièrement écrit sous R, la procédure d'estimation devient très lente si l'utilisateur augmente le nombre de nœuds, de sites ou d'itérations.
- La fonction `abba`, disponible dans le package `extRemes` (Gilleland et Katz, 2011), qui correspond au modèle CAR de l'article de Stephenson *et al.* (2015). Si cette fonction d'ajustement codée en partie en C++ est très rapide d'exécution, elle ne permet pas d'extrapoler le processus en dehors du réseau de stations observées, ce qui est un objectif majeur de cette thèse.

Il a donc été décidé d'implémenter la méthode d'estimation du modèle décrite dans la section 6.2. en écrivant une partie importante du code en C++ grâce au package `Rcpp` (Eddelbuettel et François, 2011; Eddelbuettel, 2013). Implémenter la procédure en C++, qui est un langage compilé, permet d'améliorer grandement les performances et donc le temps de calcul nécessaire.

Les méthodes développées sont inclues dans un package dédié au HKEVP afin de les rendre accessibles. Le temps d'exécution de la fonction d'estimation reste important si le nombre de sites ou de nœuds est grand, mais il est jugé acceptable pour mener l'étude comparative du chapitre 5.

Par la suite, la procédure d'estimation associée au modèle à variables latentes de Davison *et al.* (2012) a été ajoutée au package. Ce modèle est en effet un cas particulier du modèle de Reich et Shaby (2012) où seules les lois marginales sont inférées. La fonction `latent` du package `SpatialExtremes` (Ribatet, 2015) permet aussi d'ajuster ce modèle.

6.4.2 Fonctionnalités

Hormis la procédure d'estimation décrite dans la section 6.2, le package `hkevp` propose plusieurs fonctionnalités associées au modèle de Reich et Shaby (2012) qui sont décrites ici. Les fonctions ont été pensées de telle sorte que l'utilisateur ait le moins d'arguments à fournir en entrée.

Par exemple, les lois a priori et les états initiaux ont tous des valeurs par défaut choisies par expérience sur des simulations. Les amplitudes des sauts servant à générer les candidats sont aussi fixées par le package, et sont adaptées dans la procédure en utilisant la méthode suggérée par Reich et Shaby (2012) et décrite dans la section 3.2.3 de cette thèse.

La liste des fonctionnalités principales du package est donnée ci-dessous :

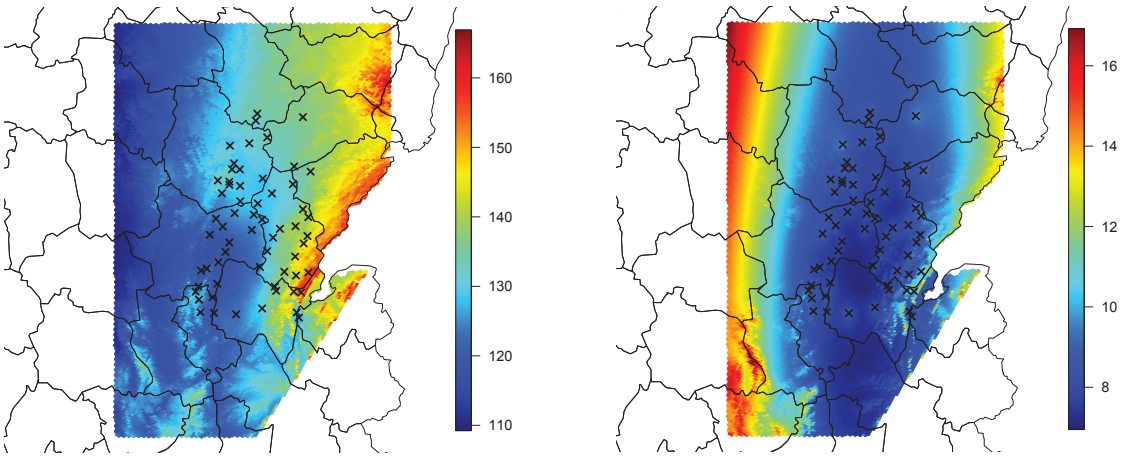
1. `hkevp.rand` permet de simuler des réalisations du HKEVP sur un ensemble de sites $\{s_1, \ldots, s_d\}$ donné,
2. `hkevp.fit` est la fonction d'ajustement du modèle décrite dans la section 6.2, où plusieurs choix sont offerts, comme celui de modéliser le paramètre d'échelle σ par $\gamma = \log(\sigma)$ ou encore de n'estimer que la structure de dépendance en prenant comme argument un processus max-stable simple,
3. `latent.fit` correspond à la fonction d'ajustement du modèle à variables latentes de Davison *et al.* (2012),
4. `mcmc.fun` et `mcmc.plot` prennent en argument la chaîne de Markov résultant de la procédure d'inférence : la première calcule une statistique (médiane, écart-type, etc.) pour tous les paramètres et la deuxième affiche les chaînes sur un graphe (cf. Figures 6.4 et 6.5),
5. `extrapol.gev` et `extrapol.return.level` utilisent la chaîne de Markov résultant de la procédure d'inférence pour estimer respectivement les paramètres (μ, σ, ξ) de la loi GEV d'une part et un niveau de retour d'autre part en un ensemble de positions cibles $\{s_1^*, \ldots, s_{d^*}^*\}$, en utilisant l'estimateur du krigeage (cf. Chapitre 2) sur les processus gaussiens latents.
6. `hkevp.predict` permet de prédire la valeur du processus max-stable Y sur des positions non jaugées $\{s_1^*, \ldots, s_{d^*}^*\}$ à partir des observations en $\{s_1, \ldots, s_d\}$ et du résultat de la procédure d'inférence en utilisant la méthode de Shaby et Reich (2012) décrite dans la section 6.1.4,
7. D'autres fonctionnalités annexes (`hkevp.expmeasure` et `return.level` par exemple) sont présentées dans le manuel associé au package et donné en annexe de ce manuscrit.

6.5 Cartes de niveaux de retour

La présentation de cartes d'interpolation du niveau de retour centennal des précipitations conclut la compétition entre les modèles spatiaux dans le chapitre 5. Les modèles LVM de Davison *et al.* (2012) et ETP de Opitz (2013) sont considérés comme les plus performants pour extrapoler les marges dans le cas général. C'est pourquoi ces deux approches sont sélectionnées pour produire les cartes d'interpolation.

La production de cartes de niveau de retour est l'objectif courant lors d'une étude spatiale des valeurs extrêmes. Par exemple, Davison *et al.* (2012) construisent de telles cartes sur des précipitations journalières en Suisse et Blanchet et Lehning (2010); Blanchet et Davison (2011); Gaume *et al.* (2013) affichent des cartes similaires dans les Alpes pour des niveaux de retour de chutes de neige dans le cadre de l'évaluation de risque d'avalanche.

Afin d'illustrer les différences obtenues, on affiche sur la figure 6.6a la même carte d'interpolation du niveau de retour obtenue avec le modèle de Reich et Shaby (2012) en prenant la médiane a posteriori. La figure 6.6b montre l'erreur associée à cette estimation à travers l'écart-type a posteriori.



(a) Médiane a posteriori.

(b) Écart-type a posteriori.

FIGURE 6.6 – Carte d'interpolation pour les précipitations de l'étude comparative obtenue avec HKEVP : (a) niveau de retour centennal et (b) écart-type associé.

Les conclusions sont similaires à celles obtenues par le modèle à variables latentes de Davison *et al.* (2012) dans le chapitre 5 : l'écart-type de l'estimation affiche un effet de bord très marqué avec une valeur faible sur la région où sont positionnées les stations et des valeurs plus fortes à mesure que l'on s'éloigne (en particulier vers l'ouest). L'erreur est aussi plus importante en haute altitude qu'en plaine : une partie des Alpes a d'ailleurs été supprimée sur la carte d'interpolation pour cette raison. La valeur élevée du niveau de retour centennal et l'erreur associée aux positions de hautes altitudes (supérieures à 2000m) a pour effet d'aplatir le reste des estimations et les rend uniforme à l'œil nu.

On peut expliquer les similitudes entre les cartes d'interpolation obtenues avec le HKEVP et celles obtenues avec le LVM en rappelant que les processus marginaux μ, σ et ξ sont modélisés par les mêmes formes, à savoir des processus gaussiens latents.

En revanche, si la forme des cartes est similaire, les estimations obtenues par le HKEVP ont une valeur plus élevée que celles obtenues avec le LVM. La figure 6.7 montre la différence d'estimation spatiale du niveau de retour entre les deux modèles : en montagne, la différence est très faible, mais en plaine (en particulier dans la zone nord-ouest de la région d'intérêt), la différence d'estimation est très importante, de l'ordre de 40mm à 50mm environ.

Au vu des conclusions tirées de la compétition sur le premier critère, le LVM peut toutefois être considéré comme plus fiable que le HKEVP pour la production de ces cartes.

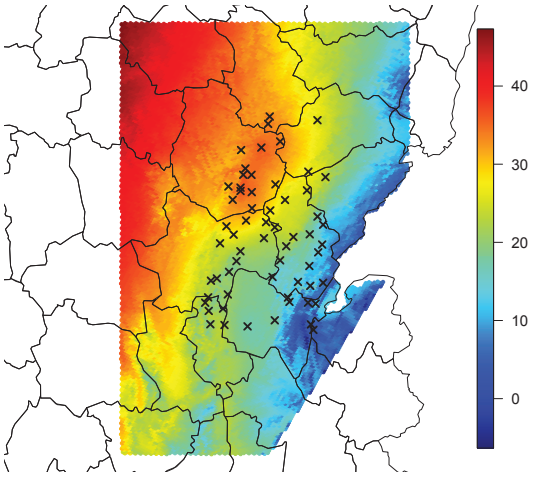


FIGURE 6.7 – Différences d'interpolation de niveau de retour centennal entre le HKEVP et le LVM.

Troisième partie

Modélisation spatiale des excès de seuil

Introduction de la partie III

La partie précédente de la thèse présente plusieurs modèles spatiaux ajustés sur les maxima par blocs d'un processus observé de façon régulière (journalière). Si cette approche permet de définir assez facilement des modèles paramétriques pour décrire le comportement extrémal d'un phénomène spatial, considérer le maximum par bloc (annuel) pour chaque station entraîne aussi une perte considérable d'information, car d'autres événements extrêmes importants ont pu apparaître dans le même bloc et être ignorés par sélection du maximum.

Dans cette partie, l'intérêt est porté sur l'extension de l'approche POT (*Peaks Over Threshold*) de la théorie des valeurs extrêmes au cadre d'un processus spatial. Ceci permet entre autres de répondre à une des problématiques de la thèse que les processus max-stables ne peuvent pas évaluer.

Événements extrêmes conditionnels

Un des objectifs de cette thèse est de proposer une méthode permettant d'estimer la probabilité qu'un jour donné, un événement extrême soit observé sur un ensemble de positions d'une région spatiale sachant que c'est le cas sur un autre ensemble de positions.

En pratique, l'intérêt lié à cette estimation peut être expliqué de plusieurs façons :

1. pour quantifier le risque joint sur une structure non jaugée, sachant un événement extrême,
2. afin d'évaluer le nombre de positions-clés à risque si un événement extrême est prévu dans une région donnée,
3. pour analyser la probabilité de contagion d'une valeur extrême : c'est-à-dire en regardant le nombre de sites frappés par des événements extrêmes si une station dépasse une certaine valeur critique.

La *probabilité d'échec conditionnelle* étudiée dans cette partie est en fait un moyen de décrire le comportement extrémal multivarié tout en répondant aux objectifs fixés au début de la thèse.

Les événements extrêmes conditionnels sont regardés de façon journalière : les processus max-stables étudiés dans la partie précédente ne sont donc pas adaptées pour répondre directement à ce type de question. Néanmoins, les modèles paramétriques associés à la fonction exposante V définis dans le chapitre 5 peuvent intervenir dans la description des excès de seuil d'un processus spatial. Ceci permet de fournir une estimation de la probabilité conditionnelle qui nous intéresse dans cette partie.

Dépassements de seuil multivariés et spatiaux

L'approche univariée des dépassements de seuil est un domaine de la théorie des valeurs extrêmes bien développé maintenant. Dans le cadre multivarié et spatial, il s'agit encore d'un axe de recherche en pleine expansion. Dans ce contexte, un dépassement de seuil peut avoir plusieurs définitions. Par exemple, on peut considérer comme extrême une observation journalière de $X = \big(X(s_1), \ldots, X(s_d)\big)$ telle que :

- toutes les composantes de X dépassent un certain seuil,
- au moins une des composantes de X dépasse un seuil fixé,
- l'image par une fonctionnelle dépasse un seuil. Cette approche permet de généraliser les deux précédentes.

La première méthode a l'avantage de regarder un vecteur dont toutes les composantes sont extrêmes, mais si d est grand ou si certaines composantes du vecteur X sont asymptotiquement indépendantes, le nombre d'observations extrêmes sera très faible, voire nul.

Pour des choix adéquats de seuils marginaux, la seconde méthode fournit assez d'observations pour permettre une inférence statistique. En revanche, certaines composantes peuvent ne pas être extrêmes, ce qui implique l'utilisation de méthodes de censure lors du calcul de vraisemblance.

Plan de la partie III

Dans un premier temps, le chapitre 7 recense plusieurs contributions scientifiques récentes dans le domaine des dépassements de seuil multivariés et spatiaux, en présentant les différentes approches envisagées et en mettant l'accent sur les processus ℓ-Pareto (section 7.2). Les problèmes rencontrés en souhaitant étudier les valeurs extrêmes d'un processus journalier sont mis en avant et des solutions (telles que la vraisemblance censurée pour l'inférence) sont données.

Dans un second temps, le chapitre 8 définit la probabilité d'échec conditionnelle et met en place plusieurs méthodes d'estimation liées à cette quantité. La section 8.4 du même chapitre propose une méthode adaptant certains estimateurs de la probabilité d'échec au cas où le processus X observé est temporellement corrélé dans les extrêmes.

Enfin, le chapitre 9 évalue à la fois les estimateurs de la probabilité d'échec conditionnelle et la méthode tenant compte de la dépendance temporelle à travers plusieurs plans de simulation.

Chapitre 7

Modélisation des dépassements de seuil multivariés et spatiaux

Ce chapitre recense différents articles récents issus de la littérature des valeurs extrêmes cherchant à modéliser les dépassements de seuil d'un processus multivarié ou spatial. Plusieurs approches sont présentées, dont les processus Pareto, un modèle de processus de Poisson pour les dépassements de seuil, un modèle adapté de la mesure angulaire et un modèle hiérarchique à variables latentes utilisé pour produire des cartes de niveaux de retour.

7.1 Dépassement de seuil pour un vecteur multivarié

Rootzén et Tajvidi (2006) étendent le théorème de Pickands-Balkema-de Haan sur la convergence des excès de seuil au cas d'un vecteur multivarié. À partir d'une loi MEV, les auteurs définissent ainsi les lois MGP (*Multivariate Generalized Pareto*) qui étendent les lois de Pareto généralisées au cas multivarié.

Définition 4 (Loi MGP). *H est la fonction de répartition d'une loi MGP si :*

$$H(x_1,\dots,x_d) = -\frac{1}{\log G(0,\dots,0)} \log\left(\frac{G(x_1,\dots,x_d)}{G(x_1\wedge 0,\dots,x_d\wedge 0)}\right) , \tag{7.1}$$

pour une loi MEV G à marges non dégénérées telle que $0 < G(0,\dots,0) < 1$.

Ainsi, $H(x_1,\dots,x_d) = 0$ si $x_j < 0$ pour tout $j \in \{1,\dots,d\}$ et

$$H(x_1,\dots,x_d) = 1 - \log G(x_1,\dots,x_d)/\log G(0,\dots,0)$$

si $x_j > 0$ pour tout $j \in \{1,\dots,d\}$.

Soit $X = (X_1,\dots,X_d) \in \mathbb{R}^d$ un vecteur aléatoire de fonction de répartition jointe F. Si F est dans le domaine d'attraction d'une loi MEV G telle que $0 < G(0,\dots,0) < 1$, le théorème 2.1 de Rootzén et Tajvidi (2006) montre qu'il existe une courbe d-dimensionnelle croissante de seuils $\{u(t) \ : \ t > 1\}$ telle que $u(1) = (0,\dots,0)$ et $F\big(u(t)\big) \xrightarrow[t\to\infty]{} 1$, et une fonction $a(u) = a\big(u(t)\big) > 0$, telles que le vecteur d'excès renormalisés $X_u = \frac{X-u}{a(u)}$ converge en loi vers une MGP si au moins une composante de X_u est positive :

$$\Pr(X_u \leqslant x \mid X_u \not\leqslant 0) \xrightarrow[t\to\infty]{} H(x) , \quad x \in \mathbb{R}_+^d ,$$

où H est définie par (7.1) avec G l'attracteur du vecteur X.

Rakonczai et Zempléni (2012) mettent en place une méthode d'inférence sur les excès de seuil multivariés pour estimer les paramètres de la loi MGP par maximum de vraisemblance. La fonction de densité de la loi MGP est obtenue en fonction de la mesure exposante V associée à la loi MEV de marges Fréchet unitaires. En utilisant la transformation

$$t_j(x_j) = \left(1 + \xi_j \frac{x_j - \mu_j}{\sigma_j}\right)_+^{-1/\xi_j} ,$$

où (μ_j, σ_j, ξ_j) sont les paramètres de la loi GEV marginale G_j associée à la composante X_j, la fonction de densité h de la loi MGP s'écrit sous la forme :

$$h(x_1,\dots,x_d) = \frac{\prod_{j=1}^d t_j'(x_j)}{V\big(t_1(0),\dots,t_d(0)\big)} \times \frac{\partial V}{\partial t_1 \dots \partial t_d}\big(t_1(x_1),\dots,t_d(x_d)\big) .$$

Un modèle paramétrique bivarié pour V est proposé par Rakonczai et Zempléni (2012) et testé avec cette procédure d'inférence sur des données de vent extrêmes.

7.2 Processus Pareto

Dans toute cette section et les suivantes, $\mathcal{S}$ désigne un sous-ensemble de $\mathbb{R}^p$, $\mathcal{C}(\mathcal{S})$ est l'ensemble des fonctions continues pour la norme uniforme sur $\mathcal{S}$ et $\mathcal{C}_+(\mathcal{S})$ est la restriction de $\mathcal{C}(\mathcal{S})$ aux fonctions positives.

7.2.1 Définition des processus Pareto

Les processus Pareto sont la généralisation des lois MGP au cas infini-dimensionnel, au même titre que les processus max-stables généralisent les lois MEV. Ils sont définis par Ferreira et de Haan (2014) comme le processus limite associé aux excès de seuil d'un processus $X \in \mathcal{C}(\mathcal{S})$.

Autrement dit, si $X \in \mathcal{C}(\mathcal{S})$ est dans le domaine d'attraction d'un processus max-stable à marges non dégénérées, alors il existe une surface croissante de seuils $\{u_t(s) \ : \ t > 1\}_{s\in\mathcal{S}} \in \mathcal{C}(\mathcal{S})$ telle que pour tout $s \in \mathcal{S}$, $F_{X(s)}\big(u_t(s)\big) \xrightarrow[t\to\infty]{} 1$, et une fonction $a(u) = a\big(u_t(s)\big) > 0$, telles que :

$$\left\{ \frac{X(s) - u_t(s)}{a(u)} \;\middle|\; \sup_{s\in\mathcal{S}} \frac{X(s) - u_t(s)}{a(u)} > 0 \right\}_{s\in\mathcal{S}} \xrightarrow[t\to\infty]{\mathcal{L}} \big\{W(s)\big\}_{s\in\mathcal{S}} \ .$$

Le processus $W \in \mathcal{C}(\mathcal{S})$ est appelé *processus GP*. Si cette limite W a de plus des marges Pareto standard, on dit que W est un *processus Pareto*.

Si W est un processus Pareto, le théorème 2.1 de Ferreira et de Haan (2014) montre qu'il vérifie les trois propriétés équivalentes suivantes :

1. W est stable en loi par seuillage. Autrement dit, le processus limite associé aux excès de seuil de W est de même loi que W.
2. La loi de W est homogène d'ordre -1 : $\Pr(W \in rA) = r^{-1}\Pr(W \in A)$ pour tout $A \subset \mathcal{S}$ et $r > 0$.
3. W admet la représentation $W(s) = YV(s)$ pour tout $s \in \mathcal{S}$, où Y est une variable aléatoire de loi Pareto standard et $V(s) \in \mathcal{C}_+(\mathcal{S})$ est un processus indépendant de Y vérifiant $\mathbb{E}[V(s)] > 0$ et $\sup_{s\in\mathcal{S}} V(s) = c$ presque sûrement, une constante positive.

L'article pionnier de Buishand *et al.* (2008) utilise déjà *l'approche constructive* correspondant à la troisième propriété du théorème 2.1 de Ferreira et de Haan (2014) pour définir un modèle de dépassements de seuil pour un processus spatial X. Cette construction permet de simuler des réalisations du processus spatial dépassant un seuil élevé au moins en un point s de la région $\mathcal{S}$, à partir des observations de X et des paramètres du modèle mis en place par Buishand *et al.* (2008). Cette approche est ensuite appliquée sur des données journalières pour estimer un niveau de retour sur le cumul régional de précipitations $\int_{s\in\mathcal{S}} X(s)ds$ observées sur une région des Pays-Bas.

7.2.2 Excès de seuil d'une fonctionnelle

Les processus Pareto ou GP étudiés par Ferreira et de Haan (2014) sont construits en définissant un dépassement de seuil spatial lorsqu'au moins une des composantes du vecteur observé $\big(X(s_1), \ldots, X(s_d)\big)$ est plus grande que son seuil marginal. Dombry et Ribatet (2015) étendent ces notions en généralisant la définition d'un dépassement de seuil par l'événement :

$$\big\{\ell(X) > u_\ell\big\} \ , \quad u_\ell \in \mathbb{R}_+ \ ,$$

pour une fonctionnelle donnée $\ell \ : \ \mathcal{C}_+(\mathcal{S}) \longrightarrow \mathbb{R}_+$, continue et homogène d'ordre $\beta > 0$: $\ell(\lambda X) = \lambda^\beta \ell(X)$.

La fonctionnelle ℓ détermine donc le type de dépassement de seuil associé à un processus X défini sur $\mathcal{S} \in \mathbb{R}^p$. Les observations correspondantes sont alors désignées sous le terme de ℓ-excès. Plusieurs exemples possibles de fonctionnelles sont donnés :

- Le cas étudié par Ferreira et de Haan (2014) (resp. par Rootzén et Tajvidi (2006) pour un vecteur multivarié) correspond à la fonctionnelle $\ell(X) = \sup_{s\in\mathcal{S}} X(s)$ (resp. $\ell(X) = \vee_{j=1}^d X_j$). Avec ce choix, un dépassement du processus X apparaît si un excès est observé en au moins un point de la région $\mathcal{S}$.
- Si un dépassement de seuil est défini lorsqu'un excès est observé sur toute la région $\mathcal{S}$ considérée, la fonctionnelle correspondante est $\ell(X) = \inf_{s\in\mathcal{S}} X(s)$.
- Le choix de fonctionnelle $\ell(X) = \int_{s\in\mathcal{S}} |X(s)|ds$ correspond aux dépassements du cumul total du processus $|X|$ observé sur la région $\mathcal{S}$.
- Si les excès de seuil de X sont conditionnés par les événements extrêmes observés sur une position $s_0 \in \mathcal{S}$ fixée, la fonctionnelle choisie est $\ell(X) = X(s_0)$.

Dombry et Ribatet (2015) montrent trois résultats principaux :

1. La convergence des ℓ-excès de seuil vers un processus ℓ-GP (Proposition 1 de l'article). Soit une fonctionnelle continue $\ell : \mathcal{C}(\mathcal{S}) \longrightarrow \mathbb{R}_+$ et $X \in \mathcal{C}(\mathcal{S})$ un processus stochastique tel que pour un processus W non dégénéré, on a la convergence faible dans $\mathcal{C}(\mathcal{S})$:

$$\Pr\big(u_\ell^{-1} X \in \cdot \mid \ell(X) > u_\ell\big) \xrightarrow[u_\ell \to \infty]{} \Pr(W \in \cdot) \ . \tag{7.2}$$

   Alors W est un processus ℓ-GP. Si de plus les marges de W sont de loi Pareto standard, W est appelé processus ℓ-Pareto.
2. Les trois propriétés des processus Pareto montrées par Ferreira et de Haan (2014) sont étendues au cas des processus ℓ-Pareto (Théorème 2 de l'article).
3. Enfin, un estimateur non-paramétrique de la structure de dépendance d'un processus ℓ-Pareto est construit à partir de la mesure spectrale associée (Proposition 3 de l'article).

7.2.3 Lien avec la mesure exposante

Thibaud et Opitz (2015) construisent une méthode d'inférence paramétrique sur les processus ℓ-GP en utilisant le modèle max-stable Extrémal-t de Opitz (2013). Cette procédure utilise le lien entre la formulation spectrale d'un processus ℓ-Pareto et la fonction exposante associée à la loi limite max-stable du processus observé. Les étapes de l'inférence sont détaillées à la section 8.2.1 du chapitre suivant, mais le lien entre la dépendance spatiale du processus ℓ-GP et le processus max-stable limite Z est explicité dans cette section.

Soit $X \in \mathcal{C}(\mathcal{S})$ le processus observé et $X_P \in \mathcal{C}_+(\mathcal{S})$ le processus de marges Pareto standard obtenu par (1.5). On suppose que X_P appartient au domaine d'attraction d'un processus max-stable simple Z non dégénéré. La dépendance spatiale du processus Z est entièrement caractérisée par la *mesure exposante* Λ vérifiant $\Lambda\big(\mathcal{C}(\mathcal{S}) \backslash \mathcal{C}_+(\mathcal{S})\big) = 0$ et :

$$\Lambda\left[\bigcup_{j=1,\dots,d} \Big\{ f \in \mathcal{C}(\mathcal{S}) \ : \ \sup_{s \in K_j} f(s) \geqslant z_j \Big\}\right] = -\log \Pr\left(\sup_{s \in K_1} Z(s) \leqslant z_1, \dots, \sup_{s \in K_d} Z(s) \leqslant z_d \right) \ , \tag{7.3}$$

pour toute collection de compacts $K_1, \dots, K_d \subset \mathcal{S}$ (Giné *et al.*, 1990).

En particulier, si $K_j = \{s_j\}$ pour tout $j \in \{1, \dots, d\}$,

$$\begin{aligned} \Lambda\left[\bigcup_{j=1,\dots,d} \Big\{ f \in \mathcal{C}(\mathcal{S}) \ : \ f(s_j) \geqslant z_j \Big\}\right] &= -\log \Pr\big(Z(s_1) \leqslant z_1, \dots, Z(s_d) \leqslant z_d\big) \\ &= V(z_1, \dots, z_d) \ , \end{aligned}$$

où V est la *fonction exposante* du processus Z évaluée sur l'ensemble de sites $\{s_1, \dots, s_d\}$.

Par la suite, les dépassements de seuil sont définis selon une fonctionnelle $\ell : \mathcal{C}_+(\mathcal{S}) \mapsto \mathbb{R}_+$ continue et homogène d'ordre 1 : les ℓ-excès sont les réalisations de X_P telles que $\ell(X_P)$ dépasse un seuil élevé fixé. Dombry et Ribatet (2015) montrent que le processus limite des ℓ-excès de X_P, s'il existe, est un processus ℓ-Pareto, noté Y_ℓ et caractérisé par la formulation spectrale :

$$Y_\ell(s) = R W_\ell(s) \ , \quad W_\ell \sim H_\ell \ ,$$

où R suit une loi de Pareto unitaire indépendamment de W_ℓ, et H_ℓ est la mesure spectrale définie sur la sphère unité $S_\ell = \big\{ f \in \mathcal{C}_+(\mathcal{S}) \ : \ \ell(f) = 1 \big\}$.

Grâce aux résultats de la théorie des valeurs extrêmes décrits dans la section 1.1, on peut relier les approches de dépassements de seuil et de maxima par blocs. On a donc le lien suivant entre la structure de dépendance de Y_ℓ et celle du processus max-stable simple Z :

$$\Lambda(df) = \kappa_\ell(\mathcal{S}) r^{-2} dr H_\ell(dW_\ell) \ ,$$

où $\kappa_\ell(\mathcal{S}) = \Lambda\big(\{ f \in \mathcal{C}(\mathcal{S}) \ : \ \ell(f) \geqslant 1 \}\big) > 0$.

Thibaud et Opitz (2015) montrent qu'on obtient ainsi la limite :

$$\Pr\big(X_P \in uB \mid \ell(X_P) \geqslant u\big) \xrightarrow[u \to \infty]{} \frac{\Lambda(B)}{\kappa_\ell(\mathcal{S})} \ , \tag{7.4}$$

où $B \subset \{ f \in \mathcal{C}_+(\mathcal{S}) \ : \ \ell(f) \geqslant 1 \}$.

7.3 Autres approches

7.3.1 Incréments extrémaux

Engelke *et al.* (2012) proposent une méthode qui peut être vue comme un cas particulier des processus ℓ-Pareto de Dombry et Ribatet (2015), avec le choix de fonctionnelle $\ell(X) = X(s_0)$. En supposant que le processus X appartienne au domaine d'attraction d'un processus max-stable Y, il est montré que les excès de seuil du vecteur multivarié observé sur un ensemble fini de stations $\{s_1, \ldots, s_d\}$, conditionnés par un excès en un site s_0, convergent en loi vers le processus spectral W apparaissant dans la construction des modèles max-stables proposés par de Haan (1984).

Plus précisément, le processus max-stable simple Z, obtenu par transformation de Y pour être à loi marginales Fréchet unitaires, peut s'écrire sous la forme :

$$Z(s) = \max_{i \geqslant 1} \frac{W_i(s)}{\zeta_i} \ , \tag{7.5}$$

où les $\{\zeta_i\}_{i\geqslant 1}$ sont les points d'un processus de Poisson unitaire sur $\mathbb{R}_+$ et les $\{W_i(s)\}_{i\geqslant 1}$ sont des répliques d'un processus spatial positif stationnaire W vérifiant $\sup_{s\in\mathcal{S}} W_i(s) < \infty$ et $\mathbb{E}\big[W_i(s)\big] = 1$ (cf. Chapitre 5).

Soit une suite $a(n) > 0$ telle que $a(n) \xrightarrow[n\to\infty]{} \infty$, Engelke *et al.* (2012) montrent la convergence des *incréments extrémaux* :

$$\left(\frac{X(s_1)}{X(s_0)}, \ldots, \frac{X(s_d)}{X(s_0)} \ \middle| \ X(s_0) > a(n)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \big(W(s_1), \ldots, W(s_d)\big) \ ,$$

où W est le processus spectral associé à Z.

La méthode proposée possède donc des points communs avec le modèle semi-paramétrique conditionnel de Heffernan et Tawn (2004) (voir aussi Heffernan et Resnick (2007)), puisque les excès de seuil sont conditionnés par les dépassements de la composante $X(s_0)$.

Engelke *et al.* (2015) utilisent ce résultat pour mettre en place une méthode d'inférence paramétrique basée sur le modèle de Brown-Resnick (Brown et Resnick, 1977; Kabluchko *et al.*, 2009).

7.3.2 Processus de Poisson

Les résultats de Coles et Tawn (1991) et de Smith *et al.* (1997) sur l'approche des valeurs extrêmes par les processus de Poisson sont adaptés au cadre spatial par Wadsworth et Tawn (2014), qui construisent une méthode d'inférence avec une vraisemblance censurée en utilisant le modèle de Brown-Resnick.

D'après la formulation spectrale (7.5), le processus $\{R_i(s)\}_{i\geqslant 1} = \{W_i(s)/\zeta_i\}_{i\geqslant 1}$ forme un processus de Poisson dont la mesure d'intensité moyenne est donnée par la fonction exposante V associée au processus max-stable Z.

Si on observe N_{u_R} réalisations de $R_i = \{R_i(s_1), \ldots, R_i(s_d)\}$ dans la région $[0, u_R]^c$ pour un seuil $u_R \in \mathbb{R}_+^d$ élevé, la vraisemblance du processus de Poisson est donnée par :

$$f_R(r) = \exp\big\{-V(u_R)\big\} \prod_{i=1}^{N_{u_R}} \big\{-\partial_{1:d} V(r_i)\big\} \ ,$$

où $\partial_{1:d} V$ correspond à la dérivée partielle de la fonction exposante V selon les indices $\{1, \ldots, d\}$.

En pratique cependant, on observe $\{X_i\}_{i\geqslant 1} = \{X_i(s_1), \ldots, X_i(s_d)\}_{i\geqslant 1}$. On peut supposer que les réalisations de X sont de marges Pareto standard, ou opérer des transformations marginales pour obtenir le processus X_P correspondant (cf. équations (1.5) dans le chapitre 1).

La théorie des valeurs extrêmes (de Haan et Resnick, 1977) montre alors que le processus $\big\{X_i/n\big\}_{i=1,\ldots,n}$ converge, quand $n \to \infty$, vers un processus de Poisson dont la mesure d'intensité moyenne est la fonction exposante V. Pour n assez grand, il est alors raisonnable d'approcher les vecteurs aléatoires $X_i(s)$ par les vecteurs $nR_i(s)$, et d'exprimer la vraisemblance de X_i grâce à celle de R_i. En notant $u = nu_R$, la vraisemblance associée aux observations de X est approximée par :

$$f(x) = \exp\big\{-nV(u)\big\} \prod_{i=1}^{N_u} \big\{-\partial_{1:d} V(x_i)\big\} \ .$$

Cependant, l'approximation du processus $\{X_i/n\}$ vers $\{R_i\}$ n'est pas raisonnable si toutes les composantes de X ne sont pas extrêmes. Pour résoudre ce problème, Wadsworth et Tawn (2014) utilisent une procédure de

censure des éléments en-dessous du seuil u, en considérant la vraisemblance censurée :

$$f_{\text{WT2014}}(x) = \exp\big\{-nV(u)\big\} \prod_{i=1}^{N_u} \big\{-\partial_{\pi_i} V\left(x_i^C\right)\big\} \ ,$$

où $x_i^C = \big(\max(x_{1,i}, u), \ldots, \max(x_{d,i}, u)\big)$ et π_i est l'ensemble d'indices de $\{1, \ldots, d\}$ tels que la composante x_i dépasse le seuil u.

Wadsworth et Tawn (2014) utilisent le modèle max-stable de Brown-Resnick et calculent la dérivée partielle $\partial_{\pi_i} V$ de la fonction exposante de ce modèle pour tout $\pi_i \subset \{1, \ldots, d\}$. Ces résultats servent à dériver un estimateur de la probabilité d'échec conditionnelle et sont présentés dans la section 8.2.2.

L'approche de Wadsworth et Tawn (2014) peut être vue comme une amélioration de la méthode de Stephenson et Tawn (2005), qui utilise l'information du temps d'occurrence du maximum annuel dans l'inférence sur les processus max-stables.

En effet, soit $\mathcal{P}$ l'ensemble des partitions de $\{1, \ldots, d\}$ et $\Pi \in \mathcal{P}$ la partition indiquant les éléments du vecteur $Y = (\vee_{i=1}^n X_i(s_1), \ldots, \vee_{i=1}^n X_i(s_d))$ qui apparaissent en même temps. Stephenson et Tawn (2005) montrent que la vraisemblance associée aux observations y de Y avec la partition Π est :

$$f_{\text{ST2005}}(y) = \exp\big\{-V(y)\big\} \prod_{j=1}^{|\Pi|} \big\{-\partial_{\pi_j} V(y)\big\} \ ,$$

où $\pi_j \in \Pi$ et $|\Pi|$ est le nombre d'éléments de Π, autrement dit le nombre de temps d'occurrences disjoints du maximum Y.

Les vraisemblances f_{WT2014} et f_{ST2005} sont similaires : la première considère toutes les valeurs du processus journalier X qui dépassent un seuil multivarié $u \in \mathbb{R}_+^d$ pour au moins une composante X_j, $j \in \{1, \ldots, d\}$, tandis que la seconde se concentre sur les valeurs maximales de chaque composante. Le modèle de Wadsworth et Tawn (2014) augmente ainsi le nombre de données considérées par le modèle de Stephenson et Tawn (2005).

Asadi *et al.* (2015) appliquent l'approche par processus de Poisson décrite précédemment à des données de débits de rivière, en définissant également une distance dite *hydrologique*. Cette mesure est construite à partir de la distance le long d'une rivière et la distance euclidienne dans l'espace géographique pour prendre en compte l'effet des précipitations sur les débits.

7.3.3 Modélisation de la mesure angulaire

Cette section traite de la modélisation de la mesure angulaire par Boldi et Davison (2007) et l'adaptation de ce modèle par Sabourin et Naveau (2014) et Sabourin (2015). Ce dernier article a servi de référence principale pour cette section.

Définition de la mesure angulaire

Soient $X = \{X(s_1), \ldots, X(s_d)\}$ le vecteur journalier observé, de fonction de répartition jointe F (et de marges $F_1, \ldots, F_d$), et X_{Fr} la transformation de X en marges Fréchet unitaires par la relation :

$$X_{\text{Fr}} = \big(X_{\text{Fr}}(s_1), \ldots, X_{\text{Fr}}(s_d)\big) = \left(-\frac{1}{\log F_1(X(s_1))}, \ldots, -\frac{1}{\log F_d(X(s_d))}\right) \ .$$

Le vecteur X est représenté en coordonnées pseudo-polaires (R, W) par la construction :

$$R = \sum_{j=1}^{d} X_{\text{Fr}}(s_j) \ \ \text{(rayon)} \ , \quad W = \frac{1}{R} X_{\text{Fr}} \in \mathcal{S}_d \ \ \text{(angle)} \ ,$$

où $\mathcal{S}_d = \{x \in \mathbb{R}_+^d \ : \ \sum_{j=1}^d x_j = 1\}$ est le simplexe unitaire de $\mathbb{R}_+^d$.

La théorie des valeurs extrêmes (cf. Chapitre 1) montre que si F appartient au domaine d'attraction d'une loi de valeurs extrêmes G, alors pour un rayon R dépassant une valeur r_0 élevée, un modèle asymptotique approprié pour les variables (R, W) est :

$$\Pr\big(R > r, W \in A \mid R > r_0\big) = \frac{r}{r_0} H(A) \ \ , \ r > r_0 \ \ , \ A \subset \mathcal{S} \ ,$$

où H est la mesure angulaire de X_{Fr}.

La mesure angulaire H est liée à la fonction exposante V du processus max-stable simple Z et vérifie la contrainte des moments :

$$\int_{\mathcal{S}_d} \omega_j dH(\omega) = \frac{1}{d} \ , \ j = 1, \ldots, d \ . \tag{7.6}$$

Il n'y a pas de forme paramétrique universelle pour cette mesure, à l'instar de la fonction exposante V, mais on peut cependant l'approcher par un modèle paramétrique tel que celui de Boldi et Davison (2007), qui est présenté ci-dessous.

Mélanges de Dirichlet

Un mélange de Dirichlet est proposé par Boldi et Davison (2007) pour modéliser la mesure angulaire H. Ce modèle est par la suite reparamétré par Sabourin et Naveau (2014) pour permettre une meilleure implémentation par une approche bayésienne.

La loi de Dirichlet est la généralisation de la loi Beta en dimension $d > 1$. Elle est déterminée par deux paramètres : $\nu \in \mathbb{R}_+$ et $\mu \in \mathcal{S}_d$ et leur fonction de densité s'écrit :

$$\mathrm{Dir}(x; \nu, \mu) = \frac{\Gamma(\nu)}{\prod_{j=1}^{d} \Gamma(\nu \mu_i)} \prod_{i=1}^{d} x_j^{\nu \mu_i - 1} \ ,$$

où $\Gamma(\cdot)$ est la fonction gamma.

Boldi et Davison (2007) modélisent la densité de la mesure angulaire par le mélange de M lois de Dirichlet :

$$h(\omega) = \sum_{m=1}^{M} p_m \mathrm{Dir}(\omega; \nu_m, \mu_m) \ ,$$

où les poids $p_m > 0$ du mélange de lois de Dirichlet satisfont la contrainte des moments (7.6) de H :

$$\sum_{m=1}^{M} p_m \mu_m = \big(1/d, \ldots, 1/d\big) \ .$$

Coordonnées cartésiennes et processus de Poisson

De façon analogue au modèle de Wadsworth et Tawn (2014), Sabourin (2015) propose de revenir en coordonnées cartésiennes et de regarder la convergence des observations $\{X_{P,i}/n\}_{i=1,\ldots,n}$, où X_P est de marge Pareto standard. Elle se réalise vers le processus de Poisson d'intensité V, la fonction exposante du processus max-stable limite. La fonction V est liée à la mesure angulaire H par la relation $\partial V(r, w) = \frac{d}{r^2} \partial r \partial H(w)$. En coordonnées cartésiennes, cette relation devient :

$$\frac{\partial V}{\partial x}(x) = d r^{-(d+1)} h(w) \ .$$

Cette formule est utilisée avec le modèle de mélanges de Dirichlet de Boldi et Davison (2007) pour estimer la mesure angulaire H et la fonction exposante V sur des données de débits multivariés.

Sabourin (2015) argumente ce retour aux coordonnées cartésiennes par trois raisons :

1. Les excès pseudo-polaires sont exprimés pour la variable transformée en marges Fréchet X_{Fr}. Un ensemble d'échec correspondant à cette variable devient alors difficile à interpréter pour les observations X.
2. Sabourin (2015) traite de certaines données de débits historiques censurées et représentées par des rectangles. Le modèle pseudo-polaire n'est pas adapté pour ce type de données, contrairement à un modèle défini en coordonnées cartésiennes.
3. Le modèle n'est plus valable près des axes car plusieurs éléments du vecteur observé ne sont alors plus extrêmes. Ce point est traité en coordonnées cartésiennes en se servant d'un modèle de censure similaire à Wadsworth et Tawn (2014) et Thibaud et Opitz (2015).

7.4 Vraisemblance censurée pour les dépassements de seuil

Asymptotiquement, les maxima composante par composante d'un vecteur multivarié X suit une loi MEV, dont la fonction de répartition est décrite entre autres par la fonction exposante V. Si les valeurs maximales annuelles (ou par bloc) composante par composante peuvent être approchées par cette loi, les choses sont différentes si l'on regarde les excès de seuil de la série journalière.

En effet, si les dépassements de seuil sont définis par les événements :

$$\left\{ \bigvee_{j=1}^{d} \frac{X_j}{u_j} > 1 \right\} ,$$

toutes les composantes du vecteur observé ne le sont pas nécessairement : l'approximation du vecteur X par une loi de valeurs extrêmes n'est alors plus justifié. C'est pourquoi la modélisation des excès de seuil multivariés requiert généralement l'utilisation d'une vraisemblance dite *censurée* (Ledford et Tawn, 1996).

Pour un vecteur bivarié (X_1, X_2) de loi Pareto standard et un seuil $u > 0$ suffisamment élevé, la vraisemblance censurée est donnée par :

$$f_{\text{cens},u}(x_1, x_2) = \begin{cases} \frac{\partial^2}{\partial x_1 \partial x_2} \exp\big(-V(x_1, x_2)\big) & \text{si} \quad x_1, x_2 > u \ , \\ \frac{\partial}{\partial x_1} \exp\big(-V(x_1, u)\big) & \text{si} \quad x_1 > u, x_2 \leqslant u \ , \\ \frac{\partial}{\partial x_2} \exp\big(-V(u, x_2)\big) & \text{si} \quad x_1 \leqslant u, x_2 > u \ , \\ \exp\big(-V(u, u)\big) & \text{si} \quad x_1, x_2 \leqslant u \ . \end{cases}$$

Autrement dit, on censure à gauche les composantes du vecteur X qui ne dépassent pas le seuil u.

Jeon et Smith (2012) utilisent la méthode d'inférence par maximum de vraisemblance composite (par paires) pour les modèles de fonction exposante V correspondant aux processus max-stables de Smith (Smith, 1990), de Schlather (Schlather, 2002) et de Brown-Resnick (Brown et Resnick, 1977; Kabluchko *et al.*, 2009), présentés pour la plupart dans le chapitre 5.

Une approche similaire est choisie par Huser et Davison (2014) pour le modèle max-stable de Davison et Gholamrezaee (2012) qui peut être vu comme une généralisation du modèle de Schlather (2002). Dans le même temps, les auteurs proposent d'étendre ce processus max-stable au cadre spatio-temporel, en utilisant des processus spectraux gaussiens ayant une fonction de corrélation adaptée aux processus spatio-temporels (Gneiting *et al.*, 2006; Gneiting, 2002).

Cette méthode est appliquée pour les processus ℓ-Pareto par Thibaud et Opitz (2015) (voir section 8.2.1), ou encore par Sabourin (2015) et Wadsworth et Tawn (2014) pour la convergence du processus $\{X_i/n\}_{i=1,\ldots,n}$, avec la vraisemblance censurée multivariée en dérivant la mesure exposante V selon les indices de $\{1, \ldots, d\}$ qui voient l'observation x_j dépasser le seuil.

7.5 Modèle à variables latentes

Une dernière approche que l'on peut citer sur les dépassements de seuil spatiaux est celle de Cooley *et al.* (2007). Dans cet article, les auteurs ont pour objectif la cartographie du niveau de retour centennal de précipitations et définissent un modèle hiérarchique bayésien qui s'apparente au modèle à variables latentes de Davison *et al.* (2012), présenté dans le chapitre 5.

Les réalisations $\{X_1, \ldots, X_n\}$ du processus spatial $X \in \mathcal{C}(\mathcal{S})$ observé sur l'ensemble de positions $\{s_1, \ldots, s_d\}$ sont supposées indépendantes en conditionnant par la valeur des paramètres $\tau \in \mathcal{C}_+(\mathcal{S})$ et $\xi \in \mathcal{C}(\mathcal{S})$ des lois GP marginales associées aux excès de seuil $u \in \mathcal{C}_+(\mathcal{S})$.

La queue de distribution est décrite, pour $u \in \mathbb{R}$ suffisamment grand :

$$\Pr\big(X(s_j) - u > x \mid X(s_j) > u\big) \approx \left(1 + \frac{\xi(s_j)x}{\exp \phi(s_j)}\right)^{-1/\xi(s_j)} ,$$

où $\phi(s) = \log \tau(s)$ pour tout $s \in \mathcal{S}$.

Dans le modèle de Cooley *et al.* (2007), les processus latents ϕ et ξ sont construits de la façon suivante :

- ϕ est un processus gaussien stationnaire dont la moyenne μ_ϕ est déterminée par un ensemble de covariables associées à la position s et dont la fonction de covariance ρ_ϕ est de forme exponentielle.
- Trois modèles spatiaux sont proposés pour le paramètre de forme ξ, augmentant en complexité :
  1. $\xi \equiv \xi_0$, une valeur constante pour toute la région étudiée,
  2. $\xi(s) = \xi_{\text{M}} \mathbb{1}_{\text{M}}(s) + \xi_{\text{P}} \mathbb{1}_{\text{P}}(s)$: le paramètre de forme prend la valeur ξ_{M} si la position est en montagne (M) et la valeur ξ_{P} si la position est en plaine (P),
  3. ξ est un processus gaussien construit de la même manière que le paramètre d'échelle ϕ.

La modélisation spatiale des paramètres marginaux $\exp(\phi)$ et ξ permet ainsi l'extrapolation spatiale de la loi des valeurs extrêmes et donc du niveau de retour centennal en toute position s dont on connaît les covariables.

Cooley *et al.* (2007) choisissent quatre covariables spatiales pour décrire les processus latents : les coordonnées géographiques (longitude et latitude), l'altitude et les pluies saisonnières moyennes (*MSP* pour *Mean Seasonal Precipitation*). Afin d'extrapoler spatialement le niveau de retour centennal, les auteurs utilisent un atlas de précipitations pour connaître la valeur de la covariable MSP en un point s donné.

Le phénomène est aussi modélisé dans l'*espace climatique*, dont les dimensions sont formées par les covariables altitude et MSP. Ce changement d'espace d'étude conduit à une légère amélioration de la qualité d'ajustement en utilisant la modélisation (2) pour le paramètre de forme : c'est à dire en l'autorisant à prendre deux valeurs possibles selon que la position est située en plaine ou en montagne.

Stephenson et Tawn (2005) suggèrent de choisir une loi a priori jointe avec une corrélation négative pour les paramètres (τ, ξ) de la loi GPD ajustée aux excès de seuil. Cooley *et al.* (2007) choisissent malgré tout deux lois a priori indépendantes, en argumentant qu'une construction dépendante est assez difficile, mais obtiennent toutefois sur des données réelles des lois a posteriori dépendantes avec une corrélation négative.

Chapitre 8

Probabilité d'échec conditionnelle

Supposons qu'un jour donné, un événement extrême de précipitations touche une ou plusieurs stations d'un réseau météorologique. On souhaite alors évaluer le risque qu'un autre endroit soit affecté par ce type d'événement. On s'intéresse à une *probabilité d'échec conditionnelle* : la probabilité qu'un événement extrême apparaisse sur un ensemble de *cibles* un jour donné, sachant qu'il est observé sur un ensemble de *stations météorologiques* le même jour.

Afin d'évaluer cette probabilité, l'aspect temporel présent dans les observations est dans un premier temps ignoré en considérant que les observations sont des répliques iid d'un processus spatial. La prise en compte de la dépendance temporelle est traitée dans un second temps à la section 8.4.

Ce chapitre utilise les mêmes notations que le précédent : $\mathcal{S}$ désigne un sous-ensemble de $\mathbb{R}^p$, $\mathcal{C}(\mathcal{S})$ l'ensemble des fonctions continues pour la norme uniforme sur $\mathcal{S}$ et $\mathcal{C}_+(\mathcal{S})$ la restriction de $\mathcal{C}(\mathcal{S})$ aux fonctions positives.

On se donne aussi deux ensembles finis A_C et A_S de positions dans $\mathcal{S}$ (A_C pour les "cibles" et A_S pour les "stations"). Soit $X \in \mathcal{C}(\mathcal{S})$ un processus aléatoire à valeurs réelles que l'on suppose stationnaire. La probabilité que le processus X dépasse une valeur élevée $u \in \mathbb{R}$ sur au moins une des cibles A_C, sachant qu'il la dépasse sur au moins une des stations A_S s'écrit :

$$p_{A_C|A_S}(u) := \Pr\left(\max_{s\in A_C} X(s) \geqslant u \;\middle|\; \max_{s\in A_S} X(s) \geqslant u\right) . \tag{8.1}$$

Dans un premier temps, la section 8.1 utilise les résultats théoriques obtenus sur les processus ℓ-Pareto par Dombry et Ribatet (2015) et Thibaud et Opitz (2015) afin de proposer une approximation de la probabilité d'échec conditionnelle (8.1). Ensuite, dans la section 8.2, la procédure inférentielle mise en place par Thibaud et Opitz (2015) est utilisée pour construire des estimateurs de $p_{A_C|A_S}(u)$ dans un cadre paramétrique. Dans un troisième temps, deux estimateurs non paramétriques sont proposés dans la section 8.3 : un premier utilisant l'approximation de la section 8.1 et un second faisant intervenir le modèle conditionnel de Heffernan et Tawn (2004).

Finalement, la section 8.4 présente les méthodes tenant compte de la dépendance temporelle inhérente au processus de précipitations journalières. Les méthodes de *declustering* construites pour une série stationnaire univariée sont recensées et une approche adaptée aux dépassements de seuil multivariés et spatiaux est présentée pour être ensuite testée sur simulations dans le chapitre 9.

8.1 Approximation de la probabilité d'échec

8.1.1 Formulation alternative de la probabilité d'échec

Pour tout $A \subset \mathcal{S}$, on note l'événement

$$E(A) = \left\{\sup_{s\in A} X(s) > u\right\} .$$

En utilisant la formule d'inclusion-exclusion et en remarquant que $E(A) \subset E(\mathcal{S})$ et que $E(A)\cup E(B) = E(A\cup B)$ pour tous $A, B \subset \mathcal{S}$, la probabilité d'échec conditionnelle (8.1) s'écrit :

$$p_{A_C|A_S}(u) = \frac{p_{A_C|\mathcal{S}}(u) + p_{A_S|\mathcal{S}}(u) - p_{A_C\cup A_S|\mathcal{S}}(u)}{p_{A_S|\mathcal{S}}(u)} . \tag{8.2}$$

Ainsi, si pour tout ensemble $A \subset \mathcal{S}$, il est possible d'approcher la probabilité conditionnelle $p_{A|\mathcal{S}}(u)$, alors on peut obtenir une approximation de la probabilité d'échec conditionnelle (8.1) grâce à la formule (8.2).
La section suivante fournit une réponse à ce problème en utilisant les processus ℓ-Pareto (Ferreira et de Haan, 2014; Dombry et Ribatet, 2015), présentés dans le chapitre précédent.

8.1.2 Utilisation d'un processus Pareto

Soient $X \in \mathcal{C}(\mathcal{S})$ le processus d'intérêt et $X_P \in \mathcal{C}_+(\mathcal{S})$ le processus transformé en marges Pareto standard. Il est facile de constater que pour tout ensemble $A \subset \mathcal{S}$, la probabilité conditionnelle $p_{A|\mathcal{S}}(u)$ s'écrit :

$$p_{A|\mathcal{S}}(u) = \Pr\left(\sup_{s\in A} X_P(s) \geqslant u_P \;\middle|\; \sup_{s\in\mathcal{S}} X_P(s) \geqslant u_P\right) , \tag{8.3}$$

où $u_P \in \mathbb{R}_+$ est obtenu par la transformation $u_P = 1/\big(1 - F_{X(s)}(u)\big)$, $s \in \mathcal{S}$.

Grâce aux résultats de Thibaud et Opitz (2015), décrits dans la section 7.2.3, la probabilité $p_{A|\mathcal{S}}(u)$ peut ainsi être approchée par une formule faisant intervenir la fonction exposante V du processus max-stable limite Z. La proposition 1 suivante fournit cette approximation.

Proposition 1. *Soient $X_P \in \mathcal{C}_+(\mathcal{S})$ un processus de marges Pareto standard et $A = \{s_1, \dots, s_d\} \subset \mathcal{S}$ un ensemble fini de positions. On a :*

$$\begin{aligned} p_{A|\mathcal{S}}(u) \quad &= \quad \Pr\left(\sup_{s\in A} X_P(s) \geqslant u_P \;\middle|\; \sup_{s\in\mathcal{S}} X_P(s) \geqslant u_P\right) \\ &\xrightarrow[u_P\to\infty]{} \quad \frac{V_A(\mathbb{1})}{\kappa(\mathcal{S})} , \end{aligned}$$

où $V_A(\mathbb{1}) = V_{(s_1,\dots,s_d)}(1,\dots,1)$ et $\kappa(\mathcal{S}) = \Lambda\big(\{f \in \mathcal{C}(\mathcal{S}) \;:\; \sup_{s\in\mathcal{S}} f(s) \geqslant 1\}\big) > 0$.

Démonstration. Soit $\ell \;:\; \mathcal{C}_+(\mathcal{S}) \to \mathbb{R}_+$ la fonctionnelle (continue et homogène) définie par $\ell(f) = \sup_{s\in\mathcal{S}} f(s)$. La probabilité conditionnelle $p_{A|\mathcal{S}}(u)$ s'écrit :

$$p_{A|\mathcal{S}}(u) = \Pr\left(\sup_{s\in A} X_P(s) > u_P \;\middle|\; \ell(X_P) > u_P\right) .$$

De plus, on a :

$$\begin{aligned} \sup_{s\in A} X_P \geqslant u_P \quad &\iff \quad X_P \in \left\{f \in \mathcal{C}_+(\mathcal{S}) \;:\; \sup_{s\in A} f(s) \geqslant u_P\right\} \\ &\iff \quad u_P^{-1} X_P \in \left\{f \in \mathcal{C}_+(\mathcal{S}) \;:\; \sup_{s\in A} f(s) \geqslant 1\right\} =: B_A \\ &\iff \quad X_P \in u_P B_A . \end{aligned}$$

Comme $A \subset \mathcal{S}$, l'ensemble B_A vérifie :

$$B_A \subset \big\{f \in \mathcal{C}_+(\mathcal{S}) \;:\; \ell(f) \geqslant 1\big\} .$$

Ainsi, en utilisant la limite (7.4), on obtient donc :

$$\begin{aligned} p_{A|\mathcal{S}}(u) \quad &= \quad \Pr\big(X_P \in u_P B_A \mid \ell(X_P) \geqslant u_P\big) \\ &\xrightarrow[u_P\to\infty]{} \quad \frac{\Lambda(B_A)}{\kappa_\ell(\mathcal{S})} = \frac{V_A(\mathbb{1})}{\kappa(\mathcal{S})} , \end{aligned}$$

où $\kappa(\mathcal{S}) = \kappa_\ell(\mathcal{S}) = \Lambda\big(\{f \in \mathcal{C}(\mathcal{S}) \;:\; \sup_{s\in\mathcal{S}} f(s) \geqslant 1\}\big) > 0$. □

Grâce à la formulation alternative (8.2) de la probabilité d'échec conditionnelle (8.1), on en déduit, pour u suffisamment élevé, l'approximation :

$$p_{A_C|A_S}(u) \approx \frac{\theta_{A_C} + \theta_{A_S} - \theta_{A_C\cup A_S}}{\theta_{A_S}} , \tag{8.4}$$

où $\theta_A = V_A(\mathbb{1})$ désigne le coefficient extrémal de Z (Schlather et Tawn, 2003) évalué aux positions de A pour tout ensemble $A \subset \mathcal{S}$.

Un estimateur de la probabilité d'échec conditionnelle (8.1) est donc le suivant :

$$\widehat{p}_{A_C|A_S}(u) := \frac{\widehat{\theta}_{A_C} + \widehat{\theta}_{A_S} - \widehat{\theta}_{A_C\cup A_S}}{\widehat{\theta}_{A_S}} , \tag{8.5}$$

8.2 Mise en place d'estimateurs via des méthodes paramétriques

La formule (8.5) permet d'estimer la probabilité d'échec conditionnelle (8.1) dès lors que l'on sait estimer la fonction exposante V du processus max-stable Z. Plusieurs modèles paramétriques existent pour V, ce qui permet de construire autant d'estimateurs de la probabilité d'échec conditionnelle.

La section 8.2.1 présente l'inférence mise en place par Thibaud et Opitz (2015), basée sur les ℓ-excès du processus journalier X transformé en marges Pareto standard. Cette méthode conduit à une estimation des paramètres de V pour un modèle max-stable choisi. Les sections 8.2.2 à 8.2.5 détaillent ensuite les calculs liés à cette méthode pour quatre modèles max-stables.

8.2.1 Inférence sur les processus Pareto

Cette section traite de la méthode inférentielle basée sur les observations d'un processus ℓ-Pareto mise en place par Thibaud et Opitz (2015) et suit donc les indications du paragraphe 3.1 de cet article.

Soit $X \in \mathcal{C}(\mathcal{S})$ un processus admettant un attracteur max-stable. On suppose qu'on observe indépendamment n fois le processus X sur un ensemble de sites $A_S = \{s_1, \dots, s_d\}$. Soit $u_{\mathrm{GP}} = \big(u_{\mathrm{GP}}(s_1), \dots, u_{\mathrm{GP}}(s_d)\big) \in \mathbb{R}^d$ un seuil assez grand et $X_P \in \mathcal{C}_+(\mathcal{S})$ le processus transformé en marges Pareto standard par la relation (1.5). L'approximation des queues de distributions marginales de X par des lois GP permet d'écrire :

$$X_P(s) \approx \frac{1}{1 - \widehat{F}_{X(s)}\big(X(s)\big)} \ ,$$

avec

$$\widehat{F}_{X(s)}(x) = \begin{cases} \widehat{F}_{n,X(s)}(x) & \text{si} \quad x \leqslant u_{\mathrm{GP}}(s) \\ 1 - \widehat{\zeta}_{u_{\mathrm{GP}}(s)} \left(1 + \widehat{\xi}(s) \frac{x - u_{\mathrm{GP}}(s)}{\widehat{\tau}(s)}\right)_+^{-1/\widehat{\xi}(s)} & \text{si} \quad x > u_{\mathrm{GP}}(s) \ , \end{cases}$$

où $\widehat{F}_{n,X(s)}(x)$ est la fonction de répartition empirique de $X(s)$, $\widehat{\zeta}_{u_{\mathrm{GP}}(s)} = \widehat{\Pr}\big(X(s) > u_{\mathrm{GP}}(s)\big)$ et $\big(\widehat{\tau}(s), \widehat{\xi}(s)\big)$ sont les estimations des paramètres de la loi GP.

Une fonctionnelle $\ell \in \mathcal{C}_+(\mathcal{S})$ est ensuite choisie et les ℓ-excès de X_P sont approchés par le processus ℓ-Pareto Y_ℓ. Sans perte de généralité, par homogénéité d'ordre 1 de la fonctionnelle ℓ, les ℓ-excès peuvent être définis par l'événement $\big\{\ell(X) \geqslant 1\big\}$.

La fonction de densité du vecteur aléatoire $(Y_\ell(s_1), \dots, Y_\ell(s_d))$ est donnée, d'après les résultats de Thibaud et Opitz (2015) décrits dans la section 8.1, par :

$$\frac{\lambda(y)}{\kappa_\ell(\mathcal{S})} \ , \quad \text{pour tout } y \in \mathbb{R}_+^d \backslash \{0\} \text{ tel que } \ell(y) \geqslant 1 \ ,$$

où λ est la densité de la mesure exposante Λ. Lorsque celle-ci est continue par la mesure de Lebesgue, on a

$$\lambda(y) = -\partial_{1:d} V(y) \ ,$$

l'opposée de la dérivée partielle de la fonction exposante V évaluée sur $A := \{s_1, \dots, s_d\} \subset \mathcal{S}$.

En spécifiant la fonctionnelle $\ell(f) = \max_{j=1,\dots,d} f(s_j)/u_j$ avec un seuil multivarié $u = (u_1, \dots, u_d)$ assez élevé, on obtient $\kappa_\ell(\mathcal{S}) = V(u)$. Si l'on se donne un modèle paramétrique pour la mesure exposante Λ, de paramètres ψ, la vraisemblance $f(x, \psi)$ obtenue sur l'ensemble des N_u ℓ-excès $X^E := \{X_1^E, \dots, X_{N_u}^E\}$ tels que $\ell\big(X_k^E\big) \geqslant 1$ pour tout $k \in \{1, \dots, N_u\}$, s'écrit :

$$f(x; \psi) = \prod_{k=1}^{N_u} \frac{\lambda(x_k^E)}{V(u)} \ . \tag{8.6}$$

Cependant, la vraisemblance (8.6) n'est pas adaptée aux ℓ-excès X^E pour deux raisons :

1. la loi de Y_ℓ n'ajuste pas les composantes de X_k^E qui ne dépassent pas leur seuil marginal,
2. si Λ a une masse positive sur les axes $\mathbb{R}_+^d \backslash \{0\}$ (ce qui est par exemple le cas du modèle Extrémal-t), elle est incompatible avec les lois marginales de X_P supposées Pareto standard et donc toutes strictement positives.

Pour ces raisons, dans la lignée de Wadsworth et Tawn (2014), Thibaud et Opitz (2015) proposent d'utiliser une vraisemblance censurée en considérant les ℓ-excès censurés $X_k^C = \max(X_k^E, u)$, le maximum étant pris composante par composante :

$$f_{\mathrm{cens}}(x; \psi) = \prod_{k=1}^{N_u} \frac{-\partial_{I_k} V(x_k^C)}{V(u)} \ ,$$

où $\partial_{I_k} V$ correspond à la dérivée partielle de V selon les indices I_k des éléments de X_k^C non censurés. Ils suggèrent de plus de rajouter l'information obtenue sur le nombre d'excès N_u qui suit, pour u assez élevé, une loi $\text{Bin}\big(n, V(u)\big)$, où n est le nombre d'observations de X. La vraisemblance construite s'écrit ainsi :

$$f_{\text{TO2015}}(x;\psi) = \big\{1 - V(u)\big\}^{n-N_u} V(u)^{N_u} \prod_{k=1}^{N_u} \frac{-\partial_{I_k} V(x_k^C)}{V(u)} \ . \tag{8.7}$$

Les sections suivantes présentent plusieurs choix possibles pour la fonction exposante V utilisée dans la vraisemblance (8.7).

8.2.2 Utilisation du modèle de Brown-Resnick

Une première possibilité est d'utiliser le modèle de Brown-Resnick (Brown et Resnick, 1977; Kabluchko *et al.*, 2009), de façon analogue à l'article de Wadsworth et Tawn (2014). On rappelle que ce processus est construit à partir des processus spectraux

$$W(s) = \exp\left(Z(s) - \frac{\sigma^2(s)}{2}\right) \ ,$$

où $Z(\cdot)$ est un processus gaussien intrinsèque de variance $\text{Var} Z(s) = \sigma^2(s)$.

Pour estimer une probabilité d'échec conditionnelle avec ce modèle, on a besoin de calculer sa fonction exposante en dimension d ainsi que ses dérivées partielles. Ces calculs ont été réalisés par Wadsworth et Tawn (2014) et sont explicités ci-dessous.

La fonction exposante V du modèle de Brown-Resnick en dimension d s'écrit :

$$V_{(s_1,\ldots,s_d)}(x_1,\ldots,x_d) = \sum_{j=1}^{d} \frac{1}{x_j} \Phi_{d-1}\left\{ \log\left(\frac{x_{-j}}{x_j}\right) + \frac{1}{2}\left(\sigma^2_{-j} + \sigma^2_j - 2\sigma_{-j,j}\right) \ \middle| \ T_j \Sigma T_j' \right\} \ , \tag{8.8}$$

où :

- A' désigne la transposée d'une matrice A afin d'éviter la confusion de notation avec T,
- $\Sigma = [\sigma_{i,j}]_{i,j=1,\ldots,d}$ est la matrice de covariance du vecteur $\big(Z(s_1),\ldots,Z(s_d)\big)$,
- $x_{-j} = (x_i)_{i\neq j}$, $\sigma^2_{-j} = (\sigma^2_i)_{i\neq j}$ et $\sigma_{-j,j} = (\sigma_{i,j})_{i\neq j}$.
- $\Phi_{d-1}\{\cdot|\Sigma\}$ est la fonction de répartition de la loi normale $(d-1)$-multivariée, de moyenne 0 et de matrice de variance-covariance Σ,
- T_j est la matrice de taille $(d-1)\times d$ formée par la combinaison entre la matrice identité $\mathcal{I}_{d-1}$ et une colonne de -1 en j-ième position :

$$T_1 = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \ldots & 1 \end{pmatrix} \quad , \quad T_2 = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -1 & 0 & \ldots & 1 \end{pmatrix} \quad , \quad \ldots$$

Pour simplifier les calculs, on se place dans le cas où le processus Z est un processus stationnaire. Autrement dit, on a $\sigma^2(s_j) = \sigma^2$ pour tout $j \in \{1,\ldots,d\}$. La dérivée partielle de la fonction exposante V selon l'ensemble d'indices $J \subset \{1,\ldots,d\}$ s'écrit :

$$\begin{aligned} -\partial_J V(x_1,\ldots,x_d) = \ & \frac{\Phi_{d-|J|}\big(\log x_{-J} - \mu_J \mid \Gamma_J\big)}{(2\pi)^{(|J|-1)/2} |\Sigma_{J,J}|^{1/2} \left(\mathbb{1}'_{|J|} q_J\right)^{1/2} \prod_{j\in J} x_j} \\ & \times \exp\left\{ -\frac{1}{2}\left[(\log x_J)' A_J (\log x_J) + (\log x_J)' \left(\frac{2q_J}{\mathbb{1}_{|J|} q_J}\right) + \sigma^2 - \frac{1}{\mathbb{1}_{|J|} q_J} \right] \right\} \end{aligned} \ , \tag{8.9}$$

où :

- $|J|$ représente le cardinal de l'ensemble d'indices J,
- $\Phi_0 = 1$,
- $\Sigma_{J,J}$ est la sous-matrice de Σ formée par les indices de J,
- $\mathbb{1}_{|J|}$ est le vecteur colonne de taille $|J|$ formé de 1 pour toutes les composantes,
- $q_J = \Sigma^{-1}_{J,J} \mathbb{1}_{|J|}$ et $A_J = \left(\Sigma^{-1}_{J,J} - \frac{q_J q_J'}{\mathbb{1}'_{|J|} q_J}\right)$,

- $\Gamma_J = K_{01}' A K_{10}$ et $\mu_J = -\Gamma_J \left(K_{01}' A K_{10} \log x_J + \frac{K_{01}' \Sigma^{-1} \mathbb{1}_d}{\mathbb{1}_d' \Sigma^{-1} \mathbb{1}_d} \right)$, où la matrice A est équivalente à A_J pour $J = \{1, \ldots, d\}$ et les matrices K_{01} et K_{10} sont formées à partir des matrices identités $\mathcal{I}_n$ de taille $n \times n$ et de zéros $\mathcal{O}_{n,m}$ de taille $n \times m$:

$$K_{01} = \begin{pmatrix} \mathcal{O}_{|J|,d-|J|} \\ \mathcal{I}_{d-|J|} \end{pmatrix}_{d \times (d-|J|)} \text{ et } K_{10} = \begin{pmatrix} \mathcal{I}_{|J|} \\ \mathcal{O}_{d-|J|,|J|} \end{pmatrix}_{d \times |J|} .$$

Le cas où Z n'est pas stationnaire d'ordre 2 a également été traité par Wadsworth et Tawn (2014).

8.2.3 Utilisation du modèle Extrémal-t

Une deuxième possibilité, choisie par Thibaud et Opitz (2015) est d'utiliser le modèle Extrémal-t (Opitz, 2013). On rappelle que ce modèle est construit par la formulation spectrale suivante :

$$Z(s) = m_\delta \max_{i \geqslant 1} \frac{W_i(s)_+^\delta}{\zeta_i} ,$$

où $\delta \in \mathbb{N} \backslash \{0\}$, où $\{\zeta_i\}_{i \geqslant 1}$ sont les points d'un processus de Poisson d'intensité $\zeta^{-2} d\zeta$ sur $(0, \infty)$, et avec la constante $m_\delta = \sqrt{\pi} 2^{1-\delta/2} \Gamma\left(\frac{\delta+1}{2}\right)^{-1}$ qui assure que les marges de Z sont bien Fréchet unitaires. Dans cette construction, les W_i sont des répliques d'un processus gaussien stationnaire de fonction de corrélation ρ.

Pour ce modèle, la fonction exposante V en dimension d s'écrit (Thibaud et Opitz, 2015) :

$$V_{(s_1,\ldots,s_d)}(x_1, \ldots, x_d) = \sum_{j=1}^{d} \frac{1}{x_j} T_{\delta+1} \left\{ \left(\frac{x_{-j}}{x_j} \right)^{1/\delta} \;\middle|\; R_{-j,j} \, , \, \frac{1}{\delta+1} \left(R_{-j,-j} - R_{-j,j} R_{-j,j}' \right) \right\} , \tag{8.10}$$

où :

- R est la matrice de corrélation : $R_{i,j} = \rho(s_i, s_j)$,
- $x_{-j} = (x_i)_{i \neq j}$, $R_{-j,j} = (R_{i,j})_{i \neq j}$ et $R_{-j,-j}$ est la matrice de taille $(d-1) \times (d-1)$ formée par R après la suppression de la i-ième ligne et de la j-ième colonne,
- et $T_{\delta+1}(\cdot | \mu, R)$ est la fonction de répartition Student multivariée, de degré de liberté $\delta + 1$, de paramètre de décentralisation μ et de matrice de corrélation R.

Pour un ensemble d'indices $J \subset \{1, \ldots, d\}$, la dérivée partielle de la fonction exposante du modèle Extrémal-t est calculée par Thibaud et Opitz (2015). La formule est donnée ci-dessous en utilisant les mêmes notations que pour le modèle de Brown-Resnick (Section 8.2.2) :

$$\begin{aligned} -\partial_J V(x_1, \ldots, x_d) = \; & T_{|J|+\delta} \left(x_{-J}^{1/\delta} \mid \mu_J, \Gamma_J \right) \delta^{1-|J|} \pi^{\frac{1-|J|}{2}} |R_{J,J}|^{-1/2} \Gamma\left(\frac{\delta+1}{2}\right)^{-1} \Gamma\left(\frac{\delta+|J|}{2}\right) \\ & \times \left(\prod_{j \in J} |x_j| \right)^{1/\delta - 1} \left[\left(x_J^{1/\delta} \right)' R_{J,J}^{-1} \left(x_J^{1/\delta} \right) \right]^{-\frac{\delta+|J|}{2}} , \end{aligned} \tag{8.11}$$

avec :

$$\mu_J = R_{-J,J} R_{J,J}^{-1} x_J^{1/\delta}$$

et

$$\Gamma_J = (|J| + \delta)^{-1} \left(x_J^{1/\delta} \right)' R_{J,J}^{-1} \left(x_J^{1/\delta} \right) \left(R_{-J,-J} - R_{-J,J} R_{J,J}^{-1} R_{J,-J} \right) .$$

Dans le cas particulier où $J = \{1, \ldots, d\}$, la formule (8.11) s'écrit :

$$-\partial_J V(x_1, \ldots, x_d) = \delta^{1-d} \pi^{\frac{1-d}{2}} |R|^{-1/2} \Gamma\left(\frac{\delta+1}{2}\right)^{-1} \Gamma\left(\frac{\delta+d}{2}\right) \left(\prod_{j=1}^{d} |x_j| \right)^{1/\delta - 1} \left(\tau_{1/\delta}(x)' R^{-1} \tau_{1/\delta}(x) \right)^{-\frac{\delta+d}{2}} ,$$

où $\tau_{1/\delta}(x) = \left(\text{sign}(x_j) |x_j|^\delta \right)_{j=1,\ldots,d}$ pour $\delta > 0$.

8.2.4 Utilisation du modèle de Reich et Shaby

Une troisième possibilité est d'utiliser le modèle HKEVP de Reich et Shaby (2012) pour décrire la fonction exposante V. On rappelle que ce modèle est introduit dans le chapitre 5 comme l'un des modèles max-stables comparés les plus performants vis-à-vis de l'estimation de la structure de dépendance spatiale (d'un coefficient

extrémal plus précisément). Contrairement aux modèles de Brown-Resnick et Extrémal-t, la fonction exposante V a aussi l'avantage d'être explicite en dimension $d \in \mathbb{N}$:

$$V_{(s_1,\dots,s_d)}(x_1,\dots,x_d) = \sum_{k=1}^{K} \left[\sum_{j=1}^{d} \left(\frac{\Omega_{jk}}{x_j} \right)^{1/\alpha} \right]^{\alpha} , \tag{8.12}$$

où Ω est la matrice $d \times K$ des noyaux : $\Omega_{jk} = \omega_k(s_j)$ avec $k = 1,\dots,K$ et $j = 1,\dots,d$, et où $\alpha \in (0,1]$ est le paramètre de dépendance du modèle.

La dérivée partielle de la fonction exposante par rapport à l'ensemble d'indices $J \subset \{1,\dots,d\}$ s'écrit :

$$-\partial_J V(x_1,\dots,x_d) = (-1)^{|J|} \frac{\alpha(\alpha-1)\dots(\alpha-|J|+1)}{\alpha^{|J|}} \sum_{k=1}^{K} \left(\prod_{j_1 \in J} \frac{\Omega_{j_1 k}^{1/\alpha}}{x_{j_1}^{1/\alpha+1}} \right) \left(\sum_{j_2=1}^{d} \frac{\Omega_{j_2 k}^{1/\alpha}}{x_{j_2}^{1/\alpha}} \right)^{\alpha-|J|} . \tag{8.13}$$

L'utilisation du modèle de Reich et Shaby (2012) permet d'améliorer considérablement le temps d'estimation par rapport aux modèles Extrémal-t ou Brown-Resnick, car il n'est pas nécessaire de calculer des fonctions de répartitions multivariées de lois normales ou Student. Cet aspect est mis en exergue sur des simulations dans le chapitre 9.

8.2.5 Utilisation du modèle logistique de Gumbel

Le modèle logistique de Gumbel (1960) est un modèle multivarié simple qui ne présente pas d'aspect spatial, dont la fonction exposante s'écrit elle aussi de façon explicite en dimension $d \in \mathbb{N}$. Ce modèle est choisi pour regarder les différences obtenues avec celui de Reich et Shaby (2012), version spatiale plus complexe du modèle logistique.

La fonction exposante du modèle logistique s'écrit :

$$V_{(s_1,\dots,s_d)}(x_1,\dots,x_d) = V(x_1,\dots,x_d) = \left(\sum_{j=1}^{d} x_j^{-1/r} \right)^{r} ,$$

où $r \in (0,1]$ est l'unique paramètre de ce modèle. La dérivée partielle de V selon l'ensemble d'indices $J \subset \{1,\dots,d\}$ se calcule facilement :

$$-\partial_J V(x_1,\dots,x_d) = (-1)^{|J|} \left(\prod_{j_1 \in J} \frac{r-j_1}{r} x_{j_1}^{-1/r-1} \right) \left(\sum_{j_2=1}^{d} x_{j_2}^{-1/r} \right)^{r-|J|} .$$

8.3 Estimation non-paramétrique de la probabilité d'échec

La formule (8.4) d'approximation de la probabilité d'échec conditionnelle a permis de construire plusieurs estimateurs à partir de modèles paramétriques max-stables. En utilisant un estimateur non-paramétrique de la fonction exposante V, on peut de la même manière construire un estimateur non-paramétrique de (8.1). La section 8.3.1 propose un premier estimateur non paramétrique naturel, tandis que la section suivante construit une seconde méthode d'estimation basée sur le modèle conditionnel semi-paramétrique de Heffernan et Tawn (2004).

Il est important de noter que les deux estimateurs présentés dans cette partie requièrent des observations à la fois sur l'ensemble de stations S et sur l'ensemble de cibles C, contrairement à ceux construits sur les processus max-stables qui permettent d'extrapoler spatialement la structure de dépendance grâce à la paramétrisation spatiale.

8.3.1 Estimateur non paramétrique simple

La fonction exposante V possède un estimateur empirique naturel, plus connu par son expression du point de vue de la fonction de dépendance caudale stable (*Stable tail dependence function* en anglais) $L(x) = V(x^{-1})$. Voir par exemple (Beirlant *et al.*, 2004, p. 257) et (de Haan et Ferreira, 2006, p. 236).

L'estimateur empirique de la fonction exposante s'exprime sous la forme :

$$\widehat{V}_k(x_1,\dots,x_d) := \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\left\{ R(X_{i,1}) \geqslant n - \frac{k}{x_1} + 1 \text{ ou } \dots \text{ ou } R(X_{i,d}) \geqslant n - \frac{k}{x_d} + 1 \right\}} ,$$

où pour $j \in \{1, \dots, d\}$ fixé, $R(X_{i,j}) = \sum_{m=1}^{n} \mathbb{1}_{\{X_{m,j} \leqslant X_{i,j}\}}$ est le rang de l'observation $X_{i,j}$, pour $i \in \{1, \dots, n\}$, et k est un nombre d'observations choisi en fonction de n. En effet $k \in \{1, \dots, n\}$ doit être assez grand mais petit comparé à n :

$$\begin{cases} k(n) & \xrightarrow[n\to\infty]{} & \infty \ , \\ k(n)/n & \xrightarrow[n\to\infty]{} & 0 \ . \end{cases}$$

On en déduit l'estimateur non-paramétrique de la probabilité d'échec conditionnelle $p_{A_C|A_S}(u)$ en utilisant (8.5) :

$$\widehat{p}_{A_C|A_S,k} := \frac{\widehat{V}_{A_C,k}(\mathbb{1}) + \widehat{V}_{A_S,k}(\mathbb{1}) - \widehat{V}_{A_C \cup A_S,k}(\mathbb{1})}{\widehat{V}_{A_S,k}(\mathbb{1})} \ . \tag{8.14}$$

8.3.2 Estimateur dérivé de Heffernan-Tawn

Le modèle conditionnel de Heffernan et Tawn (2004), décrit dans la section 1.2, peut être utilisé pour calculer la probabilité d'échec conditionnelle (8.1). On rappelle que ce modèle semi-paramétrique regarde le comportement des éléments $X_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_d)$ d'un vecteur X, sachant que l'élément marginal X_j est élevé.

Dans notre cas, on définit les variables aléatoires :

$$X_{A_S} = \max_{s \in A_S} X(s) \text{ et } X_{A_C} = \max_{s \in A_C} X(s) \ ,$$

et on modélise le comportement de $X_{A_C}|X_{A_S} > u$, pour u suffisamment grand, à l'aide du modèle de Heffernan et Tawn (2004).

La procédure d'estimation et de simulation liée à ce modèle, implémentée dans le package R `texmex` (Southworth et Heffernan, 2013) et décrite dans la section 1.2, permet d'obtenir des réalisations

$$\{\tilde{X}_{i,A_C}, \tilde{X}_{i,A_S}\}_{i=1,\dots,N}$$

de (X_{A_C}, X_{A_S}) conditionnées par l'événement $X_{A_S} > u$, où u est un quantile élevé pour la variable X_{A_S}. En utilisant ces simulations, on peut calculer empiriquement la probabilité d'échec conditionnelle par :

$$\widehat{p}_{A_C|A_S}(u) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\tilde{X}_{i,A_C} > u | \tilde{X}_{i,A_S} > u\}} \ .$$

8.4 Prise en compte de la dépendance temporelle

Dans les sections précédentes, les répliques $X_1, \dots, X_n$ du processus spatial X sont supposées indépendantes par souci de simplicité. En effet, les estimateurs de la probabilité d'échec conditionnelle $p_{A_C|A_S}(u)$ ont ainsi pu être construits grâce à cette hypothèse.

En pratique cependant, il n'est en général pas raisonnable de considérer deux observations journalières successives comme indépendantes. Thibaud et Opitz (2015) choisissent d'ignorer la dépendance temporelle éventuellement présente dans les données de pluies qu'ils utilisent, en argumentant par le fait que cet aspect est négligeable et l'est d'autant plus dans les extrêmes.

Dans cette partie, on présente tout d'abord (Section 8.4.1) la méthode dite de *declustering*, puis on propose une sélection des dépassements de seuil (Section 8.4.3) qui sera prise en compte dans la vraisemblance (8.7).

8.4.1 Méthodes de declustering

On se place dans le cas d'une série univariée stationnaire $\{X_t\}_{t \geqslant 1}$, à laquelle on associe un seuil u élevé. Une méthode usuelle permettant de prendre en compte la dépendance temporelle dans les extrêmes est celle du *declustering*. Si l'indice extrémal θ de la série est inférieur à 1, les excès de seuil auront tendance à apparaître par groupe, ou *cluster* (Beirlant *et al.*, 2004, Chapitre 10). Hsing (1987) montre que ces clusters de dépassements, lorsqu'ils sont bien identifiés et que le seuil choisi est suffisamment élevé, peuvent être considérés comme indépendants.

Les méthodes de declustering ont pour but d'identifier ces clusters et de travailler avec des éléments bien choisis afin de se ramener au cas de données indépendantes. Un exemple d'application du declustering pour calculer un niveau de retour est Cai *et al.* (2013).

Deux méthodes sont principalement utilisées dans la littérature : le *run declustering* de Smith (1989) et le *block declustering* de Tawn (1988). Une proposition d'extension du run declustering au cas multidimensionnel est faite par Sabourin et Renard (2015) et une version modifiée du block declustering est donnée par de Haan *et al.* (2014).

Le run declustering

Cette méthode, proposée par Smith (1989), consiste à définir des clusters de dépassements et à sélectionner le maximum de chaque cluster pour obtenir des observations extrêmes indépendantes.

Pour cela, on choisit un temps τ caractéristique à la série qui représente le temps nécessaire pour définir deux clusters indépendants : il s'agit du paramètre de *run*. Le cluster C_k (pour $k \in \{1, \dots, N_c\}$) est initialisé au temps t_k dès lors que $X_{t_k} > u$ et il se termine au temps $t_k + h_k$ où :

$$h_k = \arg\min_{h \in \mathbb{N}} \left\{ X_{t_k+h+h_\tau} \leqslant u \ ; \ \forall h_\tau \in \{1, \dots, \tau\} \right\} .$$

Le maximum de chaque cluster $\{C_1, \dots, C_{N_c}\}$ est sélectionné et d'après (Hsing, 1987, Théorème 4.5), ces maxima $\{\tilde{X}_1, \dots, \tilde{X}_{N_c}\}$ sont asymptotiquement indépendants sous la condition $D(u_n)$ (voir Section 1.1.5).

Extension du run declustering au cas multidimensionnel

Sabourin et Renard (2015) étendent la méthode du run declustering au cas d'observations multivariées stationnaires $\left\{(X_{1,t}, \dots, X_{d,t})\right\}_{t \geqslant 1}$. Leur algorithme d'identification des clusters de dépassements multivariés se décompose comme suit : soient $u = (u_1, \dots, u_d)$ un seuil multivarié et τ un temps caractéristique aux données observées (le paramètre de *run*) :

– Un cluster est initialisé au temps t_i ($i = 1, \dots, N_u$) si *au moins* une des marges de X dépasse son seuil :

$$\exists j \in \{1, \dots, d\} \ \mid \ X_{j,t_i} > u_j \ .$$

– Le cluster se termine au temps $X_{t_i+r_i}$ ($i = 1, \dots, N_u$) lorsque *toutes* les marges de X sont en-dessous du seuil durant un temps τ :

$$X_{j,t_i+r_i+h} \leqslant u_j \ \ , \forall j \in \{1, \dots, d\} \ \text{ et } \ \forall h \in \{1, \dots, \tau\} \ .$$

Leur méthode étant appliquée à des données de débits de rivière, le temps caractéristique τ choisi dans ce cas est de 3 jours. Pour des précipitations, il paraît raisonnable de considérer $\tau = 1$ (de Haan *et al.*, 2014).

Cette méthode pose cependant deux problèmes :

– Si la dimension d du vecteur X est élevée et que la dépendance entre les composantes est faible, la probabilité d'avoir un excès sur au moins une composante augmente, ce qui entraîne la formation de clusters de taille plus grande, mais aussi une réduction de leur nombre. Les seuils marginaux u_j (pour $j \in \{1, \dots, d\}$) doivent être ajustés en fonction de d.

– La vraisemblance est évaluée sur un vecteur qui n'est pas réellement observé puisque l'on prend le maximum pour chaque marge au sein d'un même cluster. Cette remarque est d'autant plus valable si la taille moyenne des clusters est grande.

Le block declustering

La méthode dite du *block declustering* proposée par Tawn (1988) est basée sur le théorème de convergence des r plus grandes statistiques d'ordre (Leadbetter, 1983, Chapitre 2) et sur les résultats de Smith (1986).

On suppose que la série temporelle stationnaire observée est de taille n : $\{X_1, \dots, X_n\}$ et on se donne un temps caractéristique τ qui représente la durée moyenne d'un événement extrême. L'algorithme du block declustering est le suivant :

1. On choisit $\tilde{X}_1 = \max_{i=1,\dots,n} X_i$ d'indice t_1 avec $\tilde{X}_1 > u$ et on construit le nouveau vecteur d'observations $X^{(1)}$ par :
$$X^{(1)} = \left\{ X_t \ : \ |t - t_1| > \tau \right\} .$$
Le vecteur $X^{(1)}$ est ainsi constitué des observations de X auxquelles on a supprimé $\tilde{X}_1$ et les τ observations à gauche et à droite de $\tilde{X}_1$.
2. On recommence la procédure en choisissant la plus grande valeur de $X^{(1)}$, que l'on note $\tilde{X}_2$ et en construisant le vecteur $X^{(2)}$ en supprimant $\tilde{X}_2$ ainsi que les observations distantes de τ par rapport à $\tilde{X}_2$.
3. On s'arrête à l'étape N_u lorsque :
$$\max X^{(N_u)} \leqslant u \ .$$

Cette méthode est appliquée par de Haan *et al.* (2014) sur des données de précipitations journalières en supposant un temps caractéristique $\tau = 1$: autrement dit, les données sont jugées indépendantes si elles sont séparées par au moins une journée.

8.4.2 Autres modèles

D'autres méthodes ont été proposées dans la littérature qui s'écartent des méthodes de declustering et tentent d'utiliser le plus de données possible en modélisant la dépendance entre des observations successives.

Correction de l'erreur

Le declustering a pour avantage de sélectionner des dépassements de seuil qui peuvent être considérés comme indépendants et permettre ainsi l'inférence sur les excès correspondants. Cependant, ces méthodes impliquent aussi une perte d'information due à la suppression de données extrêmes (Ferro et Segers, 2003). Deux articles suggèrent des méthodes pour résoudre ce problème :

1. Fawcett et Walshaw (2007) proposent d'utiliser tous les dépassements de seuils comme s'ils étaient indépendants, ce qui induit une sous-estimation de l'incertitude associée à l'estimation des paramètres (τ, ξ) de la loi GP. Pour prendre en compte cette erreur, ils utilisent la méthode de Smith (1991) qui corrige la matrice d'information. Cette méthode est illustrée sur des données de surcote de mer et comparée avec l'approche classique du run declustering pour montrer son efficacité.
2. Sabourin (2015) utilise la méthode du run declustering sur les données réelles mais travaille ensuite sur des simulations indépendantes du modèle de mélanges de lois de Dirichlet de Boldi et Davison (2007) pour la mesure angulaire. Pour prendre en compte la perte d'information liée à la sélection des maxima de clusters, le nombre de données simulées est :

$$n_{\text{eff}} = \left\lfloor \frac{n}{T_c} \right\rfloor ,$$

où T_c est la taille moyenne des clusters et n est le nombre d'observations.

Modèle de chaîne de Markov

Ce modèle, proposé par Coles *et al.* (1994) et Smith *et al.* (1997), utilise tous les excès observés de $\{X_1, \ldots, X_n\}$ en supposant que X est une chaîne de Markov d'ordre 1, c'est-à-dire que X_i dépend seulement de l'état précédent X_{i-1}.

La vraisemblance s'écrit alors :

$$f(x_1, \ldots, x_n) = g(x_1) \prod_{i=2}^{n} g(x_i | x_{i-1}) = \frac{\prod_{i=2}^{n} g(x_i, x_{i-1})}{\prod_{i=1}^{n-1} g(x_i)} ,$$

où la queue de la loi marginale est modélisée par une loi GP :

$$H(x) = 1 - \zeta_u \left(1 + \xi \frac{x-u}{\sigma} \right)_+^{-1/\xi} , \quad x > u ,$$

avec $\zeta_u = \Pr(X > u)$. Les lois bivariées sont modélisées par une MEV :

$$G(x_i, x_{i-1}) = \exp\big(- V(z_i, z_{i-1}) \big) ,$$

où $z_i = -1/\log(F(x_i))$. De façon similaire à Huser et Davison (2014), une procédure de censure est appliquée par Ribatet *et al.* (2009) sur les lois bivariées $g(x_i, x_{i-1})$ pour éviter de modéliser des composantes non-extrêmes par une loi max-stable :

$$g(x_i, x_{i-1}) = \begin{cases} \frac{\partial^2}{\partial x_i \partial x_{i-1}} G(x_i, x_{i-1}) & \text{si} \quad x_i, x_{i-1} > u , \\ \frac{\partial}{\partial x_i} G(x_i, u) & \text{si} \quad x_i > u, x_{i-1} \leqslant u , \\ \frac{\partial}{\partial x_{i-1}} G(u, x_{i-1}) & \text{si} \quad x_i \leqslant u, x_{i-1} > u , \\ G(u, u) & \text{si} \quad x_i, x_{i-1} \leqslant u . \end{cases}$$

Ribatet *et al.* (2009) comparent cette méthode avec l'approche conventionnelle des excès de seuil avec run declustering et Fawcett et Walshaw (2006) étendent ce modèle à des ordres supérieurs en l'appliquant sur des données de vents extrêmes.

8.4.3 Sélection de dépassements de seuil par une fonctionnelle

Les sections précédentes recensent plusieurs méthodes prenant en compte la dépendance temporelle dans l'étude des valeurs extrêmes univariées par la méthode des dépassements de seuil. Le principal inconvénient de la plupart d'entre elles est que l'extension en multivarié n'est en général pas triviale.

Dans cette partie, on propose deux méthodes de sélection des dépassements de seuil d'un processus spatial X en utilisant la définition des excès par une fonctionnelle ℓ et les approches de declustering décrites dans la section 8.4.1. Cette sélection permet par la suite d'utiliser le cadre inférentiel de la section 8.2.1 tout en prenant en compte la dépendance temporelle des observations.

On observe $X = (X_1, \ldots, X_d)$ un vecteur aléatoire multivarié et on se donne ℓ une fonctionnelle associée. Les dépassements de X au-dessus du seuil multivarié $u = (u_1, \ldots, u_d)$ sont choisis selon la fonctionnelle ℓ en définissant par exemple $\ell(X) = \max_{j=1,\ldots,d} X_j/u_j$.

Soit $\{L_i\}_{i\geqslant 1}$ la suite définie pour tout $i \in \{1, \ldots, n\}$ par $L_i := \ell(X_i)$: cette suite est stationnaire par stationnarité de X. Une méthode de declustering est directement appliquée sur la série $\{L_i\}_{i\geqslant 1}$, avec le seuil $u = 1$, afin de sélectionner $N_{\perp\!\!\!\perp}$ dépassements indépendants parmi les N_u réellement observés.

En appliquant la sélection d'excès, la vraisemblance (8.7) devient :

$$f_{\text{TO2015}}(x; \psi) = \pi(N_u|n) \prod_{k=1}^{N_{\perp\!\!\!\perp}} \frac{-\partial_{I_k} V\left(\tilde{x}_k^C\right)}{V(u)} \ ,$$

où $\pi(N_u|n)$ est la fonction de densité de la loi binomiale $\text{Bin}(n, V(u))$, et $N_{\perp\!\!\!\perp}$ est le nombre de dépassements indépendants $\{\tilde{X}_1, \ldots, \tilde{X}_{N_{\perp\!\!\!\perp}}\}$ sélectionnés selon la méthode de declustering choisie.

La vraisemblance (8.7) n'est donc modifiée que par le nombre de termes apparaissant dans le produit.

On propose de sélectionner les excès en regardant les dépassements $\{L_i > 1\}$ selon deux méthodes de declustering présentées dans la section 8.4.1.

1. En utilisant le run declustering de Smith (1989) : on détermine les N_c clusters indépendants avec le paramètre de run τ et en choisissant l'excès maximum dans chaque cluster. Avec cette méthode, $N_{\perp\!\!\!\perp} = N_c$.
2. En utilisant le block declustering de Tawn (1988) : on extrait les r statistiques d'ordre indépendantes séparées d'au moins τ observations chacune. Dans ce cas, on a $N_{\perp\!\!\!\perp} = r$.

La figure 8.1 illustre ces deux méthodes de sélection sur une même série stationnaire L, où le seuil est indiqué par la ligne horizontale pointillée. Dans les deux cas, le paramètre de declustering est de 1.

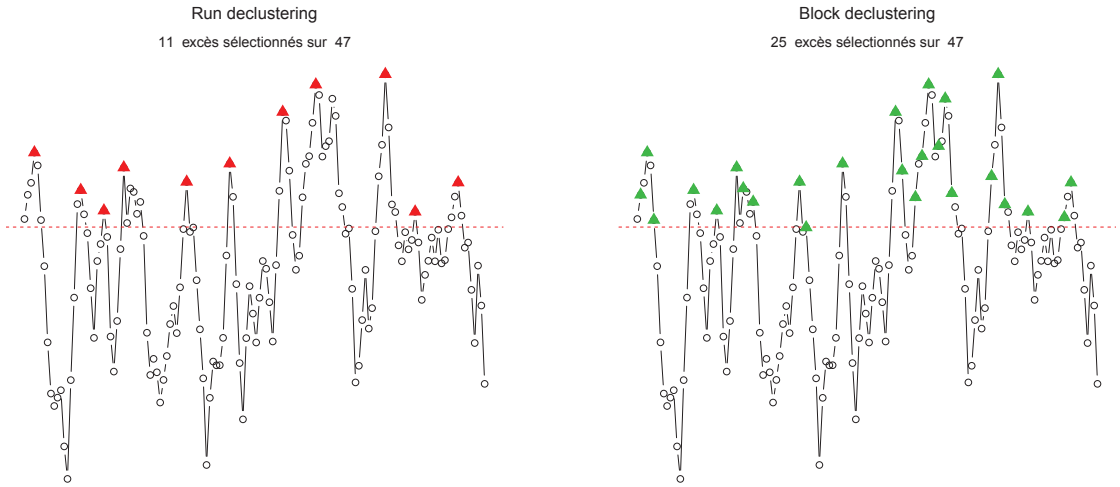


FIGURE 8.1 – Illustration des deux méthodes de sélection d'excès sur une série stationnaire. À gauche, le run declustering de Smith (1989) ; à droite le block declustering de Tawn (1988).

Chapitre 9

Étude sur simulations

Cette partie de la thèse a pour objectif l'estimation de la probabilité conditionnelle (8.1), fomulée par :

$$p_{A_C|A_S}(u) := \Pr\left(\max_{s\in A_C} X(s) > u \;\middle|\; \max_{s\in A_S} X(s) > u\right) ,$$

et correspondant à la probabilité que le processus spatial stationnaire $X \in \mathcal{C}(\mathcal{S})$ dépasse une valeur élevée $u \in \mathbb{R}$ sur au moins un site cible de l'ensemble $A_C \subset \mathcal{S}$, sachant qu'il dépasse cette valeur sur au moins une des stations météorologiques de l'ensemble $A_S \subset \mathcal{S}$.

Dans le chapitre précédent, six estimateurs de $p_{A_C|A_S}(u)$ ont été mis en place dans le cas où u est une valeur trop élevée pour permettre une estimation empirique. Cinq ont été construits en utilisant les processus ℓ-Pareto (Dombry et Ribatet, 2015) et la procédure inférentielle de Thibaud et Opitz (2015). L'approximation (8.4) de la probabilité d'échec est ainsi obtenue et plusieurs estimateurs de la forme (8.5) peuvent être dérivés en fonction du modèle utilisé pour la structure de dépendance asymptotique du processus X.
Parmi eux, quatre estimateurs paramétriques sont définis en utilisant les modèles de processus max-stables :

- de Brown-Resnick,
- Extrémal-t,
- du HKEVP,
- logistique.

Le cinquième estimateur fait appel à l'estimateur empirique de la fonction exposante (de Haan et Ferreira, 2006, p. 236). Enfin, le modèle conditionnel de Heffernan et Tawn (2004) est appliqué sur les séries univariées X_{A_C} et X_{A_S}, définies par :

$$X_{A_S} = \max_{s\in A_S} X(s) \text{ et } X_{A_C} = \max_{s\in A_C} X(s) ,$$

pour fournir un dernier estimateur de $p_{A_C|A_S}(u)$. Des simulations provenant de ce modèle permettent en effet d'estimer la probabilité d'échec conditionnelle empiriquement.

Ces six estimateurs sont respectivement notés dans tout le chapitre par :

1. $\widehat{p}_{\mathrm{BRP}}$, pour le modèle de Brown-Resnick,
2. $\widehat{p}_{\mathrm{ETP}}$, pour le modèle Extrémal-t,
3. $\widehat{p}_{\mathrm{HKEVP}}$, pour le HKEVP,
4. $\widehat{p}_{\mathrm{Log}}$, pour le modèle logistique
5. $\widehat{p}_{\mathrm{np}}$ pour la méthode non-paramétrique et
6. $\widehat{p}_{\mathrm{cond}}$ pour l'estimateur construit sur le modèle conditionnel de Heffernan et Tawn (2004).

Ces estimateurs sont testés dans un premier temps (section 9.1) à l'aide de simulations indépendantes de processus spatiaux. Dans un second temps (section 9.2), les méthodes de sélection des ℓ-excès par declustering, décrites dans la section 8.4, sont testées sur des simulations de processus temporellement corrélés.

Dans la suite de ce chapitre, la valeur de u choisie pour définir la probabilité d'échec conditionnelle est le niveau de retour centennal x_{100}.

9.1 Simulations indépendantes de processus journaliers

Dans cette partie, on s'intéresse à des simulations indépendantes d'un processus journalier X observé sur un ensemble fini $A_C \cup A_S \subset \mathcal{S}$. Trois générateurs sont proposés :

1. $\mathcal{G}_{\text{HKEVP}}(\alpha)$ simule des données de marges Fréchet unitaires selon la structure de dépendance spatiale du HKEVP. Parmi tous les modèles max-stables, celui-ci est choisi en raison de la forme explicite de sa fonction de répartition, ce qui permet de calculer la valeur exacte de la probabilité d'échec (8.1) en fonction des paramètres α et τ et de la position des nœuds $\{v_1, \ldots, v_L\}$ (cf. Section 9.1.1).
   Les paramètres du générateur $\mathcal{G}_{\text{HKEVP}}(\alpha)$ sont fixés tels que $\tau = 1.5$, avec des nœuds coïncidant avec l'ensemble $A_C \cup A_S$. Seul le paramètre de dépendance α est modifié dans ce plan de simulation.

2. $\mathcal{G}_{\text{Gauss}}(\lambda)$ produit des simulations d'un processus gaussien stationnaire et isotrope (cf. Chapitre 2), de fonction de corrélation de type *Powered Exponential* :

$$\rho(h) = \exp\left[-\left(\frac{h}{\lambda}\right)^{\nu}\right] .$$

   Le paramètre de lissage ν est fixé à 1.5 pour ce générateur : seul le paramètre de portée λ varie en fonction du type de données souhaité pour les simulations.

3. $\mathcal{G}_{\text{Student}}(\lambda)$ simule un processus Student stationnaire et isotrope. Par souci de cohérence avec le générateur précédent $\mathcal{G}_{\text{Gauss}}(\lambda)$, la fonction de corrélation est aussi de type *Powered Exponential*.
   Avec ce générateur, le paramètre de lissage ν est aussi fixé à 1.5 et le degré de liberté δ à 2 : seul le paramètre de portée λ varie pour produire des données selon différentes forces de dépendance.

Les trois générateurs choisis correspondent respectivement aux cas où le processus spatial est :
– max-stable (i.e déjà dans l'attracteur),
– asymptotiquement indépendant (processus gaussien, cf. Sibuya (1960)),
– asymptotiquement dépendant (processus Student, cf. Heffernan (2000)).

9.1.1 Valeur exacte de la probabilité d'échec

De façon analogue à la comparaison des modèles spatiaux menée dans le chapitre 5, les estimateurs de la probabilité d'échec (8.1) sont mis en compétition sur des simulations de processus selon plusieurs générateurs. Il est donc nécessaire de connaître la valeur exacte de $p_{A_C|A_S}(x_{100})$ pour chaque processus simulé.

Écrivons la probabilité d'échec conditionnelle (8.1) avec la fonction de répartition F du processus généré X. Si pour tout ensemble fini de sites $A = \{s_1, \ldots, s_d\}$, on note :

$$F_A(u) = \Pr\big(X(s) \leqslant u(s) \ : \ s \in A\big) ,$$

alors la probabilité d'échec conditionnelle (8.1) s'écrit :

$$p_{A_C|A_S}(x_{100}) = \frac{1 - F_C(x_{100}) - F_S(x_{100}) + F_{C\cup S}(x_{100})}{1 - F_S(x_{100})} . \tag{9.1}$$

Grâce à la forme explicite de la fonction exposante V du HKEVP, la probabilité d'échec conditionnelle associée aux données simulées par $\mathcal{G}_{\text{HKEVP}}(\alpha)$ peut être calculée de façon exacte en utilisant (9.1) :

$$p_{A_C|A_S}(x_{100}) = \frac{1 - \exp\big(-V_{A_C}(x_{100})\big) - \exp\big(-V_{A_S}(x_{100})\big) + \exp\big(-V_{A_C\cup A_S}(x_{100})\big)}{1 - \exp\big(-V_{A_S}(x_{100})\big)} ,$$

où V_A désigne la fonction exposante évaluée aux sites de l'ensemble fini $A \in \mathcal{S}$.
En fin de compte, on retrouve bien l'approximation (8.4) comme étant le développement à l'ordre 1 de la formule précédente.

Les fonctions de répartition multivariées associées aux générateurs $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$ ne sont pas explicites, mais il est possible de les approcher en utilisant l'algorithme de simulation de Genz (1992). Pour une région A, la fonction de répartition est approximée avec une erreur associée ε_A :

$$\widehat{F}_A(x_{100}) = F_A(x_{100}) + \varepsilon_A .$$

Cependant, même avec une erreur ε_A très faible, le calcul de la probabilité d'échec conditionnelle par (9.1) pose problème du fait de la division par $1 - F_{A_S}(x_{100})$ qui est proche de 0 pour une valeur élevée de x_{100}.
C'est pour cette raison que dans les sections suivantes, la probabilité d'échec exacte considérée pour ces deux générateurs est calculée comme la moyenne sur un ensemble d'estimations obtenues par l'algorithme de Genz (1992). Ces estimations moyennes ont été comparées en petite dimension avec les valeurs obtenues sur des simulations intensives pour s'assurer de leur validité.

9.1.2 Comparaison des estimateurs sur une petite région

Les six estimateurs de la probabilité d'échec conditionnelle (9.1) sont évalués dans cette section sur des simulations indépendantes selon les générateurs de processus spatiaux $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$. En raison de la forte demande en calcul des estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$, cette première comparaison sur simulation est faite sur un réseau de positions relativement petit ($d = |\mathcal{S}| = 9$).

Dans ce plan de simulation, toutes les positions sont supposées jaugées : des observations sont disponibles sur la cible A_C. Ceci afin de pouvoir estimer la probabilité d'échec conditionnelle avec les deux méthodes non-paramétriques.

Les données artificielles sont simulées sur la grille de taille 3×3 sites $\{0, 1, 2\}^2$. Les positions sont indiquées sur la figure 9.2a : le point $(1, 1)$ correspond à la cible A_C (en rouge) et les huit stations de l'ensemble A_S sont représentées en noir.

Les trois générateurs $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$ sont utilisés pour produire ces données artificielles sur les neuf positions de l'ensemble $A_C \cup A_S$. Pour chacun d'entre eux, on fait varier le paramètre contrôlant le degré de dépendance du processus spatial X. Les autres paramètres inhérents aux modèles de simulation restent constants, comme indiqué au début de la section 9.1.

Chaque processus artificiel est généré sur une longueur de $58 \times 91 = 5278$ jours, ce qui correspond à 58 années d'observations restreintes à l'automne. Ce choix est fait en adéquation avec la quantité de données réelles disponibles dans le deuxième jeu de données de précipitations (de longueur maximale 58 ans) et pour une saison automnale comptant 91 jours. On choisit de se restreindre à cette saison pour être cohérent avec les observations du chapitre 4 qui indiquent de plus fortes valeurs extrêmes sur la période de septembre à novembre.

Les processus simulés dans cette section n'ont de lien avec les données réelles de cumuls journaliers de précipitations automnales que par le nombre de réalisations générées. Ce choix permettra d'appliquer par la suite le ou les plus performant(s) sur les données réelles. Cette application est réalisée en utilisant la région du centre-est de la France et est disponible dans le chapitre de conclusions de cette partie III.

Pour les quatre estimateurs construits à partir de modèles max-stables et utilisant l'inférence de Thibaud et Opitz (2015), le seuil marginal $u_{\text{GP}}(s)$ choisi pour l'ajustement de la queue de distribution par une loi GP est le même que le seuil $u(s)$ permettant de définir les ℓ-excès. Ce seuil marginal unique est choisi comme le quantile empirique d'ordre 0.9.

Par souci de cohérence, un seuil équivalent est considéré pour les deux autres approches. Pour l'estimateur non paramétrique (8.14), l'ordre k est choisi par :

$$k := \lfloor 5278 \times (1 - 0.9) \rfloor = 527 \ .$$

La procédure de simulation et estimation est répétée $N = 100$ fois pour chaque type de données simulées, afin de regarder un échantillon d'estimations $\widehat{p}_{A_C|A_S,i}$ auquel il est possible d'associer une erreur quadratique moyenne (MSE), calculée par :

$$\text{MSE}(\widehat{p}_{A_C|A_S}) = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{p}_{A_C|A_S,i} - p_{A_C|A_S} \right)^2 \ ,$$

où $\widehat{p}_{A_C|A_S,i}$ représente un des six estimateurs parmi $\widehat{p}_{\text{BRP}}$, $\widehat{p}_{\text{ETP}}$, $\widehat{p}_{\text{HKEVP}}$, $\widehat{p}_{\text{Log}}$, $\widehat{p}_{\text{np}}$ et $\widehat{p}_{\text{cond}}$ sur la i-ième itération.

Les résultats sont donnés dans la table 9.1, en fonction du générateur utilisé, pour les six estimateurs de la probabilité d'échec conditionnelle. Pour faciliter la lecture, la plus faible MSE est affichée en vert et en gras, tandis que la plus forte est donnée en rouge. Les erreurs quadratiques associées à l'estimation d'une probabilité étant très petites, les valeurs de la table 9.1 sont multipliées par 1000 par souci de lisibilité.

Les boîtes à moustaches des figures 9.3, 9.4 et 9.5 (données en annexe) affichent les échantillons complets obtenus sur ces simulations, pour des données simulées respectivement par $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$.

Dans le cas des données simulées avec $\mathcal{G}_{\text{Gauss}}(\lambda)$ ou $\mathcal{G}_{\text{Student}}(\lambda)$, la valeur exacte de la probabilité d'échec est calculée sur un échantillon d'estimations obtenu par l'algorithme de Genz (1992). La moyenne de ces estimations est indiquée par une ligne rouge en tirets et l'intervalle de confiance à 95% par des lignes rouges pointillées. Ce dernier est calculé à partir des quantiles d'ordres respectifs 2.5% et 97.5%. On rappelle que la valeur moyenne est utilisée comme valeur exacte de la probabilité d'échec pour calculer les erreurs quadratiques de la table 9.1.

Plusieurs points intéressants ressortent de l'examen de ces résultats.

D'un côté, l'estimateur non paramétrique $\widehat{p}_{\text{np}}$ semble être l'un des plus compétitifs parmi ceux qui sont proposés dans ce chapitre. En effet, la MSE associée à cette méthode est la plus petite, ou très proche de

		$\widehat{p}_{\text{BRP}}$	$\widehat{p}_{\text{ETP}}$	$\widehat{p}_{\text{HKEVP}}$	$\widehat{p}_{\text{Log}}$	$\widehat{p}_{\text{np}}$	$\widehat{p}_{\text{cond}}$
$\mathcal{G}_{\text{HKEVP}}$	0.9	1.6321	0.3122	**0.2984**	1.6341	0.3134	1.112
	0.5	9.8797	4.4422	5.643	4.2803	**0.5077**	32.8597
	0.1	10.4951	57.8522	0.6776	16.0206	**0.6244**	116.4886
$\mathcal{G}_{\text{Gauss}}$	1	4.0764	2.3697	0.7892	**0.021**	3.8454	0.0737
	5	27.7131	7.2661	0.9353	**0.6094**	19.9112	14.8013
	10	7.428	5.9441	**1.0753**	1.6411	21.0339	37.9133
$\mathcal{G}_{\text{Student}}$	1	6.4995	0.5765	7.1213	12.9817	**0.1638**	17.5094
	5	2.0802	28.5404	17.2981	29.4168	**0.4387**	90.906
	10	18.4024	70.9634	11.8711	18.3643	**0.9984**	153.0321

TABLE 9.1 – MSE de six estimateurs de la probabilité d'échec conditionnelle sur une petite région (après multiplication par un facteur 1000).

la plus petite, lorsque les données sont simulées avec $\mathcal{G}_{\text{HKEVP}}(\alpha)$ ou $\mathcal{G}_{\text{Student}}(\lambda)$. Si le processus simulé est asymptotiquement indépendant (i.e simulé avec $\mathcal{G}_{\text{Gauss}}(\lambda)$), cet estimateur semble cependant rencontrer plus de difficultés.

D'un autre côté, l'estimateur $\widehat{p}_{\text{cond}}$ est très souvent celui affichant la MSE la plus élevée, en particulier lorsque la dépendance spatiale est importante : $\mathcal{G}_{\text{HKEVP}}(0.1)$, $\mathcal{G}_{\text{HKEVP}}(0.5)$, $\mathcal{G}_{\text{Gauss}}(10)$, $\mathcal{G}_{\text{Student}}(5)$ et $\mathcal{G}_{\text{Student}}(10)$. En se reportant aux figures 9.3, 9.4 et 9.5, on peut voir que cette méthode affiche la plus grande variabilité et semble avoir des difficultés à estimer correctement une probabilité d'échec conditionnelle supérieure à 0.1.

Les estimateurs $\widehat{p}_{\text{HKEVP}}$ et $\widehat{p}_{\text{Log}}$ montrent des résultats satisfaisants lorsque les données artificielles sont soit asymptotiquement indépendantes avec $\mathcal{G}_{\text{Gauss}}(\lambda)$, soit dans l'attracteur avec $\mathcal{G}_{\text{HKEVP}}(\alpha)$. En revanche, la MSE est plus importante si le processus observé est généré par $\mathcal{G}_{\text{Student}}$. La figure 9.5 indique que ces méthodes ont tendance à sous-estimer $p_{A_C|A_S}$ dans ce cas.

Enfin, les estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$ ne semblent pas faire partie des plus compétitifs. D'après les figures correspondant à cette étude, $\widehat{p}_{\text{BRP}}$ est presque systématiquement biaisé, avec une tendance à sous-estimer la probabilité d'échec lorsque celle-ci est grande, et à la surestimer lorsqu'elle est plus petite. L'estimateur $\widehat{p}_{\text{ETP}}$ présente lui aussi des résultats mitigés : le cas où il montre des résultats satisfaisants par rapport aux autres apparaît lorsque le processus observé possède une faible dépendance spatiale : $\mathcal{G}_{\text{HKEVP}}(0.9)$, $\mathcal{G}_{\text{Gauss}}(1)$ et $\mathcal{G}_{\text{Student}}(1)$.

De plus, $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$ sont extrêmement coûteux en temps de calcul, même avec un réseau de sites limité ($d = 9$ dans notre cas). Pour cette raison, et parce que les MSE obtenues par ces méthodes sont généralement plus élevées que les autres, il est décidé de les éliminer dans les études suivantes.

9.1.3 Probabilité de contagion sur une région

Dans ce plan de simulation, on s'intéresse à l'estimation de la probabilité de *contagion* d'un excès sur une région, qui décrit l'impact d'un dépassement du niveau de retour centennal en une station donnée sur le reste du réseau de positions. Cette valeur correspond à la probabilité d'échec conditionnelle (8.1) où $A_S = \{s_0\}$ est un singleton.

La question de la contagion d'un événement extrême est étudiée par Fonseca *et al.* (2012), qui définissent un indice lié pour des processus spatiaux. Cet indice (dit *de contagion*) quantifie le nombre de stations d'une région A touchées par un événement extrême sachant que la position $s_i \in A$ connaît un excès :

$$\text{CI}(A, s_0) = \lim_{u \to 1} \mathbb{E}\left[\sum_{\substack{s_j \in A \\ s_j \neq s_0}} \mathbb{1}_{\{F_j(X(s_j)) > u\}} \;\middle|\; F_0(X(s_0)) > u \right],$$

où F_j est la fonction de répartition de la variable $X(s_j)$ pour $j \in \{0, \dots, d\}$. Fonseca *et al.* (2012) établissent les liens entre cet indice et les mesures de dépendance χ et θ (cf. section 1.2). Cependant, l'article traite seulement du cas où le processus observé X est max-stable.

L'étude menée dans cette section peut être reliée à celle de Fonseca *et al.* (2012), avec toutefois deux différences notables. La première est que le processus X est simulé à partir de l'un des trois générateurs $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$. La seconde est que l'on s'intéresse à une *probabilité de contagion* (autrement dit, la probabilité d'échec conditionnelle) plutôt qu'au nombre d'excès sur un réseau.

Puisque les estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$ ne sont plus utilisés, le nombre d de positions considérées dans cette étude peut être plus grand. On utilise cette fois 25 positions sur la grille régulière $\{0, 1, 2, 3, 4\}^2$ et on fait

correspondre l'unique station $s_0 \in A_S$ avec la position centrale $(3, 3)$. Les cibles correspondent à l'ensemble du réseau restant : les positions sont indiquées sur la figure 9.2b.

Hormis ces différences, le plan de simulation est réalisé selon le même schéma que celui de la section 9.1.2 : les données sont générées indépendamment sur une période de 5278 jours et la procédure de simulation-estimation est faite $N = 100$ fois pour calculer une erreur quadratique moyenne.

Les résultats obtenus sont affichés dans la table 9.2 avec le même code couleur que précédemment. On rappelle que ces erreurs ont été multipliées par 1000 par souci de lisibilité.

Les échantillons complets sont affichés sous la forme de boîtes à moustaches sur les figures 9.6, 9.7 et 9.8 pour les simulations provenant de $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$ respectivement.

		$\widehat{p}_{\text{HKEVP}}$	$\widehat{p}_{\text{Log}}$	$\widehat{p}_{\text{np}}$	$\widehat{p}_{\text{cond}}$
$\mathcal{G}_{\text{HKEVP}}$	0.9	170.8644	156.1596	299.7585	**69.6207**
	0.7	8.2834	**2.0742**	18.6457	202.5065
	0.5	**0.0216**	6.8869	0.5545	288.3139
	0.3	**0.0571**	13.6878	4.2839	501.889
	0.1	**0.2195**	13.9111	4.2466	623.6813
$\mathcal{G}_{\text{Gauss}}$	1	593.2588	551.7471	781.5879	**0.5525**
	3	664.7213	550.1387	660.2195	**110.5625**
	5	535.682	426.549	525.7768	**181.5117**
	7	419.302	323.7025	405.9225	**93.186**
	9	334.5572	254.5491	319.5484	**91.5008**
$\mathcal{G}_{\text{Student}}$	1	353.4727	333.3476	497.7963	**257.7815**
	3	204.8691	148.0092	218.8551	**133.3699**
	5	123.5427	**77.3184**	121.2248	215.2825
	7	88.8049	**53.8943**	82.2408	291.4691
	9	61.1309	**37.1353**	55.9006	358.6587

TABLE 9.2 – MSE sur l'estimation de la probabilité de contagion d'un excès en une station sur le reste du réseau de positions (après multiplication par un facteur 1000).

Les erreurs quadratiques associées à la probabilité de contagion dans la table 9.2 sont en général beaucoup plus élevées que pour l'estimation de la probabilité d'échec conditionnelle dans la section précédente. Dans presque tous les cas considérés, l'estimation semble donc être très mauvaise.

Quelques cas se distinguent toutefois par une relative bonne performance :

1. $\widehat{p}_{\text{HKEVP}}$ sur des données générées par $\mathcal{G}_{\text{HKEVP}}(\alpha)$, pour $\alpha \leqslant 0.5$,
2. $\widehat{p}_{\text{np}}$ pour des données générées avec $\mathcal{G}_{\text{HKEVP}}(0.5)$,
3. $\widehat{p}_{\text{Log}}$ lorsque les données sont générées avec $\mathcal{G}_{\text{HKEVP}}(0.5)$ et $\mathcal{G}_{\text{HKEVP}}(0.7)$,
4. $\widehat{p}_{\text{cond}}$ quand la probabilité de contagion exacte est proche de 0 avec des données simulées par $\mathcal{G}_{\text{Gauss}}(1)$.

Pour tous les autres cas, la probabilité de contagion n'est pas bien estimée. Les figures associées à cette étude montrent clairement que les trois premières méthodes sont très fortement biaisées et surestiment la probabilité attendue : ces estimateurs ne peuvent donc pas être retenus pour estimer cette quantité.

L'estimateur $\widehat{p}_{\text{cond}}$ montre également ses limites dans l'estimation de la probabilité de contagion : la variabilité associée à cette méthode est presque toujours la plus importante. A plusieurs reprises, les estimations varient même entre 0 et 1, ce qui n'apporte aucune information sur $p_{A_C|A_S}(x_{100})$ dans ces cas.

On en conclut que dans la configuration posée par cette étude, l'estimation de la probabilité de contagion (8.1) semble très difficile : aucun des estimateurs proposés ne répond à cette question. La quantification de la contagion d'un excès sur un réseau reste un problème qui nécessiterait des recherches plus approfondies.

9.1.4 Cible non jaugée

Dans les sections précédentes, on suppose que des données sont observées sur l'ensemble des sites cibles A_C, mais en pratique on souhaite extrapoler la probabilité d'échec conditionnelle (8.1) pour une cible non jaugée. Ce cas est traité dans cette section, avec une cible unique et 24 stations météorologiques dont la disposition est donnée sur la figure 9.2c.

Si la position cible est non jaugée, il n'est pas possible d'utiliser les estimateurs $\widehat{p}_{\text{np}}$ et $\widehat{p}_{\text{cond}}$. En effet, ces deux méthodes requièrent des données observées sur A_C pour calculer :

– les statistiques d'ordre dans la formule (8.14),
– la série univariée X_{A_C} en utilisant le modèle de Heffernan et Tawn (2004).

Comme les deux estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$ ont été éliminés à cause d'une demande en temps de calcul trop importante et de performances mitigées, cette section se concentre sur les deux estimateurs restants : $\widehat{p}_{\text{HKEVP}}$ et $\widehat{p}_{\text{Log}}$. Le plan de simulation est similaire à celui des sections 9.1.2 et 9.1.3.

Les erreurs quadratiques obtenues dans cette partie, multipliées par 1000, sont données dans le tableau 9.3. Les boîtes à moustaches associées aux échantillons complets sont quant à elles affichées sur les figures 9.9, 9.10 et 9.11 pour les simulations par $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$ respectivement. Comme les estimations ont une variabilité très faible et sont souvent très proches de la valeur exacte (ligne rouge en tirets), les échelles ne correspondent pas d'une figure à l'autre pour permettre une meilleure comparaison, par type de données simulées, entre les deux estimateurs.

		$\widehat{p}_{\text{HKEVP}}$	$\widehat{p}_{\text{Log}}$
$\mathcal{G}_{\text{HKEVP}}$	0.9	3.844	**3.679**
	0.5	**0.096**	0.294
	0.1	**11.217**	18.11
$\mathcal{G}_{\text{Gauss}}$	1	6.556	**6.205**
	5	3.403	**2.51**
	10	1.78	**0.286**
$\mathcal{G}_{\text{Student}}$	1	**0.339**	0.375
	5	**14.613**	18.521
	10	**29.322**	39.922

TABLE 9.3 – MSE sur l'estimation de la probabilité d'échec conditionnelle avec une cible non jaugée (après multiplication par un facteur 1000).

Lorsque la probabilité d'échec conditionnelle est estimée sur une cible unique non jaugée, les erreurs quadratiques moyennes des deux estimateurs évalués dans ce plan de simulation sont généralement équivalentes et relativement faibles, indiquant une bonne précision d'estimation.

D'un côté, l'estimation construite à partir du modèle HKEVP est légèrement meilleure si le processus observé X est asymptotiquement dépendant ou max-stable avec une dépendance spatiale assez forte. D'un autre côté, si le processus est asymptotiquement indépendant, $\widehat{p}_{\text{Log}}$ devient légèrement plus performant.

Dans tous les cas, les résultats sont assez proches de la valeur exacte attendue et les deux estimateurs peuvent être considérés comme satisfaisants, compte tenu de la difficulté du problème. Les exceptions notables à cette conclusion concernent les données simulées avec $\mathcal{G}_{\text{Student}}(5)$ et $\mathcal{G}_{\text{Student}}(10)$, où les deux méthodes sous-estiment la valeur de la probabilité d'échec (voir figure 9.11).

9.2 Données temporellement corrélées

Les estimateurs de la probabilité d'échec conditionnelle proposés dans le chapitre précédent ont été mis en compétition sur plusieurs simulations de processus spatiaux. Jusqu'ici, les simulations étaient générées de façon indépendantes par trois générateurs : $\mathcal{G}_{\text{HKEVP}}(\alpha)$, $\mathcal{G}_{\text{Gauss}}(\lambda)$ et $\mathcal{G}_{\text{Student}}(\lambda)$.

En pratique, la plupart des données observées ne sont pas indépendantes d'un jour à l'autre. Les précipitations journalières par exemple sont souvent considérées comme indépendantes si elles sont séparées d'au moins une journée (de Haan *et al.*, 2014), mais les observations consécutives présentent tout de même une corrélation non négligeable.

On souhaite donc évaluer l'effet de la méthode de sélection des ℓ-excès décrite dans la section 8.4.3 sur l'estimation de la probabilité d'échec conditionnelle lorsque le processus X observé est temporellement dépendant.

Dans un premier temps, il est nécessaire de simuler des réalisations de processus spatiaux qui sont temporellement corrélés dans les valeurs extrêmes. Ce point est traité dans la section 9.2.1. Dans un second temps (Section 9.2.2), les estimateurs de la probabilité d'échec conditionnelle sont utilisés sur des données corrélées avec et sans sélection d'excès afin d'évaluer le gain en précision de cette méthode.

9.2.1 Simulation de données corrélées

En premier lieu, il est nécessaire de mettre en place des générateurs capables de construire des données temporellement corrélées de telle sorte que les excès de seuil forment des clusters. Autrement dit, on souhaite

que l'indice extrémal θ de la série $\{X_i(s)\}_{i\geqslant 1}$, défini dans la section 1.1.5 soit strictement inférieur à 1 pour tout $s \in \mathcal{S}$.

Comme indiqué dans (Beirlant *et al.*, 2004, Chapitre 10) une série stationnaire X peut être auto-corrélée et avoir pourtant un indice extrémal $\theta = 1$ signifiant que les excès de seuil apparaissent de façon isolée et indépendante. C'est par exemple le cas du processus autorégressif gaussien AR(1), défini pour tout $i \geqslant 1$ par :

$$X_{i+1} = \varphi X_i + \varepsilon_i \ , \tag{9.2}$$

où $\{\varepsilon_i\}_{i\geqslant 1}$ est un bruit blanc gaussien et $\varphi \in (-1, 1)$ est le coefficient d'autorégression. La série X est temporellement corrélée mais elle reste asymptotiquement indépendante, et ce, quelque soit la valeur de $\varphi \in (-1, 1)$ (Sibuya, 1960).

En revanche, en utilisant un bruit blanc $\{\varepsilon_i\}_{i\geqslant 1}$ de loi Student dans la formule (9.2), on obtient une série stationnaire asymptotiquement dépendante. Une première possibilité pour simuler un processus spatial temporellement asymptotiquement dépendant est donc de poser $X_{i+1}(s) = \varphi X_i(s) + \varepsilon_i(s)$ pour tout $s \in \mathcal{S}$, où le processus $\{\varepsilon_i\}_{i\geqslant 1}$ est obtenu par exemple en utilisant le générateur $\mathcal{G}_{\text{Student}}$ défini au début de ce chapitre.

Dans les sections suivantes, deux autres méthodes sont proposées afin de simuler des processus spatiaux dont les réalisations $\{X_i(s)\}_{i\geqslant 1}$ sont asymptotiquement corrélées pour $s \in \mathcal{S}$.

Processus autorégressif ARMAX

Un modèle de simulation envisageable est celui d'un processus spatial ARMAX d'ordre 1. Un processus ARMAX d'ordre 1 noté $\{X_i\}_{i\geqslant 1}$ est défini pour tout $i \in \mathbb{Z}$ par (Beirlant *et al.*, 2004, Exemple 10.3 p.374) :

$$X_i = \max\big(\varphi X_{i-1}, (1-\varphi)Z_i\big) \ ,$$

où $\varphi \in [0, 1)$ est le coefficient qui contrôle la force de dépendance temporelle et $\{Z_i\}_{i\geqslant 1}$ est une suite iid de variables aléatoires de loi Fréchet unitaire. Une solution stationnaire de la forme récursive est :

$$X_i = \max_{j\geqslant 0} \alpha^j (1-\alpha) Z_{i-j} \ , \quad i \in \mathbb{Z} \ .$$

De façon analogue, on construit le processus ARMAX spatial en définissant pour tout $s \in \mathcal{S}$ et pour tout $i \in \mathbb{Z}$:

$$X_i(s) = \max\big(\varphi X_{i-1}(s), (1-\varphi)Z_i(s)\big) \ , \tag{9.3}$$

où Z est un processus max-stable simple défini sur $\mathcal{S}$ et $\varphi \in [0, 1)$ est le coefficient de dépendance temporelle.

Le processus ARMAX est temporellement asymptotiquement dépendant et la valeur de φ influe sur l'indice extrémal θ par la relation (Beirlant *et al.*, 2004, Chapitre 10) : $\theta = 1 - \varphi$. Plus le coefficient φ est proche de 1, plus l'auto-corrélation est forte et plus la valeur de l'indice extrémal θ diminue : les dépassements de seuil ont alors tendance à apparaître en paquets de plus en plus gros.

Pour mener l'étude sur simulations dans la section 9.2.1, il est nécessaire de connaître la valeur exacte de la probabilité d'échec conditionnelle (8.1) que les estimateurs doivent retrouver. Pour le processus ARMAX spatial, cette valeur est la même que pour le processus indépendant Z. Autrement dit, pour tout ensemble $A = \{s_1, \ldots, s_d\} \subset \mathcal{S}$, la fonction de répartition du processus ARMAX spatial évalué sur A est :

$$\Pr\big(X_i(s_1) \leqslant x_1, \ldots, X_i(s_d) \leqslant x_d\big) = \exp\big(-V_A(x_1, \ldots, x_d)\big) \ ,$$

où V_A est la fonction exposante du processus max-stable Z servant à construire X dans (9.3), évalué aux positions de l'ensemble fini $A \subset \mathcal{S}$.

Démonstration. Pour tout $s \in \mathcal{S}$, une solution stationnaire de la construction récursive (9.3) est (Beirlant *et al.*, 2004) :

$$X_i(s) = \max_{j\geqslant 0} \varphi^j (1-\varphi) Z_{i-j}(s) \ .$$

Cette solution est utilisée pour écrire la fonction de répartition jointe du processus X :

$$\begin{aligned}
\Pr\big(X_i(s_1) \leqslant x_1, \ldots, X_i(s_d) \leqslant x_d\big) &= \Pr\left(\max_{j\geqslant 0} \varphi^j(1-\varphi) Z_{i-j}(s_1) \leqslant x_1, \ldots, \max_{j\geqslant 0} \varphi^j(1-\varphi) Z_{i-j}(s_d) \leqslant x_d\right) \\
&= \Pr\left(\bigcap_{j\geqslant 0}\left\{Z_{i-j}(s_1) \leqslant \frac{x_1}{\varphi^j(1-\varphi)}\right\}, \ldots, \bigcap_{j\geqslant 0}\left\{Z_{i-j}(s_d) \leqslant \frac{x_d}{\varphi^j(1-\varphi)}\right\}\right) \\
&= \prod_{j\geqslant 0} \Pr\left(Z_{i-j}(s_1) \leqslant \frac{x_1}{\varphi^j(1-\varphi)}, \ldots, Z_{i-j}(s_d) \leqslant \frac{x_d}{\varphi^j(1-\varphi)}\right) \\
&= \prod_{j\geqslant 0} \exp\left(-V_A\left(\frac{x_1}{\varphi^j(1-\varphi)}, \ldots, \frac{x_d}{\varphi^j(1-\varphi)}\right)\right) \\
&= \exp\left(-V_A(x_1, \ldots, x_d) \underbrace{\sum_{j\geqslant 0} \varphi^j(1-\varphi)}_{=1}\right) \\
&= \exp\big(-V_A(x_1, \ldots, x_d)\big) \ .
\end{aligned}$$

□

Cumul de données

Une autre option possible pour simuler des données asymptotiquement dépendantes est de considérer des cumuls d'observations indépendantes sur une période de h jours. Soit $\tilde{X}$ une série journalière iid, la série X des cumuls sur h jours est définie par :

$$X_i = \tilde{X}_i + \cdots + \tilde{X}_{i+h} \ .$$

Étudier des données de précipitations cumulées sur plusieurs jours présente un intérêt applicatif. L'analyse porte alors sur les épisodes pluvieux extrêmes à l'origine d'inondations telles que celles observées en Europe au printemps 2016 ou lors des épisodes cévenols qui peuvent apparaître au début de l'automne.

Pour l'étude sur simulations menée dans la section 9.2.1, le principal inconvénient à travailler avec ce type de série temporelle est que l'on ne connaît pas la valeur exacte de la probabilité d'échec conditionnelle $p_{A_C|A_S}$, ni celle du niveau de retour centennal x_{100}.

Ce problème peut être résolu en calculant empiriquement la probabilité d'échec conditionnelle observée sur des simulations intensives. Il faut cependant au préalable estimer le niveau de retour centennal associé à ces simulations. On propose de suivre la méthode décrite par Cai *et al.* (2013) pour estimer ce niveau de retour en utilisant l'indice extrémal θ :

1. Estimer l'indice extrémal θ sur la série observée $\{X_i\}_{i=1,\ldots,n}$ en utilisant par exemple la méthode de Ferro et Segers (2003).
2. Ajuster la queue de distribution de X par une loi GP comme si elle était indépendante. Les estimations des paramètres τ et ξ sont obtenus par maximum de vraisemblance.
3. En choisissant un ordre k assez grand par rapport au nombre d'observations n, le niveau de retour à T-années peut être estimé par :

$$\hat{x}_T := X_{n-k+1,n} + \hat{\tau}\frac{\left(\frac{k}{\hat{\alpha}n}\right)^{\hat{\xi}} - 1}{\hat{\xi}} \ ,$$

où $\hat{\alpha} = 1 - (1 - 1/T)^{1/(n_y\hat{\theta})}$, n_y est le nombre d'observations de X par an ($n_y = 91$ dans le cas présent puisqu'on se concentre sur les données automnales) et $X_{n-k+1,n}$ est la k-ième plus grande valeur observée.

Dans le cadre spatial, on répète la procédure ci-dessus pour chaque station s où le processus spatial X est simulé : on obtient ainsi des estimations de niveaux de retour centennaux

$$\{\hat{x}_{100}(s_1), \ldots, \hat{x}_{100}(s_d)\} \ .$$

Soit $\tilde{X} \in \mathcal{C}(\mathcal{S})$ le processus servant à générer le processus X par cumul de réalisations successives. Si les lois marginales de $\tilde{X}$ sont égales, alors il en va de même pour X. Même si les marges de X sont inconnues, on sait qu'elles sont égales d'un site à l'autre : le niveau de retour centennal théorique $x_{100}(s)$ est alors identique pour tout $s \in \mathcal{S}$ et il est donc possible de l'estimer en conservant la moyenne $\frac{1}{d}\sum_{j=1}^{d} \hat{x}_{100}(s_j)$.

9.2.2 Comparaison sur données corrélées

D'après les résultats de la section 9.2.1, quatre générateurs de données présentant une dépendance temporelle asymptotique peuvent être construits :

- $\mathcal{G}_{\mathrm{AR-Student}}(\varphi)$ simule un processus spatial de Student obtenu par $\mathcal{G}_{\mathrm{Student}}(\lambda)$, corrélé temporellement selon le coefficient d'auto-régression $\varphi \in (-1, 1)$.
- $\mathcal{G}_{\mathrm{ARMAX}}(\varphi)$ simule un processus spatial ARMAX de coefficient $\varphi \in (0, 1)$. Le modèle max-stable utilisé pour le processus de construction Z dans (9.3) est le HKEVP de Reich et Shaby (2012) en raison de la forme explicite de la fonction de répartition pour toute dimension, ce qui permet de calculer directement la valeur exacte de la probabilité d'échec conditionnelle.
- $\mathcal{G}_{\mathrm{DC-HKEVP}}(h)$ et $\mathcal{G}_{\mathrm{DC-Student}}(h)$ correspondent respectivement à des *données cumulées* (DC) sur h jours. Ces deux générateurs sont respectivement construits sur des simulations indépendantes provenant de $\mathcal{G}_{\mathrm{HKEVP}}(\alpha)$ et $\mathcal{G}_{\mathrm{Student}}(\lambda)$.

Dans cette partie, les valeurs des paramètres λ et α utilisées pour les générateurs $\mathcal{G}_{\mathrm{HKEVP}}(\alpha)$ et $\mathcal{G}_{\mathrm{Student}}(\lambda)$ sont respectivement fixées à 3 et 0.3 qui correspondent à des cas de dépendance spatiale modérée.

Les données sont simulées sur le réseau présenté sur la figure 9.2c, mais contrairement à ce qui a été fait dans la section 9.1.4, le site cible est ici considéré comme jaugé afin de pouvoir utiliser les deux méthodes d'estimation $\widehat{p}_{\mathrm{np}}$ et $\widehat{p}_{\mathrm{cond}}$.

Pour les estimateurs $\widehat{p}_{\mathrm{HKEVP}}$ et $\widehat{p}_{\mathrm{Log}}$, trois approches sont considérées :

1. un ajustement classique en appliquant la méthode d'inférence de Thibaud et Opitz (2015) comme si les données étaient indépendantes,
2. en opérant au préalable une sélection des ℓ-excès comme décrit dans la section 8.4.3 avec le *run declustering* de Smith (1989) comme choix de méthode de declustering avec un paramètre de run τ ; on note ces estimateurs $\widehat{p}^{(\mathrm{Run})}_{\mathrm{HKEVP}}$ et $\widehat{p}^{(\mathrm{Run})}_{\mathrm{Log}}$ respectivement,
3. en sélectionnant les excès comme précédemment, mais cette fois en utilisant le block declustering de Tawn (1988) avec un temps caractéristique de τ ; ces méthodes sont respectivement notées $\widehat{p}^{(\mathrm{Block})}_{\mathrm{HKEVP}}$ et $\widehat{p}^{(\mathrm{Block})}_{\mathrm{Log}}$.

Le paramètre τ choisi lors de l'utilisation des méthodes de run declustering et block declustering dépend des données générées. Plus précisément, on regarde sur la figure 9.16 l'estimation du coefficient d'auto-corrélation estimé sur un ensemble de réalisations du processus évalué en un site, et ce pour chaque générateur utilisé. Le paramètre du declustering τ est alors choisi comme le plus petit espace de temps où le coefficient d'auto-corrélation se situe dans l'intervalle $[-0.05, 0.05]$.

Pour les modèles auto-régressifs $\mathcal{G}_{\mathrm{AR-Student}}(\varphi)$ et $\mathcal{G}_{\mathrm{ARMAX}}(\varphi)$, on choisit $\tau = 4, 8$ et 26 pour les valeurs respectives de $\phi = 0.5, 0.7$ et 0.9. Pour les générateurs de données cumulées, le paramètre τ est égal à la période h sur laquelle les données sont cumulées. En effet, par construction de ces données, deux observations X_i et X_{i+h+1} sont indépendantes puisqu'elles utilisent des sommes de variables $\{\tilde{X}_i\}_{i\geqslant 1}$ disjointes et indépendantes.

Par souci de cohérence avec les trois plans de simulation réalisés dans la section 9.1, on génère des processus spatio-temporels sur une durée de 5278 jours et on répète la procédure $N = 100$ fois pour obtenir des échantillons d'estimations $\widehat{p}_{A_C|A_S}$.

Les MSE obtenues sont données dans la table 9.4, avec le même code couleur indiquant, par générateur, la meilleure méthode en vert et la moins bonne en rouge. Les échantillons complets sont une fois encore présentés sous la forme de boîtes à moustaches sur les figures 9.12, 9.13, 9.14 et 9.15 pour des données simulées avec les processus $\mathcal{G}_{\mathrm{AR-Student}}$, $\mathcal{G}_{\mathrm{ARMAX}}$, $\mathcal{G}_{\mathrm{DC-HKEVP}}$ et $\mathcal{G}_{\mathrm{DC-Student}}$ respectivement.

On s'intéresse d'abord à l'impact de la méthode de sélection des excès sur la qualité des estimations par les modèles paramétriques, puis aux résultats obtenus par les méthodes non paramétriques.

La MSE est généralement sensiblement la même entre les deux estimateurs $\widehat{p}_{\mathrm{HKEVP}}$ et $\widehat{p}_{\mathrm{Log}}$ lorsqu'aucune méthode de sélection d'excès de seuil n'est appliquée. En regardant les figures 9.13, 9.12 et 9.15, on voit que l'estimation de la probabilité d'échec conditionnelle est presque systématiquement sous-estimée. L'estimateur construit sur le modèle spatial HKEVP semble légèrement plus précis que celui défini à partir du modèle logistique, ce qui corrobore les conclusions de la section 9.1.4.

En se servant de la méthode de sélection des ℓ-excès présentée dans la section 8.4.3, la MSE est systématiquement plus faible que celle obtenue par $\widehat{p}_{\mathrm{HKEVP}}$ et $\widehat{p}_{\mathrm{Log}}$. L'intérêt d'utiliser les méthodes de declustering est donc visible par ce plan de simulation. Il semble que le run declustering (estimateurs $\widehat{p}^{(\mathrm{Run})}_{\mathrm{HKEVP}}$ et $\widehat{p}^{(\mathrm{Run})}_{\mathrm{Log}}$) soit généralement plus efficace en montrant une MSE plus faible que dans le cas du block declustering (estimateurs $\widehat{p}^{(\mathrm{Block})}_{\mathrm{HKEVP}}$ et $\widehat{p}^{(\mathrm{Block})}_{\mathrm{Log}}$).

Cependant, si les données proviennent du générateur $\mathcal{G}_{\mathrm{DC-HKEVP}}(h)$, la différence est moins flagrante. La MSE reste faible même lorsqu'aucune procédure de declustering n'est utilisée. Si le cumul de données est fait sur une période de $h = 10$ jours, réaliser une sélection d'excès semble même déteriorer les résultats.

		$\widehat{p}_{\text{HKEVP}}$	$\widehat{p}^{(\text{Run})}_{\text{HKEVP}}$	$\widehat{p}^{(\text{Block})}_{\text{HKEVP}}$	$\widehat{p}_{\text{Log}}$	$\widehat{p}^{(\text{Run})}_{\text{Log}}$	$\widehat{p}^{(\text{Run})}_{\text{Log}}$	$\widehat{p}_{\text{np}}$	$\widehat{p}_{\text{cond}}$
$\mathcal{G}_{\text{AR}-\text{Student}}$	0.5	6.3523	4.0976	4.8365	7.0545	4.5182	5.3712	**0.309**	27.846
	0.7	6.915	2.7142	3.8614	7.6171	2.9809	4.2544	**0.0956**	24.8597
	0.9	7.0953	0.7135	1.6034	7.7622	0.831	1.7841	**0.0815**	22.0042
$\mathcal{G}_{\text{ARMAX}}$	0.5	4.1731	1.6962	2.1813	5.5772	2.4392	3.0498	**0.871**	15.5102
	0.7	4.1556	**0.7129**	1.1945	5.54	1.0703	1.6831	0.9534	18.5737
	0.9	4.2283	0.2737	0.3098	5.5655	**0.2418**	0.3582	1.3151	40.8987
$\mathcal{G}_{\text{DC}-\text{HKEVP}}$	3	0.934	**0.066**	0.1313	1.4651	0.1588	0.2585	5.4229	11.5425
	5	0.4335	0.193	0.0832	0.7751	0.0958	**0.035**	7.0866	7.075
	10	**0.0445**	2.0183	1.4471	0.122	1.7529	1.2438	10.7162	12.5067
$\mathcal{G}_{\text{DC}-\text{Student}}$	3	6.693	2.4888	2.8417	7.3967	2.7757	3.1092	**0.1576**	20.9042
	5	6.7512	1.6756	2.0748	7.4315	1.8906	2.2702	**0.0978**	21.6684
	10	6.6755	0.7808	1.119	7.2998	0.9166	1.2427	**0.1581**	21.5766

TABLE 9.4 – MSE sur l'estimation de la probabilité d'échec conditionnelle avec des données temporellement corrélées.

Si l'on compare maintenant les six premières colonnes du tableau 9.4 avec les deux dernières qui correspondent aux méthodes non paramétriques, on se rend compte que malgré l'amélioration apportée par les méthodes de sélection d'excès, $\widehat{p}_{\text{np}}$ reste l'un des meilleurs pour la plupart des configurations. Le seul cas où cette méthode rencontre des difficultés est lorsque les données proviennent de $\mathcal{G}_{\text{DC}-\text{HKEVP}}(h)$, mais on a vu que la sélection de seuil ne fournissait pas non plus d'amélioration dans ce cas.

De son côté, $\widehat{p}_{\text{cond}}$ montre des résultats similaires aux études réalisées sur des données indépendantes : les estimations sont souvent sous-estimées ou ont une variance très élevée. Bien que peu coûteuse en temps de calcul, il apparaît déconseillé en pratique d'utiliser cette méthode d'estimation de la probabilité d'échec conditionnelle dans un tel contexte.

9.3 Carte de probabilité d'échec

On terminera ce travail par une illustration, en appliquant les méthodes d'estimation de la probabilité d'échec conditionnelle aux données réelles de précipitations décrites dans le chapitre 4 et plus particulièrement à la région $\mathcal{S}$ du centre-est de la France. Dans cette section, on ne conserve que les 17 séries les plus complètes parmi les 61 disponibles.

De plus, seules les données journalières automnales (mois de septembre, octobre et novembre) sont considérées suite aux remarques faites dans le chapitre 4 sur cette saison, qui comporte des valeurs extrêmes en général plus importantes.

L'objectif fixé est ici d'estimer les probabilités d'échec $\{p_{\{c_j\}|A_S}(x_{100})\}_{j=1,\dots,d_c}$ pour un ensemble de positions cibles $A_C := \{c_j\}_{j=1,\dots,d_c}$. On souhaite ainsi produire une carte de probabilités d'échec conditionnelles qui peut être traduite comme le risque qu'une valeur extrême apparaisse en la position évaluée sachant qu'une même valeur extrême est apparue sur un réseau de stations météorologiques.

La carte de probabilités est produite en utilisant l'estimateur $\widehat{p}_{\text{HKEVP}}$ car il s'agit de la seule méthode qui vérifie les trois conditions suivantes :

1. Elle permet d'extrapoler la probabilité d'échec, c'est-à-dire pour des positions non jaugées, ce qui n'est pas le cas de $\widehat{p}_{\text{np}}$ et $\widehat{p}_{\text{cond}}$.
2. Sa demande en temps de calcul est raisonnable, ce qui élimine les estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$.
3. Elle possède une structure spatiale, contrairement à $\widehat{p}_{\text{Log}}$.

Afin d'associer une erreur aux estimations $\{\widehat{p}_{\{c_j\}|S}\}_{j=1,\dots,d_c}$, les paramètres (α, τ) de la fonction exposante du modèle HKEVP sont estimés dans le cadre bayésien (cf. Chapitre 3) en utilisant la vraisemblance de Thibaud et Opitz (2015) (cf. Section 8.2.1).

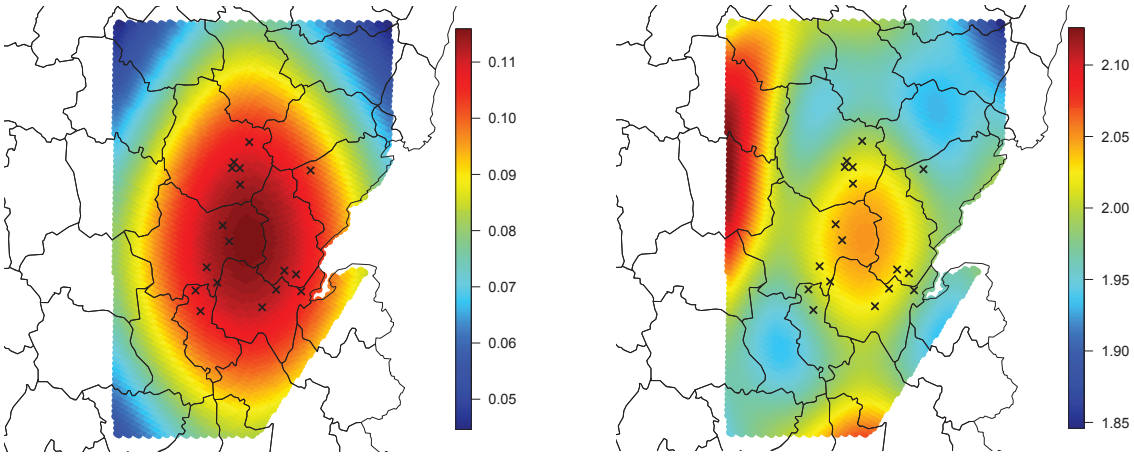
Les lois a priori choisies sur les paramètres α et τ sont équivalentes à celles définies dans le chapitre 6 pour l'inférence du modèle HKEVP sur des maxima par blocs. Autrement dit, si $D_{\max}$ est la distance maximale observée entre deux stations du réseau disponible, on définit les lois a priori :

$$\alpha \sim \text{Unif}\big([0,1]\big) \quad \text{et} \quad \frac{\tau}{2D_{\max}} \sim \text{Beta}(2,5) \ .$$

Un algorithme MCMC de type Metropolis-within-Gibbs (cf. Section 3.2) est mis en place pour simuler des chaînes de Markov $(\alpha_i, \tau_i)_{i=1,\dots,N}$ dont la loi stationnaire est la loi a posteriori $\pi(\alpha, \tau|x)$. La procédure est

utilisée pendant 20.000 itérations, les 5.000 premières sont supprimées en tant que période de chauffe et un thinning de taille 15 est appliqué pour effacer la corrélation entre les états de la chaîne (cf. Section 3.2.3).

La médiane de la loi a posteriori obtenue en chaque position cible de la région d'intérêt $\mathcal{S}$ est affichée sur la figure 9.1a. L'erreur associée à cette estimation ponctuelle est représentée sur la figure 9.1b par l'écart-type a posteriori de chaque cible. Par souci de lisibilité, l'écart-type affiché a été multiplié par 1000.



(a) Médiane a posteriori.

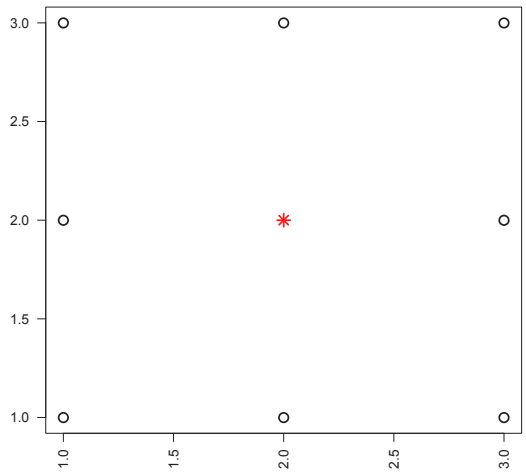
(b) Écart-type a posteriori multiplié par 1000.

FIGURE 9.1 – Carte de médianes (a) et écarts-types (a) a posteriori obtenus pour l'estimation de la probabilité d'échec par le HKEVP sur des données réelles de précipitations.

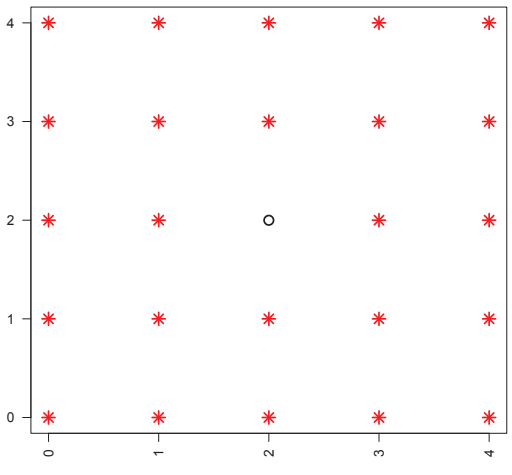
Ces deux cartes montrent que la probabilité conditionnelle d'observer une valeur extrême est plus forte au centre de la région formée par les 17 stations météorologiques de l'ensemble A_S (environ 11%). À mesure que l'on s'éloigne de cette région, cette probabilité décroît.

Les écart-types associés à l'estimation de la probabilité d'échec conditionnelle sont faibles et ont dû être multipliés par 1000 sur la figure 9.1b. On remarque que le motif spatial de cette variabilité d'estimation est lisse mais atypique : les valeurs les plus élevées de cette variabilité sont présentes sur le bord ouest et le bord sud de la région, ainsi qu'au centre, où l'estimation de la probabilité d'échec est la plus forte. Les valeurs les plus faibles sont visibles sur la bordure nord-est de la région étudiée.
Cependant, la différence entre l'écart-type le plus élevé et l'écart-type le plus bas sur tout la région est elle-aussi très petite : la variabilité d'estimation peut donc être considérée comme quasi constante sur $\mathcal{S}$.

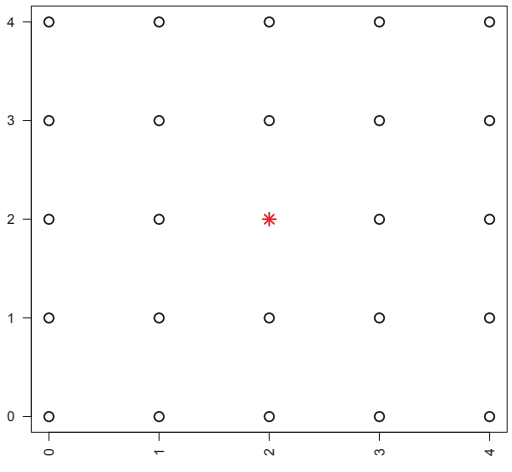
Figures associées aux plans de simulation du chapitre 9



(a) Disposition 1 : petite région

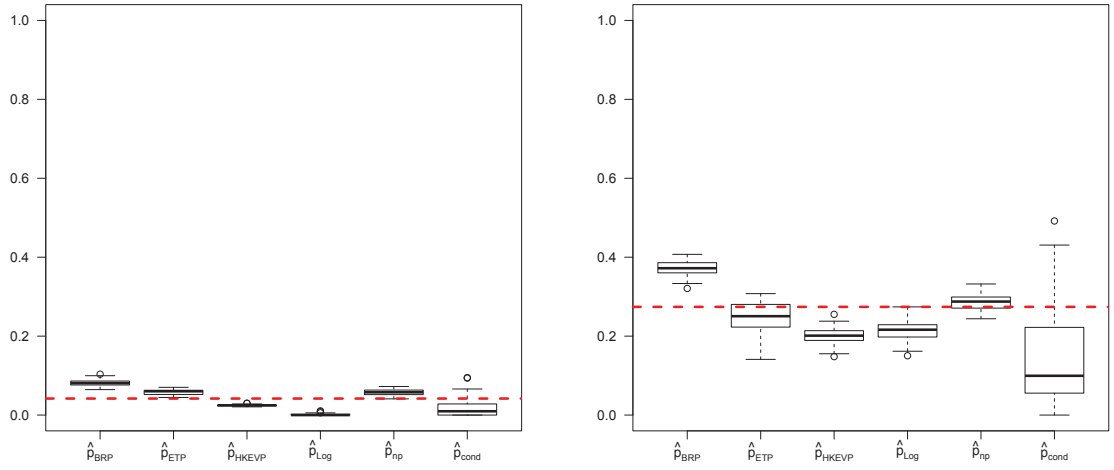


(b) Disposition 2 : une seule station



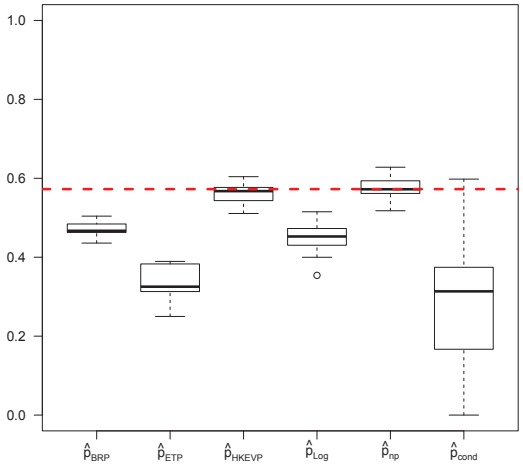
(c) Disposition 3 : un seul site cible

FIGURE 9.2 – Configuration des ensembles de cibles (étoiles) et de stations (cercles) utilisés dans les plans de simulations du chapitre 9.



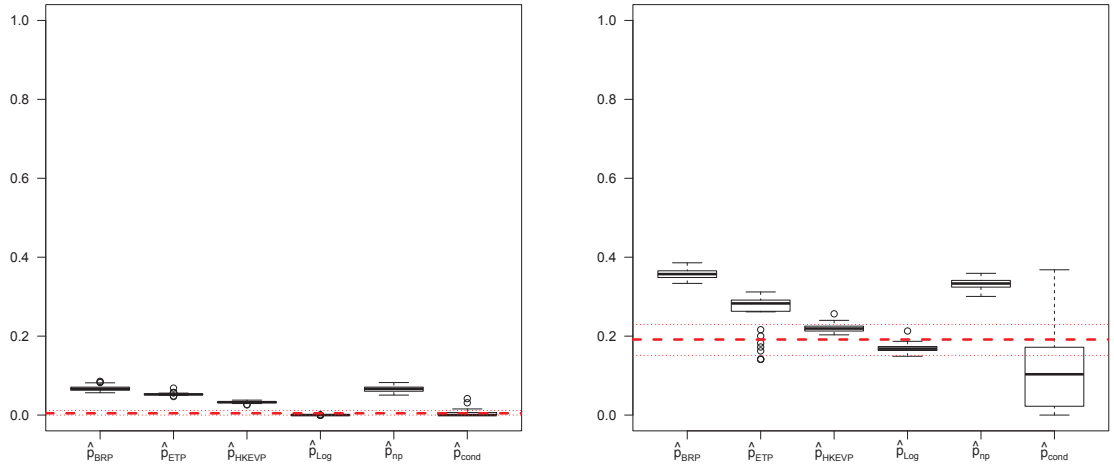
(a) $\alpha = 0.1$

(b) $\alpha = 0.5$

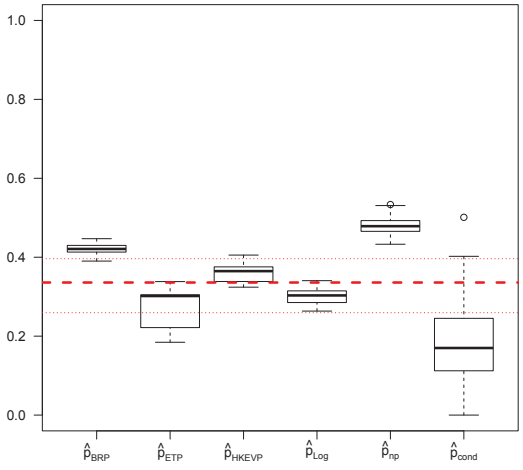


(c) $\alpha = 0.9$

FIGURE 9.3 – Estimation de la probabilité d'échec conditionnelle sur des simulations par $\mathcal{G}_{\text{HKEVP}}$.

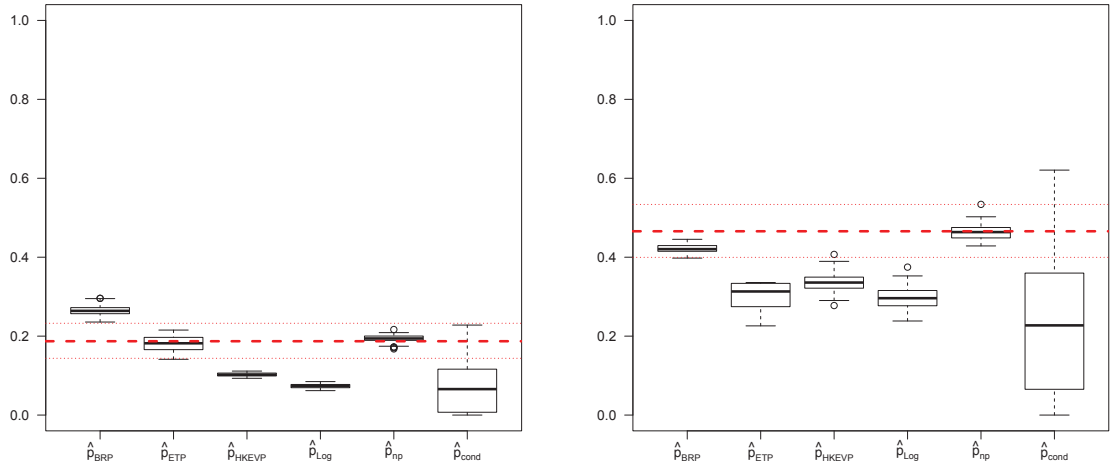


(a) $\lambda = 1$ (b) $\lambda = 5$

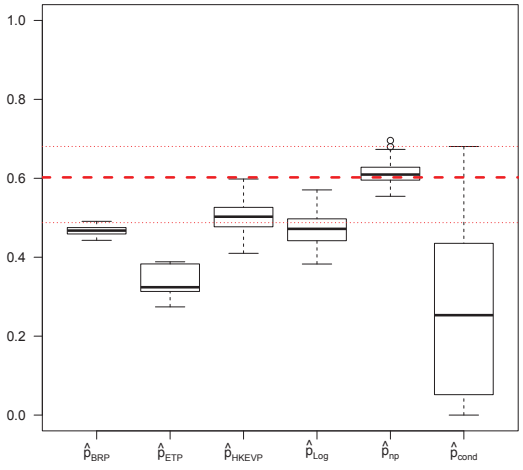


(c) $\lambda = 10$

FIGURE 9.4 – Estimation de la probabilité d'échec conditionnelle sur des simulations par $\mathcal{G}_{\text{Gauss}}$.

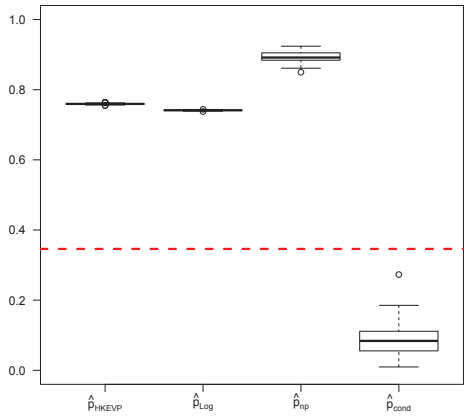


(a) $\lambda = 1$ (b) $\lambda = 5$

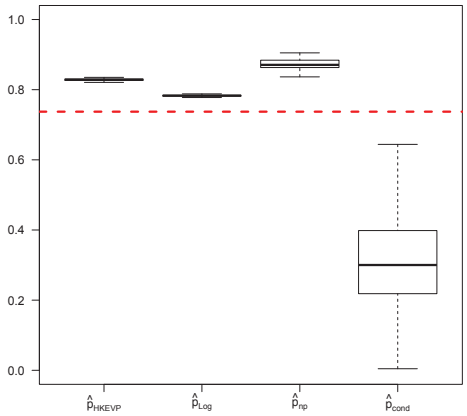


(c) $\lambda = 10$

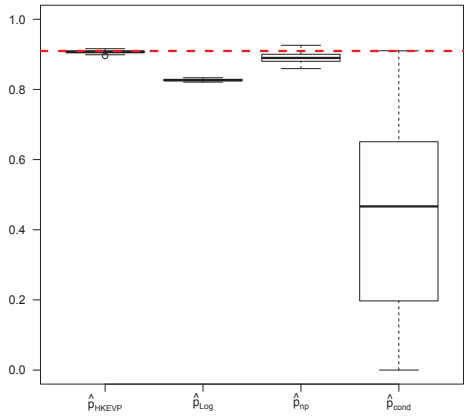
FIGURE 9.5 – Estimation de la probabilité d'échec conditionnelle sur des simulations par $\mathcal{G}_{\text{Student}}$.



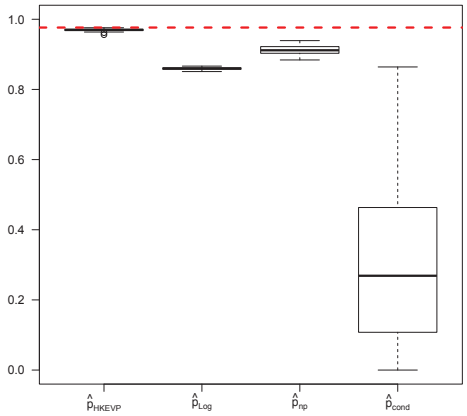
(a) $\alpha = 0.9$



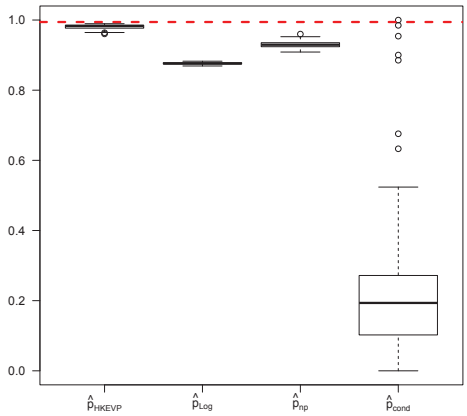
(b) $\alpha = 0.7$



(c) $\alpha = 0.5$

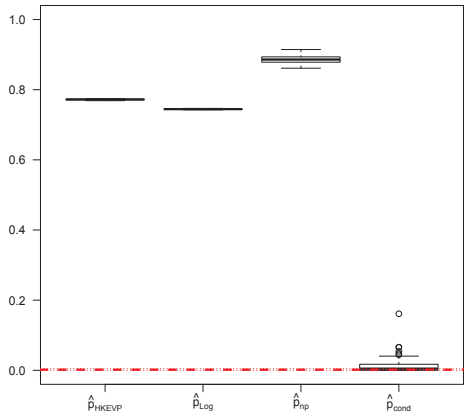


(d) $\alpha = 0.3$

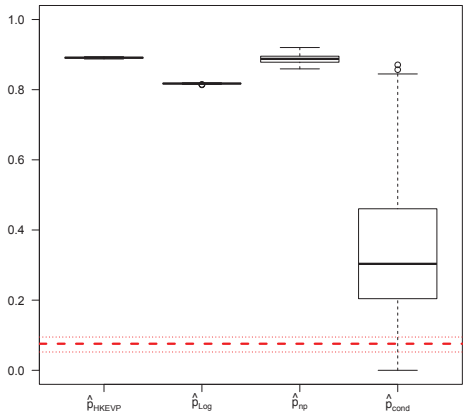


(e) $\alpha = 0.1$

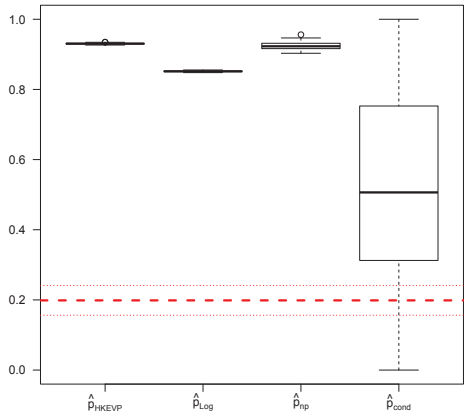
FIGURE 9.6 – Estimation de la probabilité de contagion sur des simulations par $\mathcal{G}_{\text{HKEVP}}$.



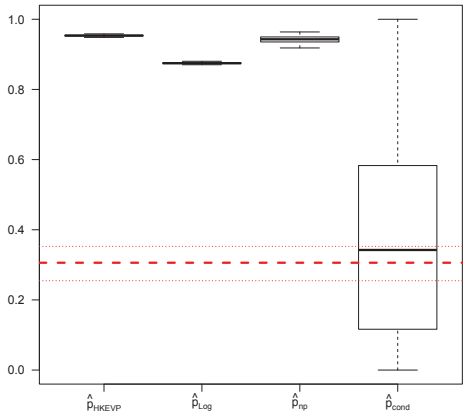
(a) $\lambda = 1$



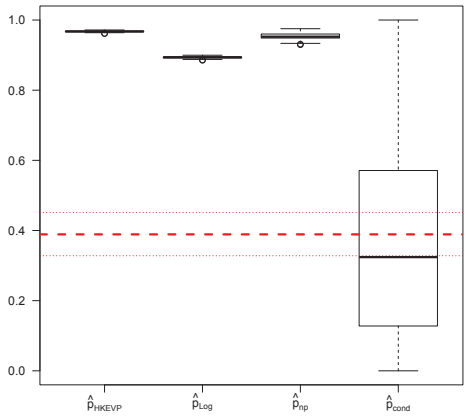
(b) $\lambda = 3$



(c) $\lambda = 5$

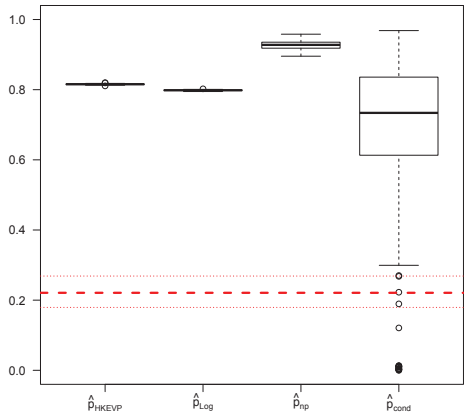


(d) $\lambda = 7$

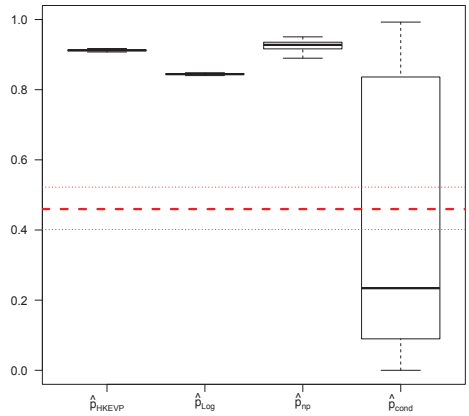


(e) $\lambda = 9$

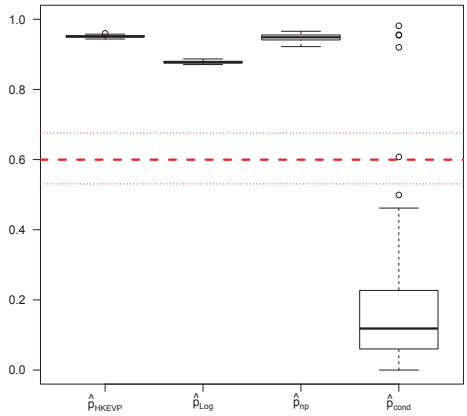
FIGURE 9.7 – Estimation de la probabilité de contagion sur des simulations par $\mathcal{G}_{\text{Gauss}}$.



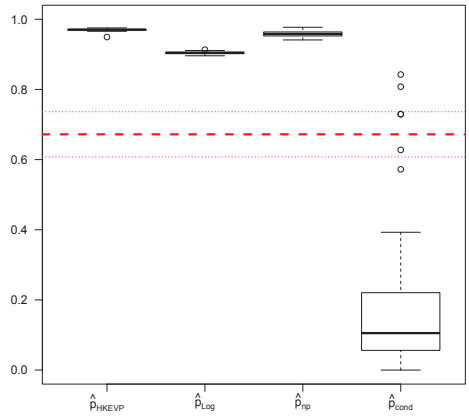
(a) $\lambda = 1$



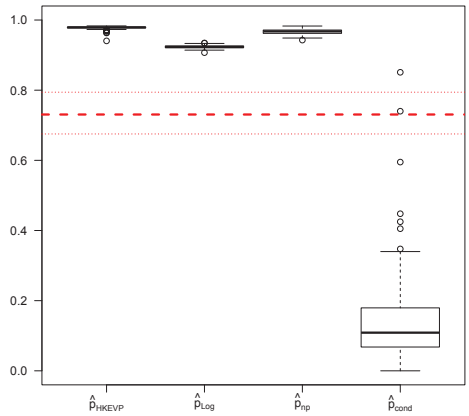
(b) $\lambda = 3$



(c) $\lambda = 5$

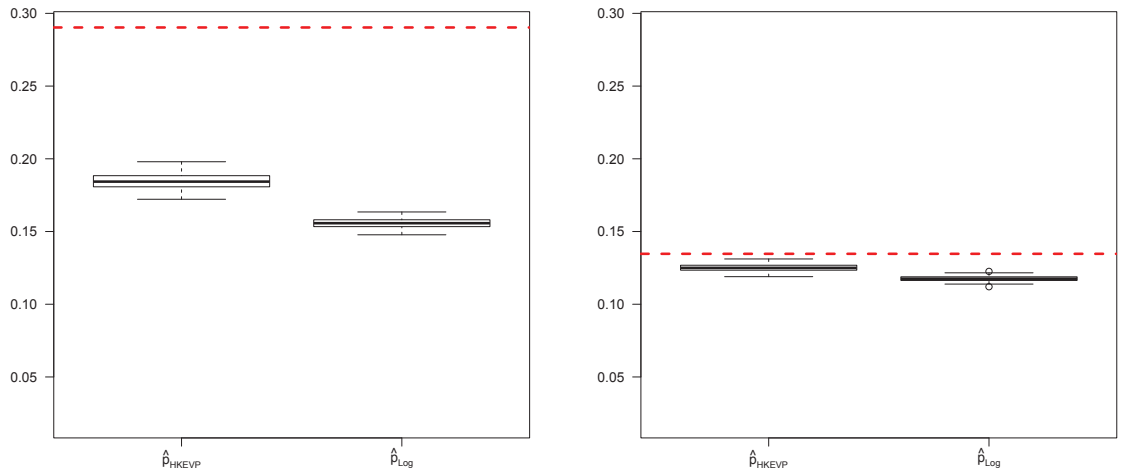


(d) $\lambda = 7$



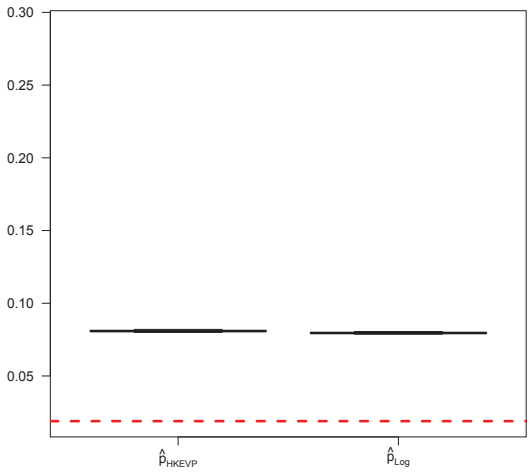
(e) $\lambda = 9$

FIGURE 9.8 – Estimation de la probabilité de contagion sur des simulations par $\mathcal{G}_{\text{Student}}$.



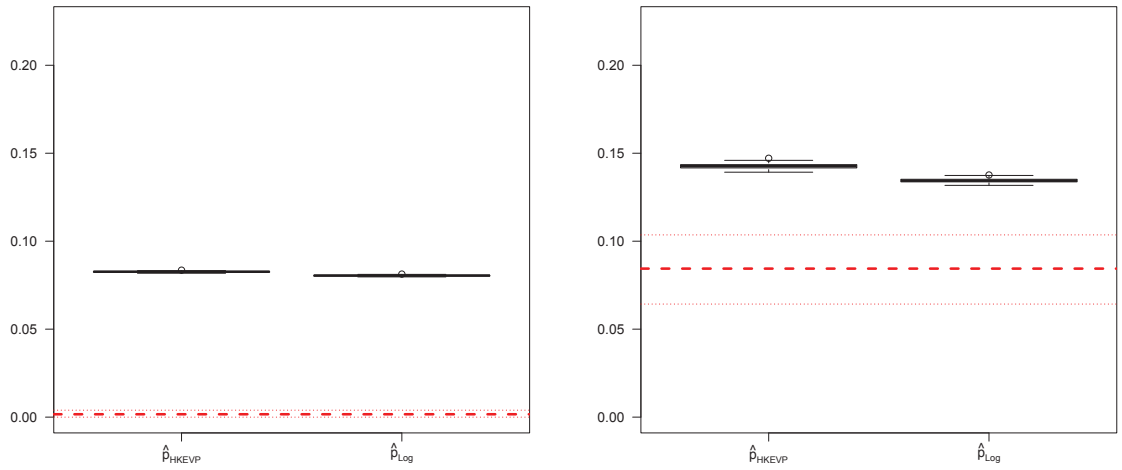
(a) $\alpha = 0.1$

(b) $\alpha = 0.5$



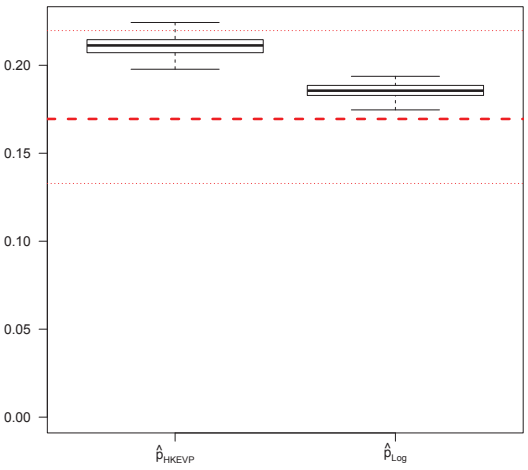
(c) $\alpha = 0.9$

FIGURE 9.9 – Estimation de la probabilité d'échec conditionnelle en une position non jaugée sur des simulations par $\mathcal{G}_{\text{HKEVP}}$.



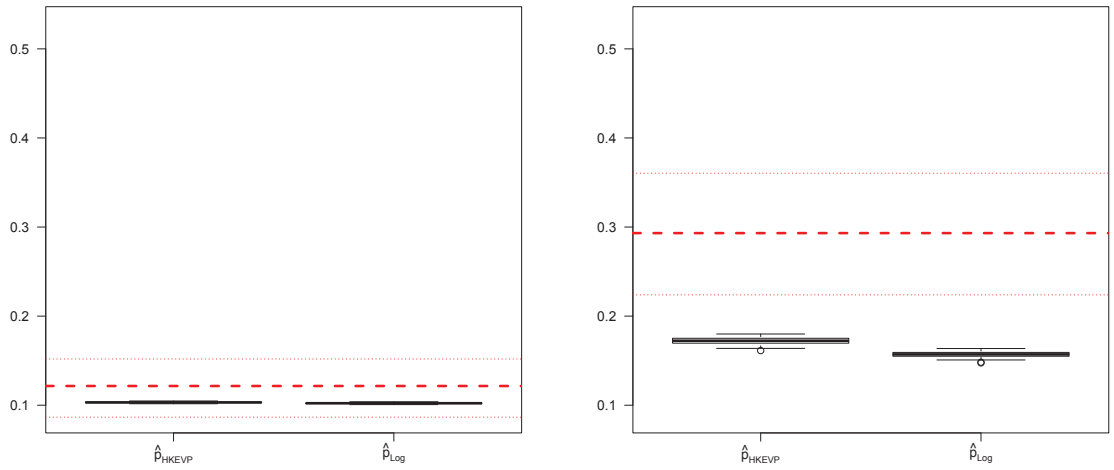
(a) $\lambda = 1$

(b) $\lambda = 5$



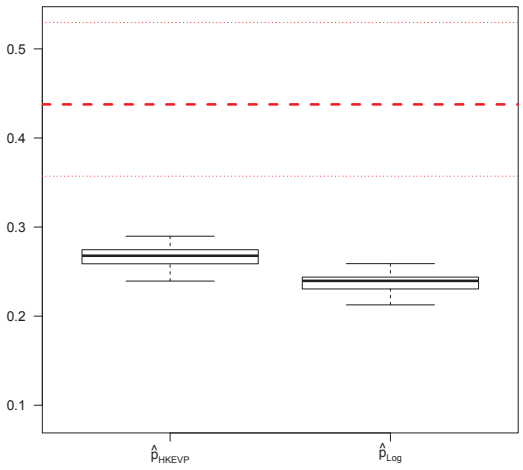
(c) $\lambda = 10$

FIGURE 9.10 – Estimation de la probabilité d'échec conditionnelle en une position non jaugée sur des simulations par $\mathcal{G}_{\text{Gauss}}$.



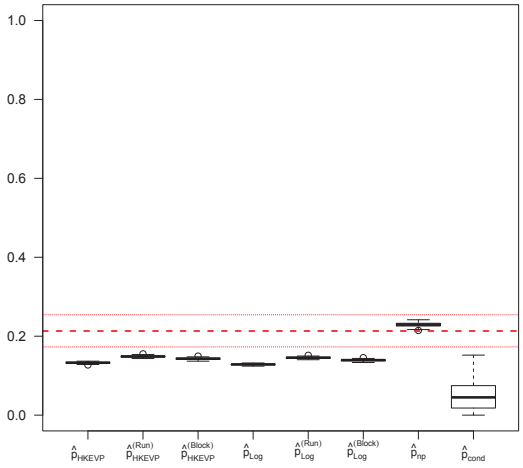
(a) $\lambda = 1$

(b) $\lambda = 5$

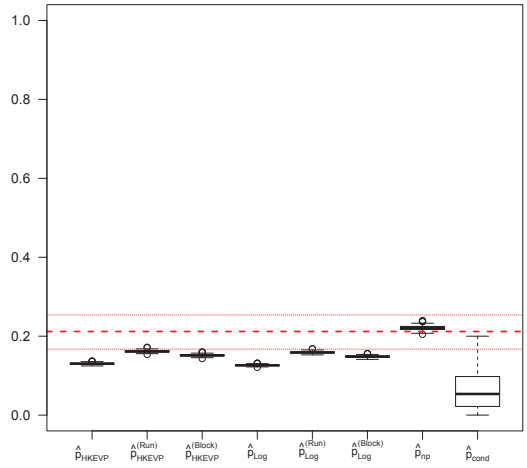


(c) $\lambda = 10$

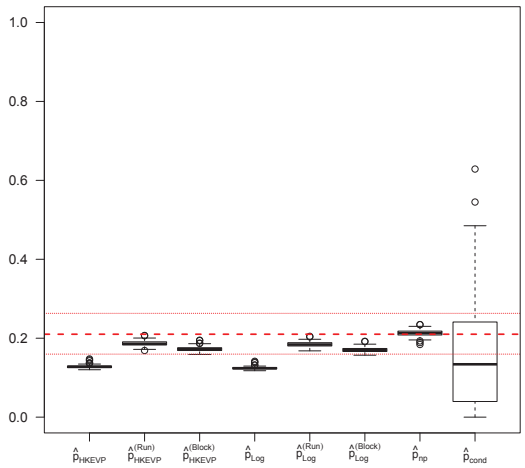
FIGURE 9.11 – Estimation de la probabilité d'échec conditionnelle en une position non jaugée sur des simulations par $\mathcal{G}_{\text{Student}}$.



(a) $\varphi = 0.5$

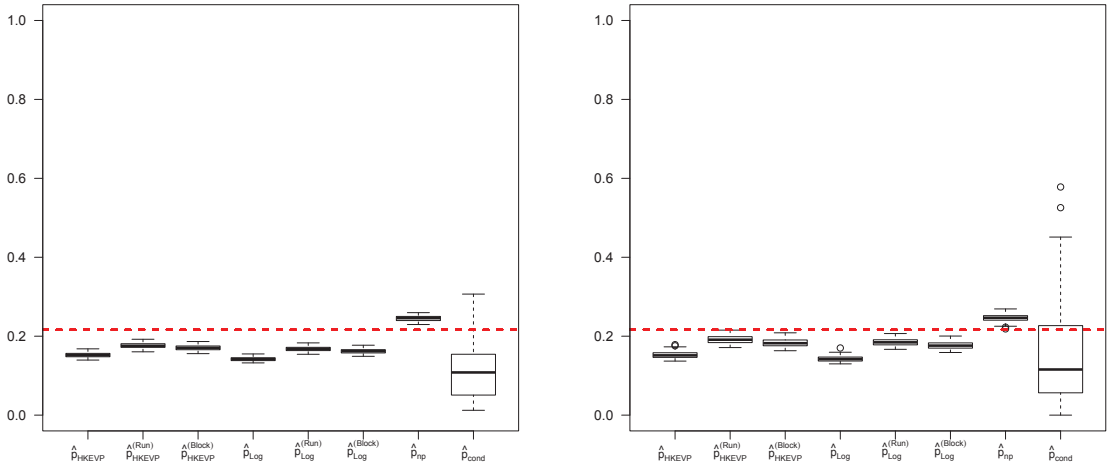


(b) $\varphi = 0.7$



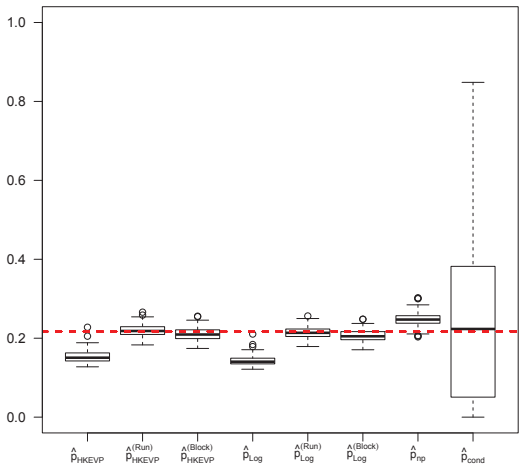
(c) $\varphi = 0.9$

FIGURE 9.12 – Estimation de la probabilité d'échec conditionnelle sur des données temporellement corrélées simulées par $\mathcal{G}_{\mathrm{AR-Student}}$.



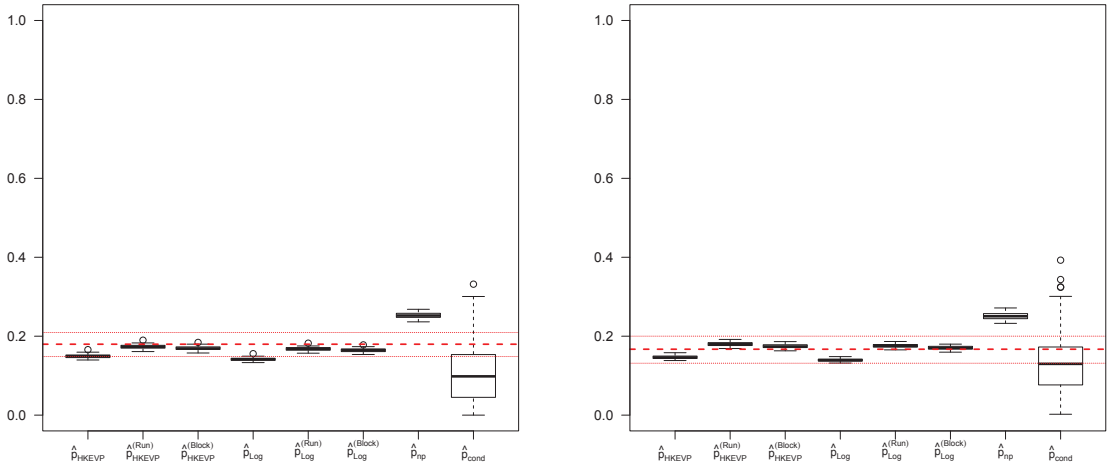
(a) $\varphi = 0.5$

(b) $\varphi = 0.7$



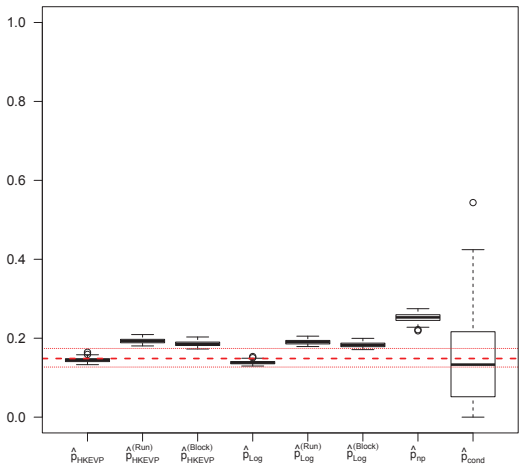
(c) $\varphi = 0.9$

FIGURE 9.13 – Estimation de la probabilité d'échec conditionnelle sur des données temporellement corrélées simulées par $\mathcal{G}_{\text{ARMAX}}$.



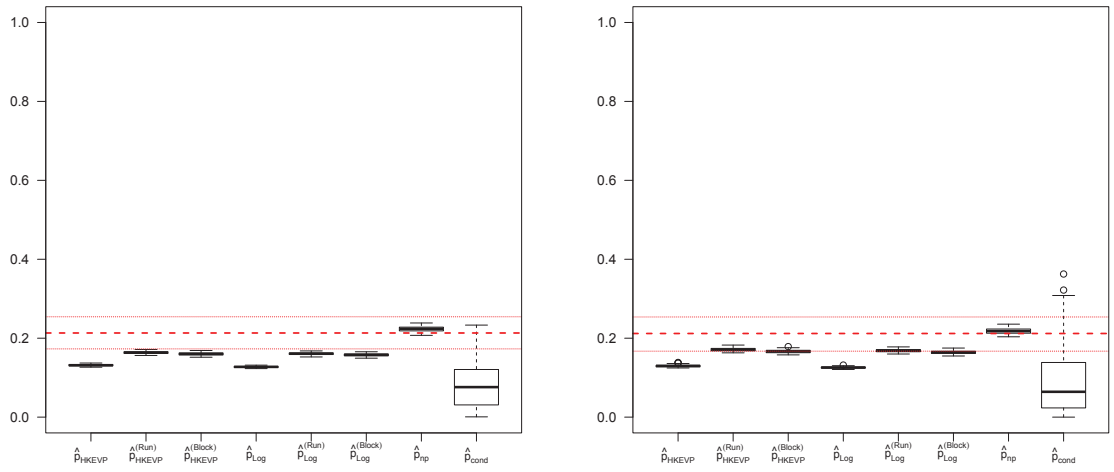
(a) Cumul sur 3 jours

(b) Cumul sur 5 jours



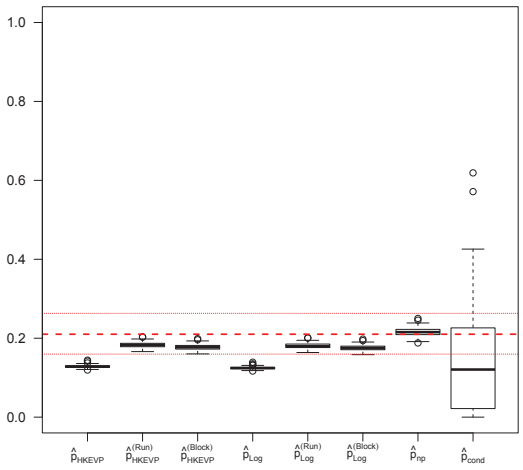
(c) Cumul sur 10 jours

FIGURE 9.14 – Estimation de la probabilité d'échec conditionnelle sur des données temporellement corrélées simulées par $\mathcal{G}_{\mathrm{DC-HKEVP}}$.



(a) Cumul sur 3 jours

(b) Cumul sur 5 jours



(c) Cumul sur 10 jours

FIGURE 9.15 – Estimation de la probabilité d'échec conditionnelle sur des données temporellement corrélées simulées par $\mathcal{G}_{\mathrm{DC-Student}}$.

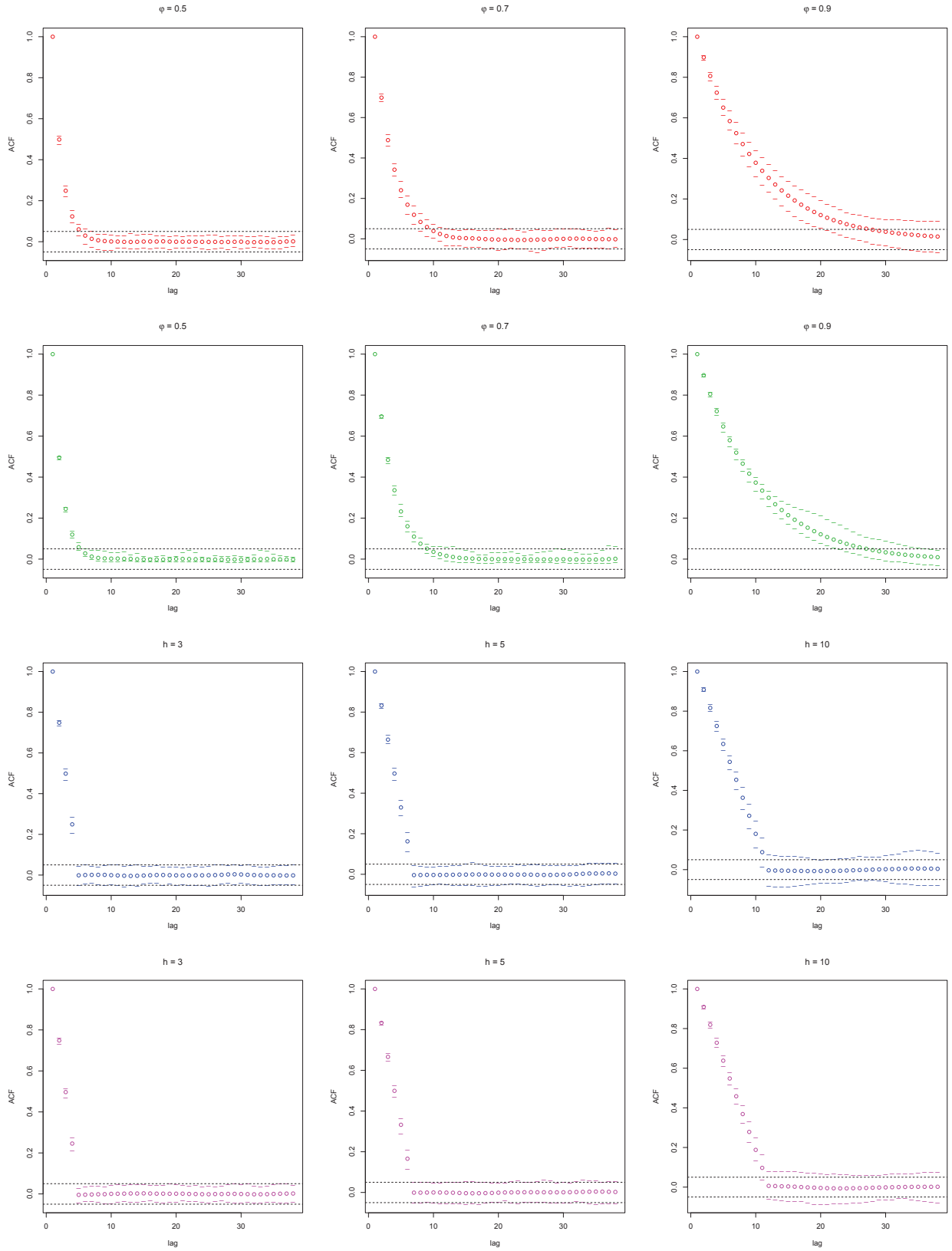


FIGURE 9.16 – Estimations du coefficient de corrélation sur des données en une position, provenant des générateurs $\mathcal{G}_{\mathrm{AR-Student}}$ (première ligne, en rouge), $\mathcal{G}_{\mathrm{ARMAX}}$ (deuxième ligne, en vert), $\mathcal{G}_{\mathrm{DC-HKEVP}}$ (troisième ligne, en bleu) et $\mathcal{G}_{\mathrm{DC-Student}}$ (quatrième ligne, en rose), en fonction de la distance temporelle entre deux observations.

Conclusions

Travaux réalisés

Comparaison de modèles max-stables

La partie II s'est appuyée sur la modélisation spatiale des valeurs extrêmes par les processus max-stables, seuls modèles limites possibles des maxima renormalisés de processus stochastiques continus. Après une présentation des modèles usuels définis dans la littérature, le chapitre 5 a fourni une étude comparative sous forme d'un article.

Une des contributions de cette thèse a été l'inclusion du modèle max-stable hiérarchique de Reich et Shaby (2012) dans une compétition avec à la fois des modèles max-stables fréquemment utilisés dans les applications et un modèle à variables latentes (Davison *et al.*, 2012) construit dans le but d'une meilleure modélisation spatiale des paramètres marginaux μ, σ et ξ de la loi GEV. Ce modèle a été présenté en détail avec la procédure d'inférence associée dans le chapitre 6 et a été implémenté au sein d'un package codé sur R et dont le manuel est disponible en annexe.

Étude comparative sur critères

L'objectif de l'étude comparative a été d'évaluer les capacités d'ajustement des modèles spatiaux sur des critères ayant un intérêt applicatif. Plusieurs conclusions sont ressorties de cette comparaison :

- Pour l'extrapolation d'un niveau de retour, le modèle à variables latentes de Davison *et al.* (2012) semble le choix le plus satisfaisant vis-à-vis de l'extrapolation des lois marginales. Le modèle ETP de Opitz (2013) a aussi montré de bonnes performances.
- Pour l'estimation du coefficient extrémal, le HKEVP de Reich et Shaby (2012) a fait preuve d'une très bonne robustesse vis-à-vis du type de données simulées, mais les modèles ETP et BRP (Brown et Resnick, 1977; Kabluchko *et al.*, 2009) ont également affiché des résultats satisfaisants.
- Quant au modèle EGP de Schlather (2002), l'étude comparative du chapitre 5 l'a placé comme le moins performant des modèles spatiaux de valeurs extrêmes, que ce soit pour l'extrapolation du niveau de retour où la MSE est généralement très élevée, ou pour l'estimation du coefficient extrémal qui est affecté par une limite théorique due à sa construction.

Estimation d'une probabilité d'échec

La partie III s'est concentrée sur les modèles adaptés aux excès de seuil d'un processus journalier. Dans un premier temps, le chapitre 7 a recensé les récents travaux de recherches portant sur cette question. Plusieurs définitions possibles pour un dépassement de seuil d'un processus spatial ont été généralisées par Dombry et Ribatet (2015) à travers les processus ℓ-Pareto.

Dans un second temps, l'objectif fixé a été l'estimation d'une probabilité d'échec conditionnelle dans le chapitre 8. Plusieurs méthodes d'estimations ont été mises en place, construites à la fois sur des approches paramétriques et non paramétriques. Les études sur simulation menées dans le chapitre 9 ont permis de comparer six estimateurs de la probabilité d'échec conditionnelle. Une méthode de sélection d'excès dans le cas d'observations multivariées a été proposée dans la section 8.4.3 et son intérêt a été illustré sur des simulations de données corrélées dans la section 9.2.

Les résultats des études menées dans la section 9.1 sur l'estimation de la probabilité d'échec conditionnelle (8.1) ont permis de tirer plusieurs conclusions que nous résumons ici :

Estimateur non paramétrique $\widehat{p}_{\mathrm{np}}$

L'estimateur non paramétrique (8.14) de la probabilité d'échec est l'une des méthodes les plus compétitives. Cette approche a fait état d'une forte robustesse vis-à-vis du type de données simulées d'une part, et de la

configuration du plan de simulations (par exemple, la disposition des stations et des cibles) d'autre part. De plus, cette méthode est très simple et demande très peu de calculs. La seule limitation de $\widehat{p}_{\text{np}}$ est que les ensembles des stations et des cibles doivent contenir des données pour qu'elle puisse être utilisée. Cette caractéristique empêche par exemple de produire une carte de probabilités conditionnelles (voir section 9.3).

Estimateurs $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$

Les méthodes paramétriques $\widehat{p}_{\text{BRP}}$ et $\widehat{p}_{\text{ETP}}$ construites avec la procédure d'inférence de Thibaud et Opitz (2015) et utilisant la mesure exposante des modèles BRP et ETP n'ont pas été conservées en raison d'une forte demande en temps de calcul et de résultats peu compétitifs sur l'étude menée dans la section 9.1.2. Dans ce plan de simulations, seules 9 positions ont été considérées et ces deux approches restaient relativement chronophages : environ une journée, et jusqu'à une semaine si le nombre de sites passe à 16. Ce défaut est dû au calcul coûteux de fonctions de répartitions gaussiennes ou Student lors de l'évaluation des dérivées partielles de la fonction exposante V (cf. (8.9) et (8.11)). De plus, les estimations de la probabilité d'échec obtenues par ces méthodes dans la section 9.1.2 n'ont pas justifié de considérer ces deux approches dans la suite de l'étude.

Estimateurs $\widehat{p}_{\text{HKEVP}}$ et $\widehat{p}_{\text{Log}}$

Le modèle HKEVP de Reich et Shaby (2012) a permis une inférence sur les dépassements de seuil spatiaux avec un temps de calcul raisonnable, alors qu'il apparaissait comme le modèle le plus demandeur dans la comparaison du chapitre 5 sur des données de maxima annuels. La raison à cela est que la formulation hiérarchique de ce modèle, détaillée dans le chapitre 6, nécessite l'estimation d'un grand nombre de paramètres avec un algorithme MCMC : entre autres, la valeurs des paramètres GEV en chaque site et la valeur de l'effet aléatoire A pour chaque nœud et chaque observation (année).

La procédure d'inférence de Thibaud et Opitz (2015) n'utilise que la fonction exposante et ses dérivées partielles, toutes explicites et calculables facilement d'après (8.12) et (8.13). L'estimateur $\widehat{p}_{\text{HKEVP}}$ a montré une assez bonne robustesse sur l'estimation de la probabilité d'échec, au même titre que $\widehat{p}_{\text{Log}}$, version plus simple mais qui ne possède pas d'aspect spatial.

Estimateur $\widehat{p}_{\text{cond}}$

L'estimateur $\widehat{p}_{\text{cond}}$ a été construit en utilisant le modèle conditionnel de Heffernan et Tawn (2004) sur les séries univariées :

$$X_{A_S} = \max_{s \in A_S} X(s) \text{ et } X_{A_C} = \max_{s \in A_C} X(s) \ .$$

Cette approche a montré une grande variabilité d'estimation tout au long des études du chapitre 9. Bien que plusieurs essais aient déjà échoué, on peut espérer qu'un perfectionnement dans le choix des paramètres de ce modèle (ordre des quantiles, nombre de réalisations conditionnelles simulées) puisse améliorer la méthode.

Probabilité de contagion

L'analyse menée dans la section 9.1.3 a révélé des estimations très loin de la réalité lorsque les rôles des stations et des cibles sont inversées, autrement dit lorsque l'on a cherché à calculer la probabilité de contagion d'un excès. D'autres simulations ont été réalisées en utilisant plusieurs stations et plusieurs cibles. Elles ne sont pas montrées dans le chapitre 9, mais il en est ressorti dans ce cas une perte de qualité pour tous les estimateurs de la probabilité d'échec.

Méthode de sélection des excès

La méthode de sélection de dépassements de seuil présentée dans la section 8.4.3 a été testée dans la section 9.2 sur des données temporellement asymptotiquement dépendantes.

Les résultats ont montré une amélioration de la précision d'estimation, surtout lorsqu'elle est associée à l'algorithme du *run declustering* de Smith (1989) décrit dans la section 8.4.1. Il a aussi été mis en évidence que l'estimateur non paramétrique (8.14) restait la plupart du temps la meilleure méthode d'estimation, même sur des données dépendantes.

Applications et perspectives

Les travaux présentés dans ce manuscrit ont permis d'évaluer différentes approches pour modéliser les valeurs extrêmes spatiales, que ce soit en regardant les valeurs maximales annuelles (partie II) ou les dépassements de seuil journaliers (partie III). Les recherches menées se sont concentrées sur le modèle de Reich et Shaby (2012) et ont permis de positionner ce dernier par rapport à plusieurs modèles fréquemment utilisés dans la littérature en pointant ses atouts et ses défauts :

- D'un côté, ce modèle permet d'estimer avec précision la structure de dépendance spatiale des valeurs extrêmes, et de fournir une réponse rapide de la probabilité d'échec conditionnelle étudiée dans la partie III. L'estimation de cette probabilité d'échec a été jugé assez satisfaisante avec cette approche vis-à-vis de données simulées, bien que l'estimateur non paramétrique $\widehat{p}_{\mathrm{np}}$ soit plus performant.
- D'un autre côté, il semble que cette approche ait tendance à surestimer la mesure de risque d'intérêt qu'est le niveau de retour centennal. De plus, l'inférence statistique liée à ce modèle peut devenir très coûteuse en temps de calcul lorsque ce dernier est ajusté à des valeurs maximales annuelles.

Les conclusions tirées dans les chapitres précédents fournissent des réponses à chacun des objectifs fixés en introduction. Les modèles recommandés varient en fonction de la mesure de risque que l'on souhaite estimer. Ces travaux peuvent en particulier être utiles à EDF pour contribuer au dimensionnement des ouvrages.

Si les aléas naturels étudiés dans ce manuscrit sont les précipitations, on peut faire remarquer que les modèles présentés peuvent être appliqués à tout type de données spatiales comme par exemple la force du vent, les températures ou les débits de rivière. Il est toutefois recommandé de procéder avec prudence, car il a été souligné que la forme de la dépendance des extrêmes joue un rôle important sur la qualité d'estimation des méthodes considérées.

Ces travaux peuvent faire l'objet d'approfondissements à l'avenir, parmi lesquels :

- réduire le temps de calcul nécessaire à certaines méthodes d'estimation,
- inclure d'autres approches de la littérature des extrêmes dans les études comparatives,
- étudier la sensibilité par rapport à la qualité du jeu de données de précipitations utilisé (par exemple le nombre d'années observées ou encore la parcimonie du réseau de stations).

Discussion post-réception des rapports

À la lumière des modifications suggérées par les rapporteurs, que je remercie chaleureusement pour leur lecture attentive, je souhaite compléter ici les perspectives issues de ce travail.

Durant ces trois années de thèse, EDF m'a offert l'opportunité d'étudier en détail les modèles spatiaux adaptés aux valeurs extrêmes. Une comparaison quantitative a mis en compétition la plupart des modèles existants et a permis la production de cartes de niveaux de retour (Figure 9 au chapitre 5 et figures 6.6 et 6.7 au chapitre 6) et de cartes de probabilités d'échec (Figure 9.1 au chapitre 9).

Une analyse qualitative de leur pertinence physique apporterait un complément d'information permettant d'asseoir plus encore les conclusions tirées dans ce manuscrit. Par manque de temps et pour mieux pouvoir se concentrer sur les méthodes traitant de dépassements de seuils spatiaux, cette analyse n'a pas été réalisée. L'expertise du groupe d'hydrologues d'EDF mise au service d'un telle examen ajouterait ainsi à cette étude un regard critique sur les risques estimés pour les précipitations. Les différents outils techniques (codes et rapports internes), qui ont été fournis durant cette collaboration, facilitera la mise en place de ce retour d'expérience.

Lors de ces recherches, une attention particulière portée au modèle HKEVP a mis en lumière deux caractéristiques qui mériteraient d'être investiguées : le problème de non-mélangeance du modèle pointé dans la section 6.2.4 et l'estimation du paramètre de forme ξ différent étonnamment de celles fournies par rapport aux autres approches.

- Une proposition intéressante est de générer les paramètres de dépendance α et τ conjointement grâce à une loi candidate bivariée judicieusement choisie. Cette technique pourrait permettre de résoudre le problème de mélangeance observé. Le package `hkevp` pourrait de plus produire simultanément plusieurs chaînes avec différents états initiaux et pour un même temps de calcul, en utilisant le calcul parallèle. Cette approche permettrait d'échantillonner entre les différentes réalisations indépendantes, toutes distribuées selon les lois a posteriori.
- Concernant l'estimation du paramètre de forme, il est envisageable d'analyser avec précision l'impact du paramètre α et du nombre de nœuds, comme étudié sur simulations par Reich et Shaby (2012).

Pour les cinq modèles spatiaux appliqués à des observations maximales annuelles, l'estimation du paramètre de forme ξ est faite en supposant que celui-ci est constant sur la région étudiée. Cette hypothèse simplificatrice, faite dans la plupart des articles de la littérature traitant de précipitations extrêmes, pourrait être remise en cause. On observe en effet une grande variabilité spatiale des estimations de ce paramètre (de -0.1 à 0.5 sur la région considérée d'après la figure 4.8c). L'étude comparative du chapitre 5 pourrait être reconsidérée en laissant ce paramètre varier spatialement selon le même modèle linéaire que celui utilisé pour les paramètres μ et σ. Une possibilité intermédiaire serait de définir des classes de valeurs équivalentes.

En menant cette analyse qualitative et quantitative, des arbitrages différents de modélisation pourraient motiver l'utilisation de méthodes alternatives, tels que les modèles max-stables inverses définis par Wadsworth et Tawn (2012). Ces processus, non investigués dans la thèse, permettent la modélisation de données exhibant une indépendance asymptotique spatiale. Elle est justifiée dans le cas de données de précipitation pour deux sites suffisamment éloignées l'un de l'autre. L'utilisation de modèles inverses est par exemple mise en avant par Davison *et al.* (2013) sur des données de précipitations en Suisse. Une manière de modéliser à la fois le cas de dépendance asymptotique (à une petite échelle) et celui de l'indépendance asymptotique (entre deux régions) est de s'appuyer sur des modèles hybrides, comme suggéré par Wadsworth et Tawn (2012). Un autre exemple illustrant cette approche est le modèle de Huser *et al.* (2016), construit sur un mélange de processus gaussiens et permettant de modéliser les deux types de dépendance asymptotique.

Bibliographie

Bibliographie

Abramowitz, M. et Stegun, I. A. (1964). *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F. *et al.* (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research : Atmospheres*, 111(D5).

Asadi, P., Davison, A. C. et Engelke, S. (2015). Extremes on river networks. *arXiv preprint arXiv :1501.02663*.

Bacro, J.-N. et Toulemonde, G. (2013). Measuring and modelling multivariate and spatial dependence of extremes. *Journal de la Société Française de Statistique*, 154(2):139–155.

Balkema, A. A. et De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, pages 318–355.

Bechler, A., Bel, L. et Vrac, M. (2015). Conditional simulations of the extremal t process : application to fields of extreme precipitation. *Spatial Statistics*, 12:109–127.

Beirlant, J., Goegebeur, Y., Segers, J. et Teugels, J. (2004). Statistics of extremes : Theory and applications.

Blanchet, J. et Davison, A. C. (2011). Spatial modeling of extreme snow depth. *The Annals of Applied Statistics*, pages 1699–1725.

Blanchet, J. et Lehning, M. (2010). Mapping snow depth return levels : smooth spatial modeling versus station interpolation. *Hydrology and Earth System Sciences*, 14(12):2527–2544.

Boldi, M.-O. et Davison, A. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69(2):217–229.

Boreux, J.-J., Parent, E. et Bernier, J. (2010). *Pratique du calcul bayésien*. Springer.

Brown, B. M. et Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probability*, 14(4):732–739.

Buishand, T., de Haan, L. et Zhou, C. (2008). On spatial extremes : with application to a rainfall problem. *The Annals of Applied Statistics*, pages 624–642.

Cai, J. J., Fougères, A.-L. et Mercadier, C. (2013). Environmental data : multivariate extreme value theory in practice. *Journal de la Société Française de Statistique*, 154(2):178–199.

Casella, G. et George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Castruccio, S., Huser, R. et Genton, M. G. (2015). High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*.

Chilès, J.-P. et Delfiner, P. (2009). Geostatistics : modeling spatial uncertainty.

Coles, S., Heffernan, J. et Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.

Coles, S. G. et Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–392.

Coles, S. G., Tawn, J. A. et Smith, R. L. (1994). A seasonal markov model for extremely low temperatures. *Environmetrics*, 5(3):221–239.

Cooley, D., Davis, R. A., Naveau, P. *et al.* (2012). Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. *The Annals of Applied Statistics*, 6(4):1406–1429.

Cooley, D., Nychka, D. et Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cooley, D. et Sain, S. R. (2010). Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(3):381–402.

Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, 4(5):613–617.

Davison, A. C. et Gholamrezaee, M. M. (2012). Geostatistics of extremes. 468(2138):581–608.

Davison, A. C., Huser, R. et Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):511–529.

Davison, A. C., Padoan, S., Ribatet, M. *et al.* (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

de Haan, L. (1984). A spectral representation for max-stable processes. *The annals of probability*, pages 1194–1204.

de Haan, L. et Ferreira, A. (2006). *Extreme Value Theory : An Introduction*. Springer Series in Operations Research and Financial Engineering. New York, NY : Springer.

de Haan, L. et Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337.

de Haan, L., Tank, A. K. et Neves, C. (2014). On tail trend detection : modeling relative risk. *Extremes*, pages 1–38.

Dombry, C., Éyi-Minko, F. et Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124.

Dombry, C. et Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*, 8(1):9–17.

Draisma, G., Dress, H., Ferreira, A. et De Haan, L. (2004). Bivariate tail estimation : dependence in asymptotic independence. *Bernoulli*, pages 251–280.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. et François, R. (2011). Rcpp : Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

El Adlouni, S., Favre, A.-C. et Bobée, B. (2006). Comparison of methodologies to assess the convergence of markov chain monte carlo methods. *Computational Statistics & Data Analysis*, 50(10):2685–2701.

Elie, L. et Lapeyre, B. (2001). Introduction aux méthodes de monte carlo. Cours.

Embrechts, P., Mikosch, T. et Klüppelberg, C. (1997). Modelling extremal events : for insurance and finance.

Engelke, S., Malinowski, A., Kabluchko, Z. et Schlather, M. (2015). Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 77(1):239–265.

Engelke, S., Malinowski, A., Oesting, M. et Schlather, M. (2012). Representations of max-stable processes based on single extreme events. *arXiv preprint arXiv :1209.2303*.

Fawcett, L. et Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Environmetrics*, 17(8): 795–809.

Fawcett, L. et Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18(2):173–188.

Ferreira, A. et de Haan, L. (2014). The generalized pareto process ; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Ferro, C. A. et Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 65(2):545–556.

Finkenstädt, B. et Rootzén, H. (2004). *Extreme values in finance, telecommunications and the environment*. Chapman & Hall/CRC, Boca Raton.

Fisher, R. A. et Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *In Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.

Fonseca, C., Ferreira, H., Pereira, L. et Martins, A. (2012). Stability and contagion measures for spatial extreme value analyses. *arXiv preprint arXiv :1206.1228*.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R. et Bernardara, P. (2010). Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences Discussions*, 14:p–951.

Gardes, L. et Girard, S. (2010). Conditional extremes from heavy-tailed distributions : An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204.

Gaume, J., Eckert, N., Chambon, G., Naaim, M. et Bel, L. (2013). Mapping extreme snowfalls in the french alps using max-stable processes. *Water Resources Research*, 49(2):1079–1098.

Gelfand, A. E. et Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Gelman, A., Carlin, J. B., Stern, H. S. et Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1(2):141–149.

Geweke, J. *et al.* (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Gilleland, E. et Katz, R. W. (2011). *`extRemes` : New software to analyze how extremes change over time*. R package.

Giné, E., Hahn, M. G. et Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probability theory and related fields*, 87(2):139–165.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T., Genton, M. G. et Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability*, 107:151.

Gottardi, F. (2009). *Estimation statistique et réanalyse des précipitations en montagne Utilisation d'ébauches par types de temps et assimilation de données d'enneigement Application aux grands massifs montagneux français*. Thèse de doctorat, Institut National Polytechnique de Grenoble-INPG.

Gottardi, F., Obled, C., Gailhard, J. et Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns : Application over french mountains. *Journal of Hydrology*, 432:154–167.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.

Guyon, X. (2007). Statistique spatiale. *In Conférence SADA (Statistique Appliquée pour le Développement en Afrique), Cotonou, Bénin*, page 96.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Heffernan, J., Stephenson, A. G. et Gilleland, E. (2016). *ismev : An Introduction to Statistical Modeling of Extreme Values*. R package version 1.41.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes*, 3(3):279–290.

Heffernan, J. E. et Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, pages 537–571.

Heffernan, J. E. et Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 66(3):497–546.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174.

Hipel, K. W. et McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*, volume 45. Elsevier.

Hsing, T. (1987). On the characterization of certain point processes. *Stochastic processes and their applications*, 26:297–316.

Huser, R. et Davison, A. (2014). Space–time modelling of extreme events. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 76(2):439–461.

Huser, R., Opitz, T. et Thibaud, E. (2016). Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *arXiv preprint arXiv :1610.04536*.

Jeon, S. et Smith, R. L. (2012). Dependence structure of spatial extremes using threshold approach. *arXiv preprint arXiv :1209.6344*.

Kabluchko, Z., Schlather, M. et de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, pages 2042–2065.

Klein Tank, A. et Können, G. (2003). Trends in indices of daily temperature and precipitation extremes in europe, 1946-99. *Journal of Climate*, 16(22):3665–3680.

Klein Tank, A., Wijngaard, J., Können, G., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C. *et al.* (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *International journal of climatology*, 22(12):1441–1453.

Klok, E. et Klein Tank, A. (2009). Updated and extended european dataset of daily climate observations. *International Journal of Climatology*, 29(8):1182–1191.

Lang, M., Naulet, R., Recking, A., Cœur, D. et Gigon, C. (2002). Etude de cas : l'analyse des pluies et crues extrêmes observées depuis 200 ans dans un bassin cévenol, l'ardèche. *La Houille Blanche*, (6-7):131–138.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Probability theory and related fields*, 28(4):289–303.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, 65(2):291–306.

Ledford, A. W. et Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80(1):221–39.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica : Journal of the Econometric Society*, pages 245–259.

Maraun, D., Osborn, T. J. et Rust, H. W. (2011). The influence of synoptic airflow on uk daily precipitation extremes. part i : Observed spatio-temporal relationships. *Climate Dynamics*, 36(1-2):261–275.

Matérn, B. (1986). Spatial variation, vol. 36 of. *Lecture Notes in Statistics*, 2.

McLeod, A. I. (2015). *Kendall : Kendall rank correlation and Mann-Kendall trend test*. R package version 2.2.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. et Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Oesting, M., Bel, L. et Lantuejoul, C. (2014). Sampling from max-stable processes conditional on a homogeneous functional via an mcmc algorithm. *In METMA VII and GRASPA14 Conference. Torino (IT), 10-12 September 2014*. IT.

Opitz, T. (2013). Extremal t processes : Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.

Padoan, S. A., Ribatet, M. et Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters*, 43(4):399–409.

Penot, D. (2014). *Cartographie des événements hydrologiques extrêmes et estimation SCHADEX en sites non jaugés*. Thèse de doctorat, Université de Grenoble.

Pettitt, A. (1979). A non-parametric approach to the change-point problem. *Applied statistics*, pages 126–135.

Pickands, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131.

Pickands, J. (1981). Multivariate extreme value distributions. *In Proceedings 43rd Session International Statistical Institute*, volume 2, pages 859–878.

Plummer, M., Best, N., Cowles, K. et Vines, K. (2006). Coda : Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.

Pohlert, T. (2016). *trend : Non-Parametric Trend Tests and Change-Point Detection*. R package version 0.2.0.

R Core Team (2013). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rakonczai, P. et Zempléni, A. (2012). Bivariate generalized pareto distribution in practice : models and estimation. *Environmetrics*, 23(3):219–227.

Reich, B. J. et Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The annals of applied statistics*, 6(4):1430.

Reich, B. J., Shaby, B. A. et Cooley, D. (2014). A hierarchical model for serially-dependent extremes : A study of heat waves in the western us. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):119–135.

Renard, B. (2006). *Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France*. Thèse de doctorat, INP Grenoble.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer.

Ribatet, M. (2015). *SpatialExtremes : Modelling Spatial Extremes*. R package version 2.0-2.

Ribatet, M., Ouarda, T. B., Sauquet, E. et Gresillon, J.-M. (2009). Modeling all exceedances above a threshold using an extremal dependence structure : Inferences on several flood characteristics. *Water Resources Research*, 45(3).

Robert, C. et Casella, G. (2009). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.

Roberts, G. O. et Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Rootzén, H. et Tajvidi, N. (2006). Multivariate generalized pareto distributions. *Bernoulli*, pages 917–930.

Roustant, O., Ginsbourger, D. et Deville, Y. (2012). DiceKriging, DiceOptim : Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.

Rust, H., Maraun, D. et Osborn, T. (2009). Modelling seasonality in extreme precipitation. *The European Physical Journal Special Topics*, 174(1):99–111.

Sabourin, A. (2015). Semi-parametric modeling of excesses above high multivariate thresholds with censored data. *Journal of Multivariate Analysis*, 136:126–146.

Sabourin, A. et Naveau, P. (2014). Bayesian dirichlet mixture model for multivariate extremes : a re-parametrization. *Computational Statistics & Data Analysis*, 71:542–567.

Sabourin, A. et Renard, B. (2015). Combining regional estimation and historical floods : A multivariate semiparametric peaks-over-threshold model with censored data. *Water Resources Research*, 51(12):9646–9664.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, P. J., Gross, S., Ober, U., Christoph Berreth, Burmeister, K., Manitz, J., Morena, O., Ribeiro, P., Singleton, R., Pfaff, B. et R Core Team (2016). *RandomFields : Simulation and Analysis of Random Fields*. R package version 3.1.12.

Schlather, M. et Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values : Properties and inference. *Biometrika*, 90(1):139–156.

Schliep, E. M., Cooley, D., Sain, S. R. et Hoeting, J. A. (2009). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.

Sebille, Q. (2016). *`hkevp` : A hierarchical model for Spatial Extremes*. R package version 1.1.4.

Shaby, B. A. et Reich, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics*, 23(8):638–648.

Sibuya, Y. (1960). Note on real matrices and linear dynamical systems with periodic coefficients. *Journal of Mathematical Analysis and Applications*, 1(3):363–372.

Smith, R. (1991). Regional estimation from spatially dependent data. *Preprint. http ://www.stat.unc.edu/postscript/rs/regest.pdf*.

Smith, R. L. (1986). Extreme value theory based on the r largest annual events. *Journal of Hydrology*, 86(1-2):27–43.

Smith, R. L. (1989). Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*.

Smith, R. L., Tawn, J. A. et Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Sneyers, R. (1984). Extremes in meteorology. *In Statistical Extremes and Applications*, pages 235–252. Springer.

Soriano, L. R., De Pablo, F. et Díez, E. G. (2001). Relationship between convective precipitation and cloud-to-ground lightning in the iberian peninsula. *Monthly Weather Review*, 129(12):2998–3003.

Southworth, H. et Heffernan, J. (2013). *`texmex` : Statistical modelling of extreme values*. R package version 2.1.

Stein, M. L. (1999). *Interpolation of Spatial Data : Some Theory for Kriging*. Springer Science & Business Media.

Stephenson, A. et Tawn, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227.

Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics*, 51(1):77–88.

Stephenson, A. G. et Ferro, C. (2015). *evd : Extreme Value Distributions*. R package version 2.3.2.

Stephenson, A. G., Shaby, B. A., Reich, B. J. et Sullivan, A. L. (2015). Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling*. *Journal of Applied Meteorology and Climatology*, 54(2):395–407.

Tapia, A., Smith, J. A. et Dixon, M. (1998). Estimation of convective rainfall from lightning observations. *Journal of Applied Meteorology*, 37(11):1497–1509.

Tawn, J. A. (1988). An extreme-value theory model for dependent observations. *Journal of Hydrology*, 101(1): 227–250.

Thibaud, E. et Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.

Van den Besselaar, E., Klein Tank, A. et Buishand, T. (2013). Trends in european precipitation extremes over 1951–2010. *International Journal of Climatology*, 33(12):2682–2689.

Wackernagel, H. (2013). *Multivariate geostatistics : an introduction with applications*. Springer Science & Business Media.

Wadsworth, J. et Tawn, J. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

Wadsworth, J. L. et Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-gaussian random functions. *Biometrika*, 101(1):1–15.

Wald, A. et Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, 14(4):378–388.

Wang, Y. et Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483.

Annexe : manuel du package `hkevp`

Package 'hkevp'

August 27, 2016

Type Package

Title Spatial Extreme Value Analysis with the Hierarchical Model of Reich and Shaby (2012)

Version 1.1.4

Date 2016-08-25

Author Quentin Sebille

Maintainer Quentin Sebille <quentin.sebille@gmail.com>

Description
Several procedures around a particular hierarchical model for extreme value: the HKEVP of Reich and Shaby (2012) <DOI:10.1214/12-AOAS591>. Simulation, estimation and spatial extrapolation of this model are available for extreme value data. A special case of this process is also handled: the Latent Variable Model of Davison et al. (2012) <DOI:10.1214/11-STS376>.

License GPL

LinkingTo Rcpp, RcppArmadillo

Depends Rcpp (>= 0.11.0)

RoxygenNote 5.0.1

NeedsCompilation yes

R topics documented:

`extrapol.gev` *Spatial extrapolation of GEV parameters with the HKEVP*

Description

Predictive distributions of the GEV parameters at a set of ungauged sites (targets), given the output from the MCMC procedures `hkevp.fit` or `latent.fit`. See details.

Usage

```
extrapol.gev(fit, targets, targets.covariates)
```

Arguments

`fit`	Output from the `hkevp.fit` procedure.
`targets`	A matrix of real values giving the spatial coordinates of the ungauged positions. Each row corresponds to an ungauged position.
`targets.covariates`	A matrix of real values giving the spatial covariates of the ungauged positions. Must match with the covariates used in `hkevp.fit` or `latent.fit`.

Details

Since the GEV parameters are modelled with latent Gaussian processes, spatial extrapolation of the marginal distributions at target positions $(s_1^*, ..., s_k^*)$ is performed with simple kriging. Estimation is done at each MCMC step to produce a sample of the predictive distribution.

Value

A named list of three elements: `loc`, `scale`, `shape`. Each one is a matrix with columns corresponding to targets positions.

Author(s)

Quentin Sebille

See Also

`extrapol.return.level`

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
loc <- sites[,1]*10
scale <- 3
shape <- 0
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, sites, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- hkevp.fit(ysim, sites, niter = 1000)

## Extrapolation:
targets <- matrix(1.5, 1, 2)
gev.targets <- extrapol.gev(fit, targets)

## True vs predicted:
predicted <- sapply(gev.targets, median)
sd.predict <- sapply(gev.targets, sd)
true <- c(targets[,1]*10, scale, shape)
# cbind(true, predicted, sd.predict)
```

`extrapol.return.level` *Spatial extrapolation of a return level.*

Description

Predictive distribution of a T-years return level at ungauged positions (targets), given the output from the MCMC procedures `hkevp.fit` or `latent.fit`.

Usage

```
extrapol.return.level(period, fit, targets, targets.covariates)
```

Arguments

`period`	An integer indicating the wished return period T.
`fit`	Output from the `hkevp.fit` procedure.
`targets`	A matrix of real values giving the spatial coordinates of the ungauged positions. Each row corresponds to an ungauged position.
`targets.covariates`	A matrix of real values giving the spatial covariates of the ungauged positions. Must match with the covariates used in `hkevp.fit` or `latent.fit`.

Details

Spatial extrapolation of the return level at target positions $(s_1^*, ..., s_k^*)$ is a two-step procedure:

- Estimation of the predictive distribution for GEV parameters at $(s_1^*, ..., s_k^*)$, by using `{extrapol.gev}`.
- Computation of the associated return level for each state of the predictive distribution.

Value

A matrix of predictive sample. Each column corresponds to a target position and each row to a predictive draw.

Author(s)

Quentin Sebille

See Also

`extrapol.gev`

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
knots <- sites
loc <- sites[,1]*10
scale <- 1
shape <- .2
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, knots, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- hkevp.fit(ysim, sites, niter = 1000)

## Extrapolation of the 100-years return level (may need more iterations and burn-in/nthin):
targets <- as.matrix(expand.grid(1.5:2.5,1.5:2.5))
pred.sample <- extrapol.return.level(100, fit, targets)
pred.mean <- apply(pred.sample, 2, mean)
pred.sd <- apply(pred.sample, 2, sd)
true <- return.level(100, targets[,1]*10, scale, shape)
# cbind(true, pred.mean, pred.sd)
```

hkevp *A hierarchical model for spatial extreme values*

Description

The HKEVP of *Reich and Shaby (2012)* is a hierarchical model which can be fitted on point-referenced block maxima across a region of space. Its acronym stands for Hierarchical Kernel Extreme Value Model.

This model fits both marginal GEV parameters and a conditional spatial dependence structure in a Bayesian framework. Estimation of all parameters is performed with a Metropolis-within-Gibbs algorithm that returns samples of posterior distributions, given prior distributions and data. See details.

Simulation and fitting procedure for this model are provided along with several other tools such as spatial extrapolation and conditional simulation, which is used for instance in *Shaby and Reich (2012)*.

A particular simpler case of the HKEVP, defined in Davison et al. (2012), is also available. This model, referred as the latent variable model, assumes conditional independence of the block maxima data (i.e. no dependence structure).

Spatial modelling of extreme values with general max-stable proceses is taken care of in other R libraries such as `RandomFields` and `SpatialExtremes`.

Details

The functions included in `hkevp` package are listed below:

1. `hkevp.fit` (resp. `latent.fit`): fits the HKEVP (resp. the latent variable model) to spatial block maxima data.
2. `hkevp.rand`: simulates data from the HKEVP.
3. `hkevp.expmeasure`: computes the exponent measure of the HKEVP. See details below.
4. `mcmc.fun`: applies a function to the main Markov chains obtained by `hkevp.fit`. Useful to compute the posterior means or quantiles for instance.
5. `mcmc.plot`: plots the resulting Markov chains, in order to visually assess convergence.
6. `extrapol.gev` (resp. `extrapol.return.level`) : computes the predictive distribution of the GEV parameters (resp. a return level) at ungauged positions.
7. `hkevp.predict`: predictive distribution of the spatial process at ungauged stes with the HKEVP given observations.

The HKEVP of *Reich and Shaby (2012)* is a hierarchical spatial max-stable model for extreme values. Max-stable models arise as the limiting distribution of renormalized maxima of stochastic processes and they generalize the Extreme Value Theory (EVT) to the infinite-dimensional case. For more information about EVT and max-stable processes, see for instance *Beirlant et al. (2004)*, *de Haan and Ferreira (2006)* and *Coles (2001)*. For an emphasis on statistical inference on spatial extremes, see for instance *Cooley et al. (2012)* and *Davison et al. (2012)*.

Let $Y(\cdot)$ be the process of block maxima recorded over a spatial region. Assume this process is max-stable, then all marginal distributions are necessarily GEV, i.e:

$$Y(s) \sim GEV\{\mu(s), \sigma(s), \xi(s)\}$$

with location, scale and shape parameters $\mu(s)$, $\sigma(s)$ and $\xi(s)$ respectively, indexed by position s in space. Without loss of generality, one may look at the *simple max-stable process* $Z(\cdot)$ with $GEV(1, 1, 1)$ margins, where spatial dependence is contained unconditionally of the marginals.

The HKEVP is thus defined by assuming that there exists $\alpha \in (0, 1]$ and a set of knots $\{v_1, ..., v_L\}$ with associated kernels $\{\omega_1(\cdot), ..., \omega_L(\cdot)\}$ such that:

$$Z(s) = U(s)\theta(s)$$

where $U(s)$ is a spatially-independent process with $GEV(1, \alpha, \alpha)$ margins and

$$\theta(s) = \left[\sum_{\ell=1}^{L} A_\ell \omega_\ell(s)^{1/\alpha}\right]^\alpha$$

is the *residual dependence process*, defined with a random variable $A_\ell \sim PS(\alpha)$, the positive stable distribution with characteristic exponent α. Note that the kernels must satisfy the condition $\sum_{\ell=1}^{L} \omega_\ell(s) = 1$, for all s.

Under those assumptions, *Reich and Shaby (2012)* showed that this model lead to an explicit formula for the distribution of a joint vector of maxima, which is not the case for max-stable processes since no general parametric representation exists (cf. *de Haan and Ferreira (2006)*). The joint distribution of the vector $\{Z(s_1), \ldots, Z(s_n)\}$ under the HKEVP assumptions can indeed be written:

$$P\{Z_1(s_1) < z_1, \ldots, Z_n(s_n) < z_n\} = \sum_{\ell=1}^{L}\left[\sum_{i=1}^{n}\left(\frac{\omega_\ell(s_i)}{z_i}\right)^{1/\alpha}\right]^\alpha .$$

The HKEVP can be seen as an approximation of the *Smith (1990)* model, but with an additional dependence parameter α which controls the strength of spatial dependence. Low value for α leads to a strong dependence structure while $\alpha = 1$ means independence.

In the article of *Reich and Shaby (2012)*, the kernels are chosen to be standardized Gaussian kernels, i.e:

$$\omega_\ell(s) = \frac{K(s_\ell|v_\ell, \tau)}{\sum_{j=1}^{L} K(s_j|v_j, \tau)} ,$$

where $K(\cdot|v, \tau)$ is the Gaussian kernel centered at v_ℓ and τ is a bandwidth parameter.

Conditionally on the marginal and the dependence parameters, we obtain independent responses which allow the computation of the likelihood. Since the model is structured via multiple layers, Bayesian inference is preferred. Details of the MCMC algorithm can be found in *Reich and Shaby (2012)*.

Marginal parameters $\mu(s)$, $\sigma(s)$ and $\xi(s)$ are modelled through latent Gaussian processes.

A simpler version of the HKEVP, namely the latent variable model of Davison et al. (2012), is also available in this package. This model assumes conditional independence and can therefore be seen as a special case of the HKEVP, with the condition $\alpha = 1$.

Note

I would like to thank Brian Reich and Benjamin Shaby for their help regarding the implementation of the inference procedure of the HKEVP, and Mathieu Ribatet for introducing me to the Bayesian world. I also acknowledge Electricite de France (EDF) for the financial support and the helpful discussions around the HKEVP with Anne Dutfoy, Marie Gallois, Thi Thu Huong Hoang and Sylvie Parey. Finally, I thank my two PhD supervisors Anne-Laure Fougeres and Cecile Mercadier for their reviews of this package.

Author(s)

Quentin Sebille

References

Beirlant, J., Goegebeur, Y., Segers, J. J. J., & Teugels, J. (2004). Statistics of Extremes: Theory and Applications. <DOI:10.1002/0470012382>

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Science & Business Media. <10.1007/978-1-4471-3675-0>

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., & Sun, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. Revstat, 10(1), 135-165.

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. Statistical Science, 27(2), 161-186. <DOI:10.1214/11-STS376>

de Haan, L., & Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer Science & Business Media. <DOI:10.1007/0-387-34471-3>

Reich, B. J., & Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics, 6(4), 1430. <DOI:10.1214/12-AOAS591>

Shaby, B. A., & Reich, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. Environmetrics, 23(8), 638-648. <DOI:10.1002/env.2178>

Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

`hkevp.expmeasure` *Exponent measure of the HKEVP*

Description

Exponent measure $V(z_1, ..., z_n)$ of the HKEVP of *Reich and Shaby (2012)*, with given model parameters or output from `hkevp.fit` or `latent.fit`.

Usage

```
hkevp.expmeasure(z, sites, knots, alpha, tau, fit)
```

Arguments

z	The vector $(z_1, ..., z_n)$ where the exponent measure is computed. Can be of length one and thus corresponds then to $(z, ..., z)$.
sites	The coordinates of the sites where the data are observed. Each row correspond to a site position.
knots	The coordinates of the knots in the HKEVP. By default, the positions of the knots coincide with the positions of the sites.
alpha	The dependence parameter α of the HKEVP: a single value in (0,1].
tau	The bandwidth parameter τ of the kernel functions in the HKEVP: a positive value.
fit	Output from the `hkevp.fit` procedure.

Details

The exponent measure describes the spatial dependence structure of a max-stable process, independently from the values of the marginal parameters. If $Z(\cdot)$ is a simple max-stable process, i.e. with unit GEV(1,1,1) margins, recorded at the set of sites $(s_1, \ldots, s_n)$, its joint cumulative probability density function is given by:

$$P\{Z(s_1) \leq z_1, \ldots, Z(s_n) \leq z_n\} = \exp(-V(z_1, \ldots, z_n)) \, ,$$

where V is the so-called exponent measure. For the HKEVP, the exponent measure is explicit for any number n of sites:

$$V(z_1, \ldots, z_n) = \sum_{\ell=1}^{L} \left[\sum_{i=1}^{n} \left(\frac{\omega_\ell(s_i)}{z_i} \right)^{1/\alpha} \right]^{\alpha} \, .$$

If argument `fit` is provided, the predictive distribution of

$$V(z_1, \ldots, z_n)$$

is computed. If not, the function uses arguments `sites`, `knots`, `alpha`, and `tau`.

Value

Either a vector if argument `fit` is provided, or a single value.

Author(s)

Quentin Sebille

References

Reich, B. J., & Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics, 6(4), 1430. <DOI:10.1214/12-AOAS591>

Examples

```
sites <- as.matrix(expand.grid(1:3,1:3))
loc <- sites[,1]*10
scale <- 3
shape <- 0
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, sites, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- hkevp.fit(ysim, sites, niter = 1000)

predict.em <- hkevp.expmeasure(1, fit = fit)
true.em <- hkevp.expmeasure(1, sites, sites, alpha, tau)
# plot(predict.em, ylim = range(predict.em, true.em), type = "l")
# abline(h = true.em, col = 2, lwd = 2)
```

`hkevp.fit` *Fitting procedure of the HKEVP with MCMC algorithm*

Description

Metropolis-within-Gibbs algorithm that returns samples from posterior distribution of all the parameters of the HKEVP.

Most of the input parameters have default values, so that the procedure can be easily handled. However, convergence of the Markov chains should be assessed by using `mcmc.plot` for instance. The experimented user can set initial states, prior hyperparameters along with the magnitude of the MCMC jumps.

Usage

```
hkevp.fit(y, sites, knots, niter, nburn, nthin, quiet, trace, fit.margins,
  gev.vary, spatial.covariates, log.scale, correlation, mcmc.init, mcmc.prior,
  mcmc.jumps)
```

Arguments

`y`	A matrix of observed block maxima. Each column corresponds to a site position.
`sites`	The coordinates of the sites where the data are observed. Each row corresponds to a site position.
`knots`	The coordinates of the knots in the HKEVP. By default, the positions of the knots coincide with the positions of the sites.
`niter`	The number of MCMC iterations.
`nburn`	The number of first MCMC iterations that are discarded. Zero by default.

`nthin`	The size of the MCMC thinning. One by default (i.e. no thinning).
`quiet`	A logical indicating if the progression of the routine should be displayed. TRUE by default.
`trace`	If `quiet` is FALSE, the log-likelihood of the model is displayed each block of `trace` MCMC steps to observe fitting progression.
`fit.margins`	A logical that indicates if the GEV parameters should be fitted along with the dependence structure. TRUE by default.
`gev.vary`	A logical vector of size three indicating if the GEV parameters (respectively the location, the scale and the shape) are spatially-varying. If not (by default for the shape), the parameter is the same at each position.
`spatial.covariates`	A numerical matrix of spatial covariates. Each row corresponds to a site position. See details.
`log.scale`	A logical value indicating if the GEV scale parameter σ is modelled by its log. FALSE by default. See details.
`correlation`	A character string indicating the form of the correlation function associated to the latent Gaussian processes that describes the marginal parameters. Must be one of `"expo"`, `"gauss"`, `"mat32"` (By default) and `"mat52"`, respectively corresponding to the exponential, Gaussian, Matern-3/2 and Matern-5/2 correlation functions.
`mcmc.init`	A named list indicating the initial states of the chains. See details.
`mcmc.prior`	A named list indicating the hyperparameters of the prior distributions. See details.
`mcmc.jumps`	A named list indicating the amplitude of the jumps to propose the MCMC candidates. See details.

Details

Details of the MCMC procedure are presented in *Reich and Shaby (2012)*. This function follows the indications and the choices of the authors, with the exception of several small changes:

- The scale parameter σ can be modelled like the two other marginal parameters as in *Davison et al. (2012)* or by its logarithm as in *Reich and Shaby (2012)*. For this, use the argument `log.scale`, set to FALSE by default.
- The Inverse-Gamma prior distributions defined for the bandwith parameter τ and for the ranges λ of the latent processes are replaced by a Beta distribution over the interval $[0, 2D_{max}]$, where D_{max} stands for the maximum distance between two sites.

The procedure can be used normally with `fit.margins = TRUE` (default) or by assuming that the observed process had GEV(1,1,1) margins already and thus ignoring the marginal estimation.

If the margins are estimated and the parameters are assumed spatially-varying, the user can provide spatial covariates to fit the mean of the latent Gaussian processes. Recall for instance for the GEV location parameter that:

$$\mu(s) = \beta_{0,\mu} + \beta_{1,\mu} c_1(s) + ... + \beta_{p,\mu} c_p(s) \ .$$

The given matrix `spatial.covariates` that represents the $c_i(s)$ elements should have the first column filled with ones to account for the intercept β_0.

The arguments `mcmc.init`, `mcmc.prior` and `mcmc.jumps` are named list that have default values. The user can make point changes in these arguments, by setting `mcmc.init = list(alpha = .5)` for instance, but must respect the constraints of each element:

- `mcmc.init`. All elements are of length one. The possibilities are:
  - `loc`, `scale` and `shape` (GEV parameters).
  - `range` and `sill` of the correlation functions.
  - `alpha`, `tau`, `A` and `B`, the dependence parameters and conditional variables of the HKEVP.
- mcmc.prior. The possible elements are:
  - `constant.gev`: a 2×3 matrix of normal parameters for spatially-constant μ, σ and ξ. The first row are the means, the second are the standard deviations.
  - `beta.sd`: the normal sd prior of all β parameters (a single value).
  - `range`, `alpha` and `tau`: the two Beta parameters.
  - `sill`: the two Inverse-Gamma parameters.
- mcmc.jumps. The possible elements are:
  - `gev` and `range`: a vector of length 3 (for each GEV parameter).
  - `tau`, `alpha`, `A`, `B`: single values for each.

Value

A named list with following elements:

- `GEV`: the Markov chains associated to the GEV parameters. The dimensions of the array correspond respectively to the sites positions, the three GEV parameters and the states of the Markov chains.
- `alpha`: the Markov chain associated to the dependence parameter α.
- `tau`: the Markov chain associated to the dependence parameter τ.
- `A`: the Markov chains associated to the positive stable random effect per site and per block. The dimensions correspond respectively to the indices of blocks, the knots positions and the states of the Markov chains.
- `llik`: the log-likelihood of the model for each step of the algorithm.
- `time`: time (in sec) spent for the fit.
- `spatial`: a named list with four elements linked to the GEV spatially-varying parameters:
  - `vary`: the argument `gev.vary`.
  - `beta`: the β parameters for each GEV parameter. The dimensions correspond respectively to the steps of the Markov chains, the p spatial covariates and the GEV parameters
  - `sills`: the Markov chains associated to the sills in the correlation functions of the latent Gaussian processes.
  - `ranges`: the Markov chains associated to the ranges in the correlation functions of the latent Gaussian processes.
- `data`: the data fitted.

- `sites`: the sites where the data are observed.
- `knots`: the set of knots.
- `spatial.covariates`: the spatial covariates.
- `correlation`: the type of correlation function for the marginal latent processes.
- `nstep`: the number of steps at the end of the routine after burn-in and thinning.
- `log.scale`: a boolean indicating if the scale parameter has been modelled via its logarithm.
- `fit.type`: either "hkevp" or "dep-only" character string to specify the type of fit.

If `fit.margins` is false, only the dependence-related elements are returned.

Author(s)

Quentin Sebille

References

Reich, B. J., & Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics, 6(4), 1430. <DOI:10.1214/12-AOAS591>

Stephenson, A. G. (2009) High-dimensional parametric modelling of multivariate extreme events. Aust. N. Z. J Stat, 51, 77-88. <DOI:10.1111/j.1467-842X.2008.00528.x>

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. Statistical Science, 27(2), 161-186. <DOI:10.1214/11-STS376>

See Also

latent.fit

Examples

```
# Simulation of HKEVP:
set.seed(1)
sites <- as.matrix(expand.grid(1:3,1:3))
loc <- sites[,1]*10
scale <- 3
shape <- 0
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, sites, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- latent.fit(ysim, sites, niter = 1000)
```

`hkevp.predict` *Predictive distribution of the max-stable process at target positions.*

Description

Computes the predictive distribution of $Y(\cdot)$ at a set of ungauged positions $(s_1^*, ..., s_k^*)$, given data at gauged positions $(s_1, ..., s_n)$, by using the output of *latent.fit* or `hkevp.fit`.

Two types of prediction are available for the HKEVP, as described in *Shaby and Reich (2012)*. See details.

Usage

```
hkevp.predict(fit, targets, targets.covariates, predict.type = "kriging")
```

Arguments

`fit`	Output from the `hkevp.fit` procedure.
`targets`	A matrix of real values giving the spatial coordinates of the ungauged positions. Each row corresponds to an ungauged position.
`targets.covariates`	A matrix of real values giving the spatial covariates of the ungauged positions. Must match with the covariates used in `hkevp.fit` or `latent.fit`.
`predict.type`	Character string specifying the type of prediction. Must be one of "`kriging`" (default) or "`climat`". See details.

Details

The spatial prediction of $Y_t(s^*)$ for a target site s^* and a realisation t of the process is described in *Shaby and Reich (2012)*. This method involves a three-step procedure:

1. Computation of the residual dependence process $\theta(\cdot)$ at the target positions.
2. Computation of the conditional GEV parameters (μ^*, σ^*, ξ^*) at the target sites. See the definition of the HKEVP in *Reich and Shaby (2012)*.
3. Generation of $Y_t(s^*)$ from an independent GEV distribution with parameters (μ^*, σ^*, ξ^*).

As sketched in *Shaby and Reich (2012)*, two types of prediction are possible: the kriging-type and the climatological-type. These two types differ when the residual dependence process θ is computed (first step of the prediction):

- The kriging-type takes the actual value of A in the MCMC algorithm to compute the residual dependence process. The prediction will be the distribution of the maximum recorded at the specified targets.
- The climatological-type generates A by sampling from the positive stable distribution with characteristic exponent α, where α is the actual value of the MCMC step. The prediction in climatological-type will be the distribution of what could happen in the conditions of the HKEVP dependence structure.

Posterior distribution for each realisation t of the process and each target position s^* is represented with a sample where each element corresponds to a step of the MCMC procedure.

Value

A three-dimensional array where:

- Each row corresponds to a different realisation of the process (a block).
- Each column corresponds to a target position.
- Each slice corresponds to a MCMC step.

Author(s)

Quentin Sebille

References

Reich, B. J., & Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics, 6(4), 1430. <DOI:10.1214/12-AOAS591>

Shaby, B. A., & Reich, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. Environmetrics, 23(8), 638-648. <DOI:10.1002/env.2178>

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
targets <- as.matrix(expand.grid(1.5:2.5,1.5:2.5))
all.pos <- rbind(sites, targets)
knots <- sites
loc <- all.pos[,1]*10
scale <- 3
shape <- 0
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, all.pos, knots, loc, scale, shape, alpha, tau)
yobs <- ysim[,1:9]

# HKEVP fit (omitting first site, used as target):
fit <- hkevp.fit(yobs, sites, niter = 1000)

# Extrapolation:
ypred <- hkevp.predict(fit, targets, predict.type = "kriging")

# Plot of the density and the true value for 4 first realizations:
# par(mfrow = c(2, 2))
# plot(density(ypred[1,1,]), main = "Target 1 / Year 1")
# abline(v = ysim[1,10], col = 2, lwd = 2)
# plot(density(ypred[2,1,]), main = "Target 1 / Year 2")
# abline(v = ysim[2,10], col = 2, lwd = 2)
# plot(density(ypred[1,2,]), main = "Target 2 / Year 1")
# abline(v = ysim[1,11], col = 2, lwd = 2)
# plot(density(ypred[2,2,]), main = "Target 2 / Year 2")
# abline(v = ysim[2,11], col = 2, lwd = 2)
```

`hkevp.rand` *Simulation of the HKEVP*

Description

Simulation procedure of the HKEVP with given sites and knots positions and marginal and spatial dependence parameters.

Usage

```
hkevp.rand(nrep, sites, knots, loc, scale, shape, alpha, tau)
```

Arguments

`nrep`	A positive integer. Number of realisations of the block mashapema process.
`sites`	The coordinates of the sites where the data are observed. Each row corresponds to a site position.
`knots`	The coordinates of the knots in the HKEVP. By default, the positions of the knots coincide with the positions of the sites.
`loc`	A numerical value or a vector of real values for the GEV location parameter. If a vector, its length must coincide with the number of sites. The value by default is 1.
`scale`	A numerical value or a vector of real values for the GEV scale parameter. If a vector, its length must coincide with the number of sites. The value by default is 1.
`shape`	A numerical value or a vector of real values for the GEV shape parameter. If a vector, its length must coincide with the number of sites. The value by default is 1.
`alpha`	The dependence parameter α of the HKEVP: a single value in (0,1].
`tau`	The bandwidth parameter τ of the kernel functions in the HKEVP: a positive value.

Details

Simulating one realisation of the block mashapema process $Y(\cdot)$ from the HKEVP involves three steps:

1. The *nugget process* $U(\cdot)$ is generated independently at each position, by simulating a random variable with $GEV(1, \alpha, \alpha)$ distribution.

2. The *residual dependence process* $\theta(\cdot)$ is computed by using the kernel functions centered at the set of knots, the bandwidth parameter τ and the simulations of the positive stable $PS(\alpha)$ random effect A.
3. The process $Z = U\theta$ is computed and its margins are transformed to the general GEV distribution with $\mu(s), \sigma(s)$ and $\xi(s)$ parameters.

Value

A numerical matrix of real values. Each column corresponds to a position and each row to a realisation of the process.

Author(s)

Quentin Sebille

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
loc <- sites[,1]*10
scale <- 3
shape <- 0
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, sites, loc, scale, shape, alpha, tau)
```

`latent.fit` *Fitting procedure of the latent variable model*

Description

Metropolis-within-Gibbs algorithm that returns posterior distribution (as Markov chains) for the marginal GEV parameters of an observed spatial process. This function is close to `hkevp.fit` but with less parameters since conditional independence is assumed and only the margins are estimated. In `SpatialExtremes` library, a similar function can be found under the name `latent`.

Usage

```
latent.fit(y, sites, niter, nburn, nthin, quiet, trace, gev.vary,
  spatial.covariates, log.scale, correlation, mcmc.init, mcmc.prior, mcmc.jumps)
```

Arguments

`y`	A matrix of observed block maxima. Each column corresponds to a site position.
`sites`	The coordinates of the sites where the data are observed. Each row corresponds to a site position.
`niter`	The number of MCMC iterations.
`nburn`	The number of first MCMC iterations that are discarded. Zero by default.
`nthin`	The size of the MCMC thinning. One by default (i.e. no thinning).
`quiet`	A logical indicating if the progression of the routine should be displayed. TRUE by default.
`trace`	If `quiet` is FALSE, the log-likelihood of the model is displayed each block of `trace` MCMC steps to observe fitting progression.
`gev.vary`	A logical vector of size three indicating if the GEV parameters (respectively the location, the scale and the shape) are spatially-varying. If not (by default for the shape), the parameter is the same at each position.
`spatial.covariates`	A numerical matrix of spatial covariates. Each row corresponds to a site position. See details.
`log.scale`	A logical value indicating if the GEV scale parameter σ is modelled by its log. FALSE by default. See details.
`correlation`	A character string indicating the form of the correlation function associated to the latent Gaussian processes that describes the marginal parameters. Must be one of `"expo"`, `"gauss"`, `"mat32"` (By default) and `"mat52"`, respectively corresponding to the exponential, Gaussian, Matern-3/2 and Matern-5/2 correlation functions.
`mcmc.init`	A named list indicating the initial states of the chains. See details.
`mcmc.prior`	A named list indicating the hyperparameters of the prior distributions. See details.
`mcmc.jumps`	A named list indicating the amplitude of the jumps to propose the MCMC candidates. See details.

Details

Details of the MCMC procedure are presented in *Davison et al. (2012)*. This function follows the indications and the choices of the authors, with the exception of several small changes:

- The scale parameter σ can be modelled like the two other marginal parameters as in *Davison et al. (2012)* or by its logarithm as in *Reich and Shaby (2012)*. For this, use the argument `log.scale`, set to FALSE by default.
- The Inverse-Gamma prior distributions defined for the bandwith parameter τ and for the ranges λ of the latent processes are replaced by a Beta distribution over the interval $[0, 2D_{max}]$, where D_{max} stands for the maximum distance between two sites.

If the the parameters are assumed spatially-varying, the user can provide spatial covariates to fit the mean of the latent Gaussian processes. Recall for instance for the GEV location parameter that:

$$\mu(s) = \beta_{0,\mu} + \beta_{1,\mu} c_1(s) + ... + \beta_{p,\mu} c_p(s) \ .$$

The given matrix spatial.covariates that represents the $c_i(s)$ elements should have the first column filled with ones to account for the intercept β_0.

The arguments mcmc.init, mcmc.prior and mcmc.jumps are named list that have default values. The user can make point changes in these arguments, by setting mcmc.init = list(loc = .5) for instance, but must respect the constraints of each element:

- mcmc.init: all elements are of length one. The possible elements are:
  - loc, scale and shape (GEV parameters).
  - range and sill of the correlation functions.
- mcmc.prior: the possible elements are:
  - constant.gev: a 2×3 matrix of normal parameters for spatially-constant μ, σ and ξ. The first row are the means, the second are the standard deviations.
  - beta.sd: the normal sd prior of all β parameters (a single value).
  - range: the two Beta parameters.
  - sill: the two Inverse-Gamma parameters.
- mcmc.jumps: the possible elements are gev and range, vectors of length 3 (for each GEV parameter).

Value

A named list with following elements (less elements if fit.margins is FALSE):

- GEV: the Markov chains associated to the GEV parameters. The dimensions of the array correspond respectively to the sites positions, the three GEV parameters and the states of the Markov chains.
- llik: the log-likelihood of the model for each step of the algorithm.
- time: time (in sec) spent for the fit.
- spatial: a named list with four elements linked to the GEV spatially-varying parameters:
  - vary: the argument gev.vary.
  - beta: the β parameters for each GEV parameter. The dimensions correspond respectively to the steps of the Markov chains, the p spatial covariates and the GEV parameters
  - sills: the Markov chains associated to the sills in the correlation functions of the latent Gaussian processes.
  - ranges: the Markov chains associated to the ranges in the correlation functions of the latent Gaussian processes.
- data: the data fitted.
- sites: the sites where the data are observed.
- spatial.covariates: the spatial covariates.
- correlation: the type of correlation function for the marginal latent processes.
- nstep: the number of steps at the end of the routine after burn-in and thinning.
- log.scale: a boolean indicating if the scale parameter has been modelled via its logarithm.
- fit.type: "latent" character string to specify the type of fit.

Author(s)

Quentin Sebille

References

Reich, B. J., & Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. The annals of applied statistics, 6(4), 1430. <DOI:10.1214/12-AOAS591>

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. Statistical Science, 27(2), 161-186. <DOI:10.1214/11-STS376>

See Also

hkevp.fit

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:4,1:4))
loc <- sites[,1]*10
scale <- 3
shape <- .2
alpha <- 1
tau <- 2
ysim <- hkevp.rand(15, sites, sites, loc, scale, shape, alpha, tau)

# Latent Variable Model fit:
set.seed(1)
fit <- latent.fit(ysim, sites, niter = 10000, nburn = 5000, nthin = 5)



mcmc.plot(fit, TRUE)
par(mfrow = c(2,2))
apply(fit$GEV[,1,], 1, acf)
```

---

mcmc.fun *Point estimates of HKEVP fit*

Description

Application of a function to the main Markov chains resulting from the procedure `hkevp.fit`. May be used to obtain point estimates on posterior distribution (e.g. the mean, the median). See details.

Usage

```
mcmc.fun(fit, FUN, ...)
```

Arguments

fit	A named list. Output from the hkevp.fit procedure.
FUN	The function applied to the Markov chains in fit. The median by default. The output from FUN must be a single value.
...	Optional arguments of the function to be applied on the Markov chains (e.g. na.rm = FALSE).

Details

A function is applied to the main Markov chains resulting from the MCMC procedures hkevp.fit or latent.fit. These chains correspond to the three GEV parameters, the dependence parameter α and the bandwidth τ.

The value returned by FUN must be a single value.

Value

If fitted model is the HKEVP, a named list with three elements:

- GEV: A numerical matrix. Result of the function FUN for each GEV parameter (columns) and each site position (rows).
- alpha: A numerical value. Result of the function FUN on the Markov chain associated to the dependence parameter α.
- tau: A numerical value. Result of the function FUN on the Markov chain associated to the bandwidth parameter τ.

If fitted model is the latent variable model, the functions returns the GEV matrix only.

Author(s)

Quentin Sebille

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
knots <- sites
loc <- sites[,1]*10
scale <- 3
shape <- .2
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, knots, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- hkevp.fit(ysim, sites, niter = 1000)
```

```
# Posterior median and standard deviation:
# mcmc.fun(fit, median)
# mcmc.fun(fit, sd)
```

`mcmc.plot` *Markov chains plotting*

Description

Plots of the resulting Markov chains obtained by the MCMC procedures `hkevp.fit` or `latent.fit`. May be used to assess graphically convergence of the chains.

Usage

```
mcmc.plot(fit, plot.spatial, mfrow)
```

Arguments

`fit`	Output from the `hkevp.fit` procedure.
`plot.spatial`	Logical indicating if the Markov chains of the sills and ranges hyperparameters should be plotted. FALSE by default.
`mfrow`	Optional vector of two numerical values indicating the parameter of the window plotting called by the `plot(...)` function.

Author(s)

Quentin Sebille

Examples

```
# Simulation of HKEVP:
sites <- as.matrix(expand.grid(1:3,1:3))
knots <- sites
loc <- sites[,1]*10
scale <- 3
shape <- .2
alpha <- .4
tau <- 1
ysim <- hkevp.rand(10, sites, knots, loc, scale, shape, alpha, tau)

# HKEVP fit:
fit <- hkevp.fit(ysim, sites, niter = 1000)

# Markov chains plot:
# mcmc.plot(fit)
```

mcmc_hkevp — *C routine for fitting either the HKEVP or the latent variable model with MCMC.*

Description

This set of functions are called by `hkevp.fit` or `latent.fit` and should not be used directly by the user (high risk of segmentation fault)!

Author(s)

Quentin Sebille

See Also

`hkevp.fit`

return.level — *The associated return level*

Description

Computation of the associated return level with given period and GEV parameters.

Usage

```
return.level(period, loc, scale, shape)
```

Arguments

`period`	An integer indicating the wished return period T.
`loc`	A numerical value or vector for the GEV location parameter. Must be of length one or same length as `scale` and/or `shape`.
`scale`	A numerical value or vector for the GEV scale parameter. Must be of length one or same length as `loc` and/or `shape`.
`shape`	A numerical value or vector for the GEV shape parameter. Must be of length one or same length as `loc` and/or `scale`.

Details

The T-year return level is a common value of risk in Extreme Value Theory. It represents the value that is expected to be exceeded once over T years by the annual maxima. Given the parameters μ, σ and ξ of the GEV distribution associated to the yearly maxima, we can compute the associated T-return level y_T by:

$$y_T := \mu + \frac{\sigma}{\xi}\left[\log\left(\frac{T}{T-1}\right)^{-\xi} - 1\right] .$$

Value

A numerical value or a numerical vector, depending on the input arguments `loc`, `scale`, `shape`

Author(s)

Quentin Sebille

Examples

```
return.level(period = 100, loc = 1, scale = 1, shape = 1)
return.level(period = 200, loc = 1:10, scale = 1, shape = 0)
```

Index

Modélisation spatiale de valeurs extrêmes
Application à l'étude de précipitations en France

Résumé: Les précipitations extrêmes en France sont responsables de phénomènes d'inondations entraînant la perte de vies humaines et des millions d'euros en dégâts matériels. Une manière de mesurer le risque associé à ces événements météorologiques rares est de faire appel à la théorie statistique des valeurs extrêmes, qui propose plusieurs approches permettant d'évaluer des scénarios catastrophes. Cette thèse s'intéresse en particulier à trois mesures de risque faisant intervenir à la fois des lois de probabilité jointes et des méthodes de prédiction spatiale liées à la géostatistique.
Dans un premier temps, plusieurs modèles spatiaux de valeurs extrêmes construits sur des données de maxima annuels sont évalués dans une étude comparative ayant conduit à l'écriture d'un article. La comparaison des méthodes est menée en se servant de simulations construites à partir de données réelles de maxima annuels de précipitations en France. Les critères choisis sont le niveau de retour centennal et le coefficient extrémal. Un des modèles - le processus max-stable et hiérarchique de Reich et Shaby (2012)- fait l'objet d'une étude particulière. De plus, il a ensuite été implémenté dans un package R permettant à la fois sa simulation et son inférence.
Dans un second temps, les données journalières dépassant un seuil élevé sont modélisées dans un cadre spatial dans le but d'estimer une probabilité d'échec conditionnelle. Plusieurs estimateurs de cette mesure sont proposés grâce à des modèles paramétriques (processus Pareto) d'une part et à des approches non paramétriques d'autre part. Les méthodes construites tiennent compte de la dépendance temporelle observable dans les valeurs journalières.
Tout au long de la thèse, les méthodes développées sont appliquées sur des données réelles (cumuls journaliers de précipitations en France).

Mots clés: Valeurs extrêmes; Processus spatiaux; Mesures de risque; Précipitations extrêmes.

Spatial modeling of extreme values. Application to precipitation in France

Abstract: Extreme precipitation in France are responsible for flooding events that cause people's deaths and billions of euros in material damage. Measuring the risk associated to these rare meteorological events is possible thanks to the extreme value theory which allows the estimation of such catastrophic scenarios. This thesis focus on three risk measures involving joint probabilities and spatial prediction methods related to geostatistics.
In a first time, several spatial models for extreme values built on annual maxima are evaluated in a comparative study in the form of an article. This comparison is performed using simulated data from real annual maxima of precipitation in France. The two criteria used are linked to risk measures: the hundred years return level and the extremal coefficient. One particular model is presented in details: the one of Reich and Shaby (2012). This model is implemented under a R package entirely dedicated to its estimation and simulation procedures.
In a second time, exceedances of spatial daily data are modelled in order to estimate a conditional failure probability. Several estimators of this measure are proposed, based on either parametric methods (Pareto processes) or non parametric approaches. The temporal dependence is also considered with care when estimating this probability.
Along this thesis, the methods are applied on daily data of precipitation in France.

Keywords: Extreme values; Spatial processes; Risk measures; Extreme precipitation.

Image en couverture : Simulation d'un processus spatial selon le modèle de Reich et Shaby (2012).

UNIVERSITÉ DE LYON