

Functional Peaks-over-threshold Analysis for Complex Extreme Events

THÈSE N° 8927 (2018)

PRÉSENTÉE LE 7 DÉCEMBRE 2018
À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE DE STATISTIQUE
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Raphaël Gérard Théodore Michel Marie
DE DELOYË ET FOURCADE DE FONDEVILLE

acceptée sur proposition du jury:

Prof. V. Panaretos, président du jury
Prof. A. C. Davison, directeur de thèse
Prof. H. Rootzén, rapporteur
Prof. C. Dombry, rapporteur
Prof. S. Morgenthaler, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

*“Usage de la fenêtre: inviter la beauté à entrer et
laisser l’inspiration sortir.”*

— Sylvain Tesson

Acknowledgements

My gratitude goes first to my supervisor, Anthony Davison, without whom none of this work would have been possible. He gave me the freedom to conduct my own research while supporting me with his positivism and entire trust, two qualities which contributed to make my time at EPFL such a wonderful and unique experience.

Also I would like to thank Clément Dombry, Stephan Morgenthaler, Victor Panaretos, and Holger Rootzén for having accepted to be members of my thesis committee and for the time they dedicated to reading it. Exchanging with them about the research that has occupied my mind during the past four years was both really exciting and pleasant. Warm thanks also go to my co-authors, Sebastian Engelke and Marco Oesting; collaborating with them has been great and considerably improved my understanding of extreme value theory.

Thank you to all my colleagues, Yousrah, Thomas, Léo, Hélène and Sophie, with whom I shared evenly joys and doubts both professional and existential. Coffee breaks and after work meetings with Marie-Hélène, Rémy, and Linda have been precious moments that were necessary to counter the loneliness that a Ph.D. candidate can face; I have been lucky to meet you. Also, I would like to thank Nadia for her kindness and availability: it is good to know that the door of your office is always open for a conversation.

Running at the sport center paced my weeks over the past four years, giving me the opportunity to lower the pressure and forget about work. There I met numerous nice and caring people who became good friends. Silvia, Hugo, Maxime, Tara, ... just to name a few. Also, I would like to thank Audrey-Anne and Julien for the time spent together: moving to Lausanne has really been the best decision.

This thesis would not have been possible without the unswerving support of my family, who have always been present for me at every step of my life.

Finally, moving in Lausanne gave me the opportunity to meet Laure, the most impor-

Acknowledgements

tant event of the past four years. She had to deal with me in every possible state of mind, from great happiness to deep discouragement. In these moments, she has been cheerful and attentive to me, and I know she will always be. Thank you so much.

Lausanne, November 2018

Raphaël de Fondeville

Abstract

Most current risk assessment for complex extreme events relies on catalogues of similar events, either historical or generated artificially. In the latter, no existing methods produce completely new events with mathematically justified extrapolation above observed level of severity. This thesis contributes to the development of stochastic generators of events based on extreme value theory, with a special focus on natural hazards.

The sources of historical meteorological records are multiple but climate model output is attractive for its spatial completeness and homogeneity. From a statistical perspective, these are massive gridded data sets, which can be exploited for accurate estimation of extreme events. The first contribution of this thesis describes methods of statistical inference for extremal processes that are computationally tractable for large data sets. We also relate the extremal behaviour of aggregated data to point observations, a result that we use to downscale gridded data to local tail distributions. These contributions are illustrated by applications to rainfall and heatwaves.

Building stochastic generators of extreme events requires the extension of classical peaks-over-threshold analysis to continuous stochastic processes. We develop a framework in which characterization of complex extremes can be motivated by field-specific expertise. The contribution includes the description of the theoretical limiting distribution of functional exceedances, called the generalized r -Pareto process, the functional equivalent of the generalized Pareto distribution, for which we describe statistical inference procedures, simulation algorithms and goodness-of-fit diagnostics. We apply these results to build a stochastic weather generator of extreme wind storms over Europe.

Keywords: Censored likelihood; Downscaling; Extreme value theory; Generalized r -Pareto process; Gradient score; High-dimensional inference; Natural hazards; Spatio-temporal statistics; Stochastic processes; Wind storm.

Résumé

Pour chaque type d'évènements extrêmes, l'estimation du risque associé repose sur l'utilisation de deux catégories de catalogues : historiques et artificiels. Pour cette dernière, il n'existe aucune méthode qui soit en mesure de produire des évènements dont la sévérité est mathématiquement extrapolée au-delà des seuils historiques. La présente thèse a pour but de contribuer au développement de générateurs aléatoires d'évènements fondés sur la théorie des valeurs extrêmes, avec une attention particulière portée aux catastrophes naturelles.

Les bases de données retraçant l'histoire de notre climat sont d'origines multiples, mais celles produites par les modèles climatiques sont attrayantes pour leur large couverture spatiale et leur homogénéité. D'un point de vue statistique, ces modèles fournissent une gigantesque quantité de données quadrillant l'ensemble du globe qui peut être exploitée pour estimer précisément le risque climatique. D'une part, la présente thèse décrit des méthodes pour l'estimation de processus extrêmes, dont la complexité computationnelle est suffisamment faible pour être applicable à des jeux de données de grande taille. D'autre part, le comportement extrême limite des données agrégées est relié à celui des observations ponctuelles, comme celles produites par les stations météorologiques. Ce résultat est ensuite utilisé pour estimer le comportement local des queues de distributions à partir d'observations agrégées.

Enfin, la construction de générateurs aléatoires d'évènements extrêmes nécessite d'étendre les résultats classiques sur l'analyse des excès de seuil aux processus stochastiques continus. Nous développons une méthodologie dans laquelle la caractérisation d'évènements extrêmes complexes peut être motivée par une expertise spécifique au domaine d'application. La contribution décrit la distribution théorique limite des excès de seuils fonctionnels, appelée processus de r -Pareto généralisé. Ce dernier est présenté comme l'équivalent fonctionnel de la distribution de Pareto généralisée, et ses méthodes d'estimation, de simulation et de validation de modèle sont discutées. Ces résultats sont ensuite appliqués pour construire un générateur aléatoire de tempêtes extrêmes en Europe.

Acknowledgements

Mots clés : Catastrophes naturelles; Inférence en haute dimension; Processus de r -Pareto généralisé; Réduction d'échelle; Score de gradient; Statistiques spatio-temporelles; Tempêtes; Théorie des valeurs extrêmes; Vraisemblance censurée.

Contents

Acknowledgements	v
Abstract (English/Français)	vii
Introduction	1
Motivation	1
Outline of the thesis	3
1 Asymptotic tail distributions: theory and models	5
1.1 Probabilistic functional spaces	5
1.2 Univariate asymptotic distributions	6
1.2.1 Preliminary: convergence type and regular variation	6
1.2.2 Limit distribution of block maxima	8
1.2.3 Limit distribution of threshold exceedances	11
1.3 Multivariate and functional limits of componentwise maxima	12
1.3.1 Max-domain of attraction	12
1.3.2 Max-stable process	14
1.3.3 Exponent and spectral measure	15
1.3.4 Spectral representation	17
1.4 Multivariate and functional limits of threshold exceedances	18
1.4.1 Functional regular variation	18
1.4.2 Limit distribution of functional r -exceedances	23
1.5 Characterization of bivariate tail dependence	28
1.5.1 Asymptotic independence	28
1.5.2 Asymptotic dependence	30
1.6 Spatio-temporal extremal processes based on Gaussian random functions	32
1.6.1 Classical geostatistics	32
1.6.2 Extremal processes based on Gaussian random functions	42
1.7 Discussion	47

Contents

2	High-dimensional peaks-over-threshold inference	49
2.1	Introduction	49
2.2	Modelling exceedances over a high threshold	50
2.2.1	Functional regular variation	50
2.2.2	r -Pareto processes	52
2.2.3	Extreme value processes associated to log-Gaussian random functions	54
2.3	Inference for r -Pareto processes	55
2.3.1	Generalities	55
2.3.2	Efficient censored likelihood inference	57
2.3.3	Score matching	59
2.4	Simulation study	62
2.4.1	Exact simulation	62
2.4.2	Domain of attraction	63
2.5	Extreme rainfall over Florida	65
2.5.1	General	65
2.5.2	Multivariate extremal dependence model	65
2.5.3	Model checking and simulation	67
3	Extremal behaviour of aggregated data	71
3.1	Introduction	71
3.2	Limit results for extremes of aggregated data	73
3.2.1	Background on extremes	73
3.2.2	Univariate limiting distributions of aggregated data	74
3.2.3	Multivariate limiting distributions of aggregated data	77
3.3	Statistical Inference	79
3.3.1	Setting	79
3.3.2	Fitting based on marginal estimates	80
3.3.3	Censored likelihood for the joint tail behaviour	81
3.4	Simulation of Extreme Events	82
3.5	Application: downscaling extremes	85
3.5.1	Statistical downscaling	85
3.5.2	Application to extreme temperature in the South of France	86
4	Functional peaks-over-threshold analysis	91
4.1	Introduction	91
4.2	Modelling exceedances over a high threshold	94
4.2.1	Univariate model	94
4.2.2	Functional regular variation and \mathcal{R} -Pareto processes	95
4.3	Limiting distributions of r -exceedances	97

4.4	Generalized r -Pareto processes	100
4.4.1	Definition	100
4.4.2	Construction and marginal properties	102
4.4.3	Simulation	103
4.4.4	Link to max-stable processes	105
4.5	Statistical inference	106
4.6	Model validation	110
4.7	Discussion	111
5	Statistical modelling of extreme windstorms over Europe	113
5.1	Introduction	113
5.2	Risk estimation for extreme windstorms	115
5.3	Data set and region of study	115
5.4	Storm definition and frequency modelling	116
5.5	Marginal model	118
5.6	Dependence model	121
5.7	Simulations	124
5.8	Discussion	125
6	Conclusion and future work	129
A	Proofs for Chapter 1	131
A.1	Theorem 1.6	131
A.2	Corollary 1.2	132
A.3	Corollary 1.3	132
A.4	Theorem 1.8	133
A.5	Equation (1.30)	133
B	Supplementary material for Chapter 2	135
B.1	High-dimensional censored likelihood	135
B.1.1	Computational considerations	135
B.1.2	Algorithm for multivariate normal distribution function estimation	136
B.2	Properness of the gradient scoring rule	137
B.3	Gradient score for Brown–Resnick processes	138
B.4	Average computation times of the fitting procedures	139
B.5	Detailed results for simulations in Section 2.4	139
B.6	Proof of Proposition 2.1	140
B.7	Pareto process simulation	143

Contents

C	Supplementary material for Chapter 3	145
C.1	Proof of Theorem 3.2	145
C.2	Background and formula related to Hüsler–Reiss distributions	147
C.2.1	Hüsler–Reiss distributions	147
C.2.2	Explicit formulas for extremes of aggregated data	148
C.3	Simulation study	151
C.3.1	Gumbel case	151
C.3.2	Weibull case	153
C.3.3	Fréchet case	154
C.4	Model assessment for the downscaling application	156
D	Proofs of theoretical results in Chapter 4	157
D.1	Equation (4.5)	157
D.2	Theorem 4.1	157
D.3	Corollary 4.1	159
D.4	Derivation of (4.13)	160
D.5	Derivation of (4.22)	160
D.6	Derivation of (4.27)	161
D.7	Poisson process and binomial representation equivalence	162
E	Supplementary material for Chapter 6	165
E.1	Diagnostic plots for modelling the frequency of wind storms	166
	Bibliography	183
	Curriculum Vitae	185

Introduction

Motivation

'It's rough out there'.

It was with this title that in January 2008 the front page of *The Economist* compared the global sub-prime crisis to a category 5 cyclone. In the past few decades, the number of unexpected and extreme events has seemed to multiply, a trend partly conveyed by a large increase in media coverage mainly focused on tracking down responsibilities. For instance, a broad consensus attributes the roots of the 2007 financial crisis to faulty mathematical models embodied by the now famous 'formula that killed Wall Street'. In 2009, Italy adopted a different point of view and opted for human culprits when sentencing six scientific experts for their failure to give an 'adequate warning' before the earthquake of l'Aquila, which caused 300 deaths and more than 1500 other casualties.

These stories illustrate how the expectations of the general public changed over the past few years: citizens are no longer satisfied by the assurance of a quick recovery provided by national solidarity, but expect such catastrophes to be, if not avoided, at least mitigated as much as possible; if not, regulators are held responsible. Thus risk mitigation for natural catastrophes such as floods, cyclones, earthquakes, heatwaves and droughts has taken a central place in the political decision process, especially since the assessment of climate change, whose consequences are far from fully understood.

To get an idea of the human and financial impact of natural hazards on our society, consider the case of extreme wind storms. During summer 2005, hurricane Katrina, in Figure 1, struck the Gulf coast of the United States. Its unusually deep trajectory into the land caused more than 1000 fatalities and around 125 billion dollars of damage. Subsequently about 80% of the city of New Orleans was flooded due to fatal engineering flaws in the flood protection system. Twelve years later, hurricanes Harvey, Irma and Maria consecutively swept over the east coast of the U.S. causing a total of 278 billion dollars of damage within a month. Understanding the frequency and

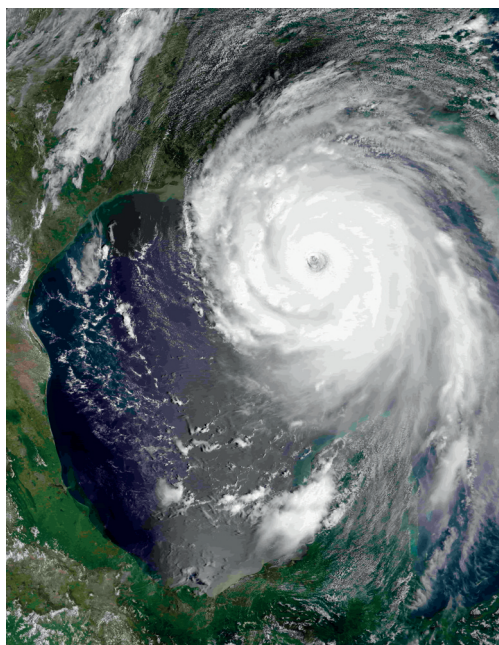


Figure 1 – Satellite picture of hurricane Katrina. The storm struck the east coast of the U.S. during summer 2005, causing more than 1000 fatalities and about 125 billion dollars of damage.

intensity of extreme windstorms is key to ensuring the safety of people and infrastructure. For instance, the likelihood of flooding caused by extreme windstorms is expected to increase with the forecasted rise in sea levels caused by global warming, and flood protection systems must be modified accordingly. Similarly, for (re-)insurers, understanding and quantifying their risk exposure is essential to ensure sufficient financial resources and avoid potential bankruptcy following a series of unusually severe disasters.

Most current risk assessment for extreme windstorms relies on catalogues of events that are used as ‘stress tests’ for human infrastructure or insurance portfolios. These catalogues usually consist of historical records or are artificially generated by climate models. However, we know of no existing method that can generate completely new extreme events and allow mathematically justified extrapolation above observed intensity levels. Stochastic weather generators are mathematical models that create random but realistic events which could be used to enlarge or even create catalogues. However, existing generators have not been designed for extreme events.

Extreme value theory describes the statistical behaviour of extreme events, and provides a mathematical framework to extrapolate above the intensity of histori-

cal records. This thesis develop a methodology to analyse complex extreme events and build stochastic generators that go beyond the Gaussian approximation, the central component of the formula that killed Wall Street. From numerical complexity to theoretical development, this work tackles the various challenges posed by risk estimation of natural hazards, with a focus on extreme windstorms. Ultimately, we build a generator of storms over Europe that can be used to assess the resilience of human infrastructure and serve as a basis for political and for economic decisions.

Outline of the thesis

Chapter 1 gives notions on probabilistic spaces and summarizes classical results of extreme value theory. We present the generalized extreme value and generalized Pareto distributions, which respectively describe the univariate limiting distribution of block maxima and threshold exceedances. Max-stable processes are introduced as the functional limits of component-wise block maxima, and we give a new derivation of existing results on functional exceedances. For applications to the environment, specific measures and models of dependence are required; we describe them both for classical spatio-temporal statistics and for extremal processes.

In environmental applications, data sets are likely to be large, and thus statistical inference methodologies need to be tractable and computationally efficient. Indeed, high-dimensional applications have been mostly limited to Gaussian models due to their relative computational simplicity. Chapter 2 reviews existing inference procedures for extremal processes. One of the most successful methodologies, the peaks-over-threshold censored likelihood, requires heavy computations, so we develop an efficient algorithm that makes inference tractable for larger data sets than with previous implementations. However, even with optimized code, inference is limited to ~ 500 dimensions, although it is not rare in environmental applications to deal with thousands of measurements. Thus we introduce an alternative inference procedure based on proper scoring rules, which is both computationally cheaper and robust. We compare these methods in a simulation study and show their practical applicability by estimating the extremal dependence of two types of severe rainfall.

The largest source of spatial data in environmental applications is climate models, which are defined on grids. One way to conceptualize their gridded nature is to suppose that they are the result of the aggregation of an underlying and non-observed physical process. In Chapter 3 we derive the limiting tail distribution of aggregated data. Our results allow us to generalize a classical measure of tail dependence, namely the extremal coefficient, to reflect the impact of aggregation on the observations. We

Contents

apply our theory to estimate the local tail distribution of extreme temperatures in the South of France using only coarse gridded data, a procedure known as downscaling.

The classical approach for peaks-over-threshold analysis of stochastic processes relies on specific notions of exceedance defined once the marginal behaviour of the process has been standardized, for instance to unit Fréchet. In Chapter 4, we generalize this methodology to allow more general definitions of exceedance and derive the limiting distribution of functional excesses when they are defined directly on the original process. We introduce the generalized r -Pareto process, a generalization of the generalized Pareto distribution to functions, that uniquely describes the functional limit tail distribution.

Finally, Chapter 5 illustrates the methodology of Chapter 4 with the construction of a stochastic weather generator for extreme windstorms over Europe. We propose a complete analysis with preliminary data exploration, marginal and dependence modelling and fitting diagnostics. The model is convincing but further work is required to quantify its uncertainties and to build more realistic dependence structures.

1 Asymptotic tail distributions: theory and models

This chapter introduces the background necessary to a proper mathematical description of asymptotic tail distributions, followed by a presentation of univariate extreme value theory. Then these results are generalized to functional component-wise maxima. In Section 1.4, we review existing results on functional peaks-over-threshold analysis, and give new derivations of the convergence results. We highlight the limits of the current state of the literature on functional threshold exceedances to motivate Chapter 4. Section 1.5 introduces tools for measuring dependence, while Section 1.6 is a broad survey of spatio-temporal extremal processes built using Gaussian random functions.

1.1 Probabilistic functional spaces

In this section, we introduce the mathematical objects required to derive the asymptotic distributions of component-wise maxima and threshold exceedances. While these concepts are necessary for formal mathematical derivations, the practitioner could just skip this section. However we give insights on the practical implications of these notions when using the statistical models described in the following chapters, in order to make these abstract notions more concrete.

A *metric* is a non-negative symmetric function $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ satisfying the identity of indiscernibles, i.e.,

$$d(x, y) = 0 \Rightarrow x = y, \quad x, y \in \mathcal{M},$$

and the triangle inequality

$$d(x, z) \leq d(x, y) + d(y, z), \quad x, y, z \in \mathcal{M}.$$

Chapter 1. Asymptotic tail distributions: theory and models

A set \mathcal{M} associated with a metric is called a *metric space*. If \mathcal{M} contains a countable, dense subset, i.e., there exists a sequence $\{x_n\}_{n=1}^{\infty}$ of elements of the space such that every nonempty open subset of \mathcal{M} contains at least one element of the sequence, then \mathcal{M} is *separable*. The space \mathcal{M} is *complete* if every Cauchy sequence in \mathcal{M} converges in \mathcal{M} . To study environmental phenomena, we consider functions defined on a subset S of a complete separable metric space, e.g., $\mathcal{M} = \mathbb{R}^2$ for spatial applications or $\mathcal{M} = \mathbb{R}^3$ for the spatio-temporal case. For instance, when modelling extreme windstorms in Chapter 5, we chose S to represent the region of interest, western Europe, and its state over a 24-hour window.

A *vector space* is a space \mathcal{M} for which $x+y \in \mathcal{M}$ and $ax \in \mathcal{M}$ for any $x, y \in \mathcal{M}$ and $a \in \mathbb{R}$. A *Banach space* is a complete vector space associated with a distance $d(x, y) = \|x - y\|$ ($x, y \in \mathcal{M}$), called a norm, such as the infinity norm $\|x\|_{\infty} = \sup_{s \in S} x(s)$. To derive functional limit distributions, we consider the Banach space $C(S)$ of continuous functions $x : S \rightarrow \mathbb{R}$ endowed with the infinity norm. With this choice, we suppose that the underlying physical process that we aim to model, such as a wind field over Europe, is continuous in space and time.

A *Borel set* is any set in \mathcal{M} that can be formed from open sets through countable unions, countable intersections, and relative complements. The collection of all Borel sets on \mathcal{M} is known as the *Borel σ -algebra* and denoted $\mathcal{B}(\mathcal{M})$.

A *probability space* is a triplet including a sample space, e.g., the space $C(S)$ of real-valued continuous functions over S , a σ -algebra containing all possible events, in our case $\mathcal{B}\{C(S)\}$, and a probability measure P assigning probabilities to these events.

These are the basic mathematical notions required to derive the functional limit distributions of extreme events.

1.2 Univariate asymptotic distributions

1.2.1 Preliminary: convergence type and regular variation

Let \mathcal{M} be a complete, separable metric space and let $C(\mathcal{M}, \mathbb{R})$ denote the space of real-valued continuous functions on \mathcal{M} . Given a sequence $\{X_n\}_1^{\infty}$ of random variables taking values in \mathcal{M} with distribution functions $F_n(x) = \Pr(X_n \leq x)$, $x \in \mathcal{M}$, we say that $\{X_n\}_1^{\infty}$ *converges weakly* to X , if for any bounded function $f \in C(\mathcal{M}, \mathbb{R})$,

$$\int_{\mathcal{M}} f(x) F_n(dx) \rightarrow \int_{\mathcal{M}} f(x) F(dx), \quad n \rightarrow \infty. \quad (1.1)$$

1.2. Univariate asymptotic distributions

A simpler characterization is obtained when $\mathcal{M} = \mathbb{R}$ and restricting \mathcal{M} to the subset

$$\mathcal{M}(F) = \{x \in \mathcal{M} : F(x) < \infty \text{ and } F \text{ is continuous at } x\};$$

in this case, weak convergence is equivalent to

$$F_n(x) \rightarrow F(x), \quad n \rightarrow \infty, \quad x \in \mathcal{M}(F).$$

In other words, univariate weak convergence is equivalent to the convergence of probability measures, but, in practice and for a multivariate extension, dealing with distribution functions can be inconvenient, especially from a statistical point of view, and for this reason (1.1) is usually preferred.

In extreme value theory, probability measures with infinite masses on certain sets, such as the origin $\{0\}$, are common, with the consequence that equation (1.1) may not be finite. *Vague convergence* (Resnick, 2007, p. 49) is a generalized notion of convergence capable of handling infinite measures. Let \mathcal{M} now be a locally compact space with a countable basis and σ -field $\mathcal{B}(\mathcal{M})$. A measure Λ is called *Radon* if $\Lambda(K) < \infty$, for any compact subset K of \mathcal{M} . Let $M_+(\mathcal{M})$ be the set of non-negative Radon measures on \mathcal{M} . For a sequence of measures $\Lambda_n \in M_+(\mathcal{M})$, we say that $\{\Lambda_n\}_1^\infty$ converges vaguely to Λ , if for all continuous, real-valued functions f with compact support on \mathcal{M} ,

$$\int_{\mathcal{M}} f(x) \Lambda_n(dx) \rightarrow \int_{\mathcal{M}} f(x) \Lambda(dx), \quad n \rightarrow \infty.$$

When studying tails of distributions, as in Section 1.2.2 for instance, vague convergence can be linked with the class of regularly-varying functions. A measurable function $F : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is *regularly-varying* at ∞ with index α , written $F \in RV_\alpha$, if

$$\lim_{t \rightarrow +\infty} \frac{F(tx)}{F(t)} = x^\alpha, \quad x > 0.$$

Similarly, a random variable X , with survival function $\bar{F} = 1 - F$, is regularly varying of index α if

$$\bar{F}(x) = x^\alpha L(x), \quad x > 0,$$

where L is a slowly-varying function, i.e., for any $c > 0$, $\lim_{x \rightarrow +\infty} L(cx)/L(x) = 1$. The notions described above are key components in deriving the theoretical limiting distributions of block maxima and threshold exceedances.

1.2.2 Limit distribution of block maxima

In this Section, X_1, \dots, X_n denote independent identically distributed copies of a random variable X with distribution function F . The statistical analysis of extremes was first developed for block maxima (Gumbel, 1958, Section 5.1): The variable

$$M_n = \max_{i=1, \dots, n} X_i$$

is called a block maximum and its distribution function is $\Pr(M_n \leq x) = F^n(x)$. Let x^* denote the upper bound $\sup\{x : F(x) < 1\}$ of the support of X , which is not necessarily finite. Then

$$M_n \rightarrow x^*, \quad n \rightarrow \infty; \quad (1.2)$$

in other words, the distribution of M_n degenerates to a distribution with unit mass on x^* as $n \rightarrow \infty$. When studying the sample average, the central limit theorem states that the distribution of $\sum_{i=1}^n X_i$ converges to standard normal as $n \rightarrow \infty$, under mild conditions and suitable affine normalization. Following the same principle, the distribution M_n is studied for rescaling sequences chosen such that the degeneracy in (1.2) is avoided.

Theorem 1.1 (Fisher and Tippett (1928), Gnedenko (1943)) *Suppose there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$ such that*

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

where G is not degenerate. Then G belongs to one of the following three classes:

- *Fréchet*: $G(x) = \begin{cases} 0, & x < b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x \geq b, \end{cases}$
- *Gumbel*: $G(x) = \exp\left\{-\exp\left(-\frac{x-b}{a}\right)\right\}, \quad x \in \mathbb{R},$
- *Weibull*: $G(x) = \begin{cases} \exp\left\{-\left(-\frac{x-b}{a}\right)^\alpha\right\}, & x < b, \\ 1, & x \geq b. \end{cases}$

for $\alpha, a > 0$ and $b \in \mathbb{R}$.

These three classes constitute the family of extreme-value distributions and fully describe the class of max-stable distributions, which we now introduce.

1.2. Univariate asymptotic distributions

Definition 1.1 (Distribution of same type) *Two distribution functions F and G are of the same type, if for some $a > 0$, $b \in \mathbb{R}$,*

$$F(x) = G(ax + b), \quad x \in \mathbb{R}.$$

A random variable X is said to be *max-stable* if for any $n \geq 1$, F^n is of the same type as F , i.e., if there exist $a_n > 0$, and $b_n \in \mathbb{R}$, such that

$$\frac{M_n - b_n}{a_n} \stackrel{D}{=} X,$$

where $\stackrel{D}{=}$ denotes equality in distribution. In Theorem 1.1, the class of max-stable distributions is divided into three categories, but Proposition 1.1 motivates a unified parametrization of this family.

Proposition 1.1 (Resnick (1987), p. 7) *Let F and G be non-degenerate distribution functions. Suppose that for a sequence $\{F_n\}_1^\infty$ of distributions there exist $a_n, a'_n > 0$ and $b_n, b'_n \in \mathbb{R}$ such that*

$$F_n(a_n x + b_n) \rightarrow F(x), \quad F_n(a'_n x + b'_n) \rightarrow G(x),$$

weakly as $n \rightarrow \infty$, then

$$\frac{a_n}{a'_n} \rightarrow a > 0, \quad \frac{b'_n - b_n}{a_n} \rightarrow b \in \mathbb{R}, \quad n \rightarrow \infty,$$

and

$$G(x) = F(ax + b).$$

Proposition 1.1 implies that the limit distribution of a sequence $\{F_n\}_1^\infty$ is unique up to affine transformations. For this reason, the choice of rescaling sequences can be made so that we obtain a single family of distributions, the generalized extreme-value distributions (Fisher and Tippett, 1928),

$$G_\xi(x) = \begin{cases} \exp \left[- \left\{ 1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right\}_+^{-1/\xi} \right], & \xi \neq 0, \\ \exp \left\{ - \exp \left(- \frac{x-\mu}{\sigma} \right) \right\}, & \xi = 0, \end{cases} \quad (1.3)$$

where $x \in \{x \in \mathbb{R} : 1 + \xi(x - \mu)/\sigma \geq 0\}$, $\sigma > 0$ and $\mu \in \mathbb{R}$. In this family, the parameters σ and μ are respectively the scale and location of the distribution and the tail decay regime is determined by the shape parameter ξ (Jenkinson, 1955). The Fréchet type corresponds to $\xi > 0$, and is characterized by polynomial tail decay and a support with

Chapter 1. Asymptotic tail distributions: theory and models

infinite upper bound. For $\xi < 0$, the generalized extreme value distribution is bounded above by $\mu - \sigma/\xi$, giving the Weibull type. The Gumbel class is interpreted as the limit of the Weibull and the Fréchet distributions when $\xi \rightarrow 0$.

For an arbitrary random variable X with distribution function F , we say that X is in the *max-domain of attraction* of an extreme value distribution G if there exist sequences $\{a_n\}_{n=1}^{\infty} > 0$ and $\{b_n\}_{n=1}^{\infty} \in \mathbb{R}$ such that $F^n(a_n x + b_n)$ converges to G , i.e,

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G, \quad n \rightarrow \infty, \quad (1.4)$$

and we write $X \in \text{MDA}(G)$. Characterization of max-domain of attraction has been extensively studied (von Mises, 1964; Leadbetter, 1983; Resnick, 1987, Chapter 1); for the limits of classical distribution functions, see Beirlant et al. (2004, p. 59, 62, 72).

As mentioned in Section 1.2.1, a possible characterization of the max-domain of attraction is obtained within the framework of regularly varying functions. We focus on this characterization, as regular variation will be a key component for Section 1.4 and Chapter 4. Following Resnick (1987, p. 54), a random variable X belongs to the Fréchet domain of attraction if and only if its survival function $1 - F$ is regularly varying with index $-\alpha < 0$. In this case, the sequence $\{a_n\}_1^{\infty}$ tends to infinity as $n \rightarrow \infty$ and thus, using Proposition 1.1, $\{b_n\}_1^{\infty}$ can equivalently be chosen equal to 0 for any $n \geq 1$. These properties are key components in Chapter 3 to derive our results in the Fréchet domain of attraction. For the Weibull domain of attraction, regular variation of the random variable $x^* - X^{-1}$ with index $-\alpha < 0$ is necessary and sufficient for membership of this class of tail decay. In this case, the sequence $\{b_n\}_1^{\infty}$ can be chosen constant and equal to the finite upper bound x^* of X .

From a practical perspective, for independent identically distributed copies X_i ($i = 1, \dots, mn$) of an arbitrary random variable X , we create blocks maxima

$$M_{n,j} = \max_{i=(j-1)n+1, \dots, jn} X_i, \quad j = 1, \dots, m,$$

for sufficiently large n , and then we estimate $a_n > 0$ and $b_n, \xi \in \mathbb{R}$ such that

$$\Pr\left(\frac{M_{n,i} - b_n}{a_n} \leq x\right) \approx G_{\xi}(x). \quad (1.5)$$

Equation (1.5) emphasises that the generalized extreme value distribution provides a unified framework to model block maxima of a random variable X for a sufficiently large n .

1.2.3 Limit distribution of threshold exceedances

The statistical analysis of extremes using block maxima (Gumbel, 1958, Section 5.1) is widely used, but the reduction of a complex dataset to maxima with a block size such that approximation (1.5) is sufficiently good can lead to a significant loss of information (Madsen et al., 1997), so modelling exceedances over a high threshold is often preferred in applications (Davison and Smith, 1990).

Let X be a random variable in the max-domain of attraction of a generalized extreme value distribution G_ξ . Then by Taylor expansion of the logarithm, equation (1.4) is equivalent to

$$n\Pr(X > b_n + a_n x) \rightarrow -\log G_\xi(x), \quad n \rightarrow \infty, \quad (1.6)$$

for all $x \in \mathbb{R}$ such that $1 + \xi(x - \mu)/\sigma > 0$. Equation (1.6) can also be formulated in terms of convergence of measures and in this case, vague convergence is required.

Theorem 1.2 (Threshold exceedances) *Suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$, and $\xi \in \mathbb{R}$ such that*

$$\left. \begin{aligned} n\Pr \left\{ \left(1 + \xi \frac{X - b_n}{a_n} \right)_+^{1/\xi} > x \right\} \\ n\Pr \left\{ \exp \left(\frac{X - b_n}{a_n} \right)_+ > x \right\} \end{aligned} \right\} \rightarrow \Lambda(x), \quad x > 0, \quad n \rightarrow \infty,$$

where $(x)_+ = \max(0, x)$. Then Λ is either degenerate or equal to $\Lambda(x) = x^{-1}$.

Similarly to block maxima, conditions for the existence of sequences $\{a_n\}_1^\infty$ and $\{b_n\}_1^\infty$ such that Λ is not degenerate are linked to the notion of regular variation: If X is regularly varying with index $-1/\xi < 0$, then there exist a sequence $\{a_n\}_{n=1}^\infty > 0$ and $\{b_n\}_1^\infty \in \mathbb{R}$ such that

$$n\Pr \left(\frac{X - b_n}{a_n} \geq x \right) \rightarrow \left(1 + \xi \frac{x}{\sigma} \right)^{-1/\xi}, \quad n \rightarrow \infty,$$

in $M_+(0, \infty]$ (Resnick, 2007, Theorem 3.6), where similarly to block maxima, $a_n \rightarrow \infty$ as $n \rightarrow \infty$ and $\{b_n\}_1^\infty$ can be chosen equal to 0. Similar results can be obtained for the Weibull case when the variable $x^* - X^{-1}$ is regularly varying with index $-1/\xi$.

Form a practical point of view, Theorem 1.2 provides a basis for a unified description of the tails of distributions in terms of threshold exceedances. For a large enough threshold $u < \inf\{x : F(x) = 1\}$, the tail distribution of X can be approximated by a generalized Pareto distribution $H_{(\xi, \tilde{\sigma})}$ (Davison and Smith, 1990), i.e.,

$$\Pr(X - u > x \mid X > u) \approx H_{(\xi, \tilde{\sigma})}(x) = \begin{cases} (1 + \xi x / \tilde{\sigma})_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x / \tilde{\sigma}), & \xi = 0, \end{cases} \quad (1.7)$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu) > 0$ and $a_+ = \max(a, 0)$. Similarly to the max-stable case, if the shape parameter ξ is negative, then x must lie in the interval $[0, -\sigma/\xi]$, whereas x can take any positive value with positive or zero ξ . The random variable X is said to belong to the Weibull, the Gumbel or the Fréchet family if its shape parameter is respectively negative, zero or positive. These results can also be motivated on the basis of a point process characterization of high-level exceedances; see Leadbetter (1983), Falk et al. (1994) and Embrechts et al. (1997, pp. 237–247) for more details.

As a consequence of equation (1.7), Davison and Smith (1990) propose to approximate the upper tail of the distribution function of a random variable X by

$$F(x) \approx 1 - \zeta_u H_{(\xi, \sigma, \mu)}(x - u), \quad x > u,$$

where $u > 0$ is a sufficiently high threshold, and ζ_u , the probability that X exceeds the threshold u , is determined by u .

In their simplest form, models for univariate extreme value theory apply to independent and identically-distributed variables, but the theory has been used for time series (Leadbetter, 1983; Hsing et al., 1988; Beirlant et al., 2004, p. 383), non-stationary (Smith, 1989; Chavez-Demoulin and Davison, 2005, 2012) and spatial data (Davison and Gholamrezaee, 2012). The analysis of block maxima and threshold exceedances can also be generalized to random vectors and continuous processes.

1.3 Multivariate and functional limits of componentwise maxima

Due to recent events and because the impact of global warming is not well understood, there has been a surge of interest in environmental applications, especially regarding severe climatic events such as floods, windstorms, and heatwaves, which cannot be modelled using only univariate extreme value theory. Multivariate and functional extreme value theory was first developed for component-wise maxima by extending the generalized extreme value distribution. In this section, we describe only functional asymptotic distributions because the multivariate case can be derived by replacing the random process X by a random vector.

1.3.1 Max-domain of attraction

Let S be a compact subset of a complete separable metric space \mathcal{M} , let $C(S)$ denote the space of real-valued continuous functions on S equipped with the supremum

1.3. Multivariate and functional limits of componentwise maxima

norm $\|\cdot\|_\infty$, defined by $\|x\|_\infty = \sup_{s \in S} |x(s)|$, and let $\mathcal{B}\{C(S)\}$ be the Borel σ -algebra associated to $C(S)$.

For independent copies X_i ($i = 1, \dots, n$) of a stochastic process $X = \{X(s) : s \in S\}$, the process of component-wise maxima M_n ($n = 1, 2, \dots$) is defined as the local maximum over a block of n observations, i.e.,

$$M_n(s) = \max_{i=1, \dots, n} X_i(s), \quad s \in S.$$

For instance, if X represents the daily mean temperature over S , e.g., Switzerland, then, for $n = 360$, M_n corresponds to the pointwise annual maximum daily mean temperature over the region. The stochastic process X is said to belong to the *max-domain of attraction* of some process Z , if there exist sequences of functions $a_n : S \rightarrow (0, \infty)$, $b_n : S \rightarrow \mathbb{R}$, all continuous in $s \in S$, and $\xi \in \mathbb{R}$, such that $(M_n - b_n)/a_n$ converges in distribution to Z on the space $C(S)$ as $n \rightarrow \infty$, i.e.,

$$\mathcal{L} \left\{ \max_{i=1, \dots, n} \frac{X_i(s) - b_n(s)}{a_n(s)}, s \in S \right\} \longrightarrow \mathcal{L}\{Z_\xi(s), s \in S\}, \quad n \rightarrow \infty, \quad (1.8)$$

where $\mathcal{L}(\eta)$ denotes the law of a process η and

$$Z_\xi(s) = \begin{cases} \operatorname{sgn}(\xi) Z(s)^\xi, & \xi \neq 0, \\ \log Z(s), & \xi = 0, \end{cases} \quad s \in S.$$

It follows that for any $s \in S$, the univariate random variable $X(s)$ belongs to the max-domain of attraction of an extreme value distribution

$$G_\xi(x) = \begin{cases} \exp[-\{\operatorname{sgn}(\xi)x\}^{-1/\xi}], & \xi \neq 0, \\ \exp\{-\exp(-x)\}, & \xi = 0, \end{cases} \quad (1.9)$$

for all $x \in \mathbb{R}$ with $x\xi \geq 0$. Equation (1.9) explicitly links the functional generalization (1.8) with the univariate results presented in Section 1.2.2. Here, for simplicity, we present results for $\xi \in \mathbb{R}$, but the definitions (1.8) and (1.9) remain valid if the scalar ξ is replaced by a continuous function $\xi : S \rightarrow \mathbb{R}$; see de Haan and Ferreira (2006, pp. 294–296).

As in the univariate case, the existence of the scaling functions a_n and b_n for the Fréchet domain of attraction is linked to the notion of regular variation; see Resnick (2007, Chapter 6) for the multivariate case, and de Haan and Lin (2001) for functions. In Section 1.4, regularly varying sequences of measures are introduced to describe the limit distribution of functional exceedances, and their link with max-stable processes

is given in Chapter 4.

1.3.2 Max-stable process

Like with the univariate theory, the right-hand side of equation (1.8) describes the class of max-stable processes: for any $\xi \in \mathbb{R}$, the limit process Z_ξ is *max-stable*, i.e., for $Z_{\xi,i}$ ($i = 1, \dots, n$) independent copies of Z_ξ , there exist functions $a_n(s) > 0$ and $b_n(s) \in \mathbb{R}$ such that

$$\mathcal{L} \left\{ \frac{\max_{i=1, \dots, n} Z_{\xi,i}(s) - b_n(s)}{a_n(s)}, s \in S \right\} = \mathcal{L}\{Z_\xi(s), s \in S\}, \quad (1.10)$$

and then for any fixed location $s \in S$, $Z_\xi(s)$ is a generalized extreme-value random variable.

The introduction of the affine transformation in Theorem 1.1 was motivated by an analogy with the central limit theorem. For stochastic processes, $\sum_{i=1}^n X_i$ also converges, as $n \rightarrow \infty$, to a Gaussian random process with zero mean and unit variance but after subtraction of its mean function $\mu(s)$ and division by the square root of the variance function $\sigma^2(s)$ of X . When studying the tails of stochastic processes, not only may the location and scale vary over space, but also the tail regime with a functional shape parameter ξ . Thus, it is usually convenient to transform the process X to have unit Fréchet marginals (de Haan and Ferreira, 2006, Chapter 9). In this case, the normalizing sequences are known: $a_n(s) = n$ and we can conveniently choose $b_n(s) = 0$; see Proposition 1.1. Also, de Haan and Ferreira (2006, Theorem 9.2.1) justify this by proving that the convergence of M_n to a max-stable process is equivalent to, first, the convergence of $X(s)$ to a generalized extreme value distribution for any $s \in S$ and second, the convergence of the standardized process X^* ,

$$\mathcal{L} \left[X^*(s) = \frac{1}{1 - F_s\{X(s)\}}, s \in S \right] \rightarrow \mathcal{L}\{Z(s), S\}, \quad n \rightarrow \infty,$$

where F_s is the distribution function of X at $s \in S$ and $Z = Z_1$, known as a simple max-stable process. This two-step procedure, standardization followed by the analysis of the normalized process X^* , is similar to copula modelling (Nelsen, 2006), for which before studying the dependence, marginals are first standardized to a common distribution, for instance to unit Fréchet or unit Pareto (Klüppelberg and Resnick, 2008).

In practice, normalizing the process requires knowledge of the distribution function F_s , which is unknown in general. Coles et al. (1999) used a non-parametric approach

1.3. Multivariate and functional limits of componentwise maxima

based on the empirical distribution function

$$F_s^n(x) = (n+1)^{-1} \sum_{i=1}^n \mathbf{1}_{X_i(s) < x},$$

where the use of $n+1$ avoids mapping the maximum of the observed values to infinity, while Coles and Tawn (1991) propose the semi-parametric model

$$F_s(x) = \begin{cases} n^{-1} \sum_{i=1}^n \mathbf{1}_{X_i(s) < x} & x \leq u, \\ 1 - \zeta_u H_{(\xi, \sigma)}(x - u), & x > u, \end{cases} \quad (1.11)$$

where ζ_u is the probability that X exceeds the threshold u . Pre-processing X by normalizing the margins is a common practice in multivariate extreme-value theory, especially because statistical inference of the marginal tail behaviour jointly with the extremal dependence can be difficult. For this reason, we now suppose that the process X has been standardized to X^* , with unit Fréchet margins, and focus on the description of the properties of simple max-stable processes. Generalizations of max-stable processes to more general spaces, for instance to the space of real-valued càdlàg functions on $[0, 1]$ (Lindskog et al., 2014), exist, but are of limited interest for environmental applications.

1.3.3 Exponent and spectral measure

Let X^* be a stochastic process in the domain of attraction of a simple max-stable process Z and define the sequence of measures

$$\Lambda_n(A) = n \Pr \left(\frac{X^*}{a_n} \in A \right), \quad A \subset C_+(S)$$

where $C_+(S) = \{x \in C(S) \setminus \{0\} : x(s) \geq 0, s \in S\}$. The sequence $\{\Lambda_n\}_{n=1}^\infty$ converges to a limit measure Λ as $n \rightarrow \infty$ (de Haan and Ferreira, 2006, Theorem 9.3.1), i.e., for any Borel set A in $C_+(S)$

$$\Lambda_n(A) \rightarrow \Lambda(A), \quad n \rightarrow \infty.$$

The limiting measure Λ characterizes the distribution of the max-stable process Z ; more precisely, for any $u > 0$,

$$\Pr[Z \in \{x \in C_+(S) : \|x/u\|_\infty \leq 1\}] = \exp[-\Lambda\{A_\infty(u)\}], \quad (1.12)$$

where $A_\infty(u) = \{x \in C_+(S) : \|x/u\|_\infty \geq 1\}$. In multivariate extreme value theory, the quantity $\Lambda\{A_\infty(u)\}$ is known as the exponent measure and is used to characterize extremal dependence; see Section 1.5 for more details. Also, the limit measure is

Chapter 1. Asymptotic tail distributions: theory and models

homogeneous of order -1 , i.e., for any scalar $t > 0$,

$$\Lambda(tA) = t^{-1} \Lambda(A), \quad A \subset C_+(S). \quad (1.13)$$

Equation (1.13) is a key property of extremal processes, as it enables extrapolation of the tail probabilities above observed intensity levels. For instance consider a closed set A containing at least one observation of the sample Z_i ($i = 1, \dots, n$). Then for a large enough $t > 0$, the set $tA = \{tx : x \in A\}$ does not include any observation, but its probability $\Pr(tA)$ can be directly obtained from $\Pr(A)$ using equation (1.13). Also, combining (1.12) and (1.13), it is easy to retrieve the max-stability of Z , i.e., to prove that

$$\Pr\{n^{-1}Z \in A_\infty(r)\}^n = \Pr\{Z \in A_\infty(r)\}, \quad n = 1, 2, \dots$$

The representation of the limit measure Λ is not unique and the homogeneity property is key to alternative representations.

Theorem 1.3 (Giné et al. (1990)) *Let X^* lie in the max-domain of attraction of a simple max-stable process Z with limiting measure Λ . Then there exists a measure $\sigma_{\|\cdot\|}$ on $C_{\|\cdot\|}(S) = \{x \in C_+(S) : \|x\| = 1\}$ such that*

$$\int_{C_{\|\cdot\|}(S)} x(s) d\sigma(x) = 1, \quad s \in S, \quad (1.14)$$

and

$$\Lambda\left(\left\{x \in A_{\|\cdot\|}(r) : \frac{x}{\|x\|} \in A\right\}\right) = r^{-1} \sigma_{\|\cdot\|}(A), \quad (1.15)$$

where $A_{\|\cdot\|}(r) = \{x \in C_+(S) : \|x\| \geq r\}$.

In decomposition (1.15), r^{-1} measures the intensity of the process, while $\sigma_{\|\cdot\|}$, called the angular measure, characterizes the dependence of Z . Choosing different norms in (1.15) yields different decompositions. For instance, with the L_1 norm $\|x\|_1 = \int_S |x| dx$, the variable $W = Z/\|Z\|_1$ is called the pseudo-angle, and $\sigma_{\|\cdot\|_1}$ satisfies

$$\begin{aligned} \sigma_{\|\cdot\|_1}\{C_1(S)\} &= \Lambda\{C_+(S)\} \\ &= \int_{C_+(S)} r^{-2} dr d\sigma_{\|\cdot\|_1}(w) \\ &= \int_{C_1(S)} \int_{1/\|w\|_1} r^{-2} dr d\sigma_{\|\cdot\|_1}(w), \\ &= \int_S \int_{C_1(S)} w(s) d\sigma_{\|\cdot\|_1}(w) ds, \\ &= \|S\|_1, \end{aligned}$$

1.3. Multivariate and functional limits of componentwise maxima

which is thus independent of the distribution of W . For this reason, choosing the L_1 norm is often convenient because the computation of the measure $\Lambda\{A_{\|\cdot\|}(r)\}$, for instance to perform statistical inference, is straightforward. Finally in Theorem 1.3, any measure $\sigma_{\|\cdot\|}$ on $C_{\|\cdot\|}(S)$ satisfying condition (1.14) yields a valid limit measure Λ and thus a simple max-stable process on S , so the family of max-stable processes is not finite.

1.3.4 Spectral representation

The multiple decompositions obtained with (1.15) also impact the possible representations of a max-stable process with limit measure Λ .

Theorem 1.4 (de Haan (1984); Giné et al. (1990); Penrose (1992)) *Let Z be a simple max-stable process with continuous sample path on $C_+(S)$. Then Z can be written*

$$Z(s) = \max_{i \in \mathbb{N}} U_i W_i(s), \quad s \in S, \quad (1.16)$$

where $\{U_i : i \in \mathbb{N}\}$ are the points of a Poisson point process on $(0, \infty)$ with intensity measure $u^{-2} du$ and the spectral functions W_i ($i \in \mathbb{N}$), are independent copies of some non-negative, continuous process $\{W(s), s \in S\}$ with $E\{W(s)\} = 1$ for all $s \in S$ and $E\{\sup_{s \in S} W(s)\} < \infty$.

By making a parallel between the decomposition of Λ in (1.15) and Theorem 1.4, we use that for any simple max-stable process Z , we can write

$$Z(s) = \max_{i \in \mathbb{N}} U_i W_i(s), \quad s \in S, \quad (1.17)$$

where $\{U_i : i \in \mathbb{N}\}$ are the points of a Poisson point process on $(0, \infty)$ with intensity measure $u^{-2} du$, and the spectral functions W_i ($i \in \mathbb{N}$) now refer simply to a stochastic process on $C_1(S)$ with $E\{W(s)\} = 1$ ($s \in S$). Equation (1.16) is more useful in practice than (1.17) because the requirements on the process W are weaker, making it easier to build models.

In environmental applications, the decomposition (1.16) was compared to rainfall-storms (Smith, 1990): each couple (U_i, W_i) ($i = 1, \dots$) represents a storm contributing to the field of maxima, whose severity is given by U_i and whose spatial pattern is determined by the angular process W_i . This interpretation was exploited by Huser and Davison (2014) to model hourly rainfall using a spatio-temporal max-stable process.

Finally the representation (1.16) is the key component for most algorithms to simulate max-stable processes, but the maximum over an infinite set of points yields, without further restriction, inefficient or approximate algorithms (Schlather, 2002; Engelke et al., 2011; Oesting et al., 2011). Exploitation of either alternative but equivalent representations or more restrictive formulations of (1.16), for instance by Dieker and Mikosch (2015) and Dombry et al. (2016), leads to exact and efficient methods for simulation. Conditional simulation of max-stable process is also possible; see Dombry et al. (2013), Oesting and Schlather (2013), Bechler et al. (2015a), and Oesting et al. (2017a) for more details.

1.4 Multivariate and functional limits of threshold exceedances

Section 1.3 presented the extension of univariate block maxima and its limit distribution, namely the generalized extreme value distribution, to functions. For peaks-over-threshold analysis, the notion of exceedances for vectors or functions is not unique and thus a functional extension of the univariate theory is more delicate. In this section we present existing results on functional peaks-over-threshold analysis, which we revisit to make the parallel with their generalization in Chapter 4 easier. Similarly to the current literature, we focus on the case where the process X belongs to the Fréchet domain of attraction. In this case, generalization of peaks-over-threshold analysis to functions is possible within the framework of functional regular variation. Proofs of the theoretical results in this section can be found in Appendix A.

1.4.1 Functional regular variation

Definition

Let S be a compact metric space, such as $[0, 1]^2$ for spatial applications. We write $\mathcal{F}_+ = C\{S, [0, \infty)\}$ for the closed subset of the Banach space of continuous functions $x : S \rightarrow \mathbb{R}$ endowed with the uniform norm $\|x\|_\infty = \sup_{s \in S} |x(s)|$. A measurable closed subset \mathcal{C} of \mathcal{F}_+ is called a cone if $tx \in \mathcal{C}$ for any $x \in \mathcal{C}$ and $t > 0$. In the study of extremes, the cones $\mathcal{C} = \{0\}$ or $\mathcal{C} = \{x \in \mathcal{F}_+ : \inf_{s \in S} x(s) = 0\}$ are often excluded from \mathcal{F}_+ to avoid the appearance of limiting measures with infinite masses at the origin or on the coordinate axes. Let $M_{\mathcal{F}_+ \setminus \mathcal{C}}$ denote the class of Borel measures on $\mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ for any cone \mathcal{C} , and say that a set $A \in \mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ is bounded away from \mathcal{C} if $d(A, \mathcal{C}) = \inf_{x \in A, y \in \mathcal{C}} \|x - y\|_\infty > 0$.

A sequence of measures $\{\Lambda_n\}_1^\infty \subset M_{\mathcal{F}_+ \setminus \mathcal{C}}$ is said to converge to a limit $\Lambda \in M_{\mathcal{F}_+ \setminus \mathcal{C}}$,

1.4. Multivariate and functional limits of threshold exceedances

written $\Lambda_n \xrightarrow{\hat{w}} \Lambda$ (Hult and Lindskog, 2005), if $\Lambda_n(A) \rightarrow \Lambda(A)$ as $n \rightarrow \infty$, for all $A \in \mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ bounded away from \mathcal{C} with $\Lambda(\partial A) = 0$, where ∂A denotes the boundary of A . For equivalent definitions of this so-called \hat{w} -convergence, see Lindskog et al. (2014, Theorem 2.1).

A stochastic process X with sample paths in $\mathcal{F}_+ \setminus \mathcal{C}$ is *regularly varying* (Hult and Lindskog, 2005) if there exist a sequence of strictly positive continuous functions $\{a_n\}_{n=1}^{\infty}$ with $a_n(s) \rightarrow \infty$ as $n \rightarrow \infty$ for all $s \in S$ and a measure $\Lambda \in M_{\mathcal{F}_+ \setminus \mathcal{C}}$ such that

$$n \Pr(a_n^{-1} X \in \cdot) \xrightarrow{\hat{w}} \Lambda(\cdot), \quad n \rightarrow \infty; \quad (1.18)$$

then we write $X \in \text{RV}(\mathcal{F}_+ \setminus \mathcal{C}, a_n, \Lambda)$. The limiting measure Λ satisfies a homogeneity property of order $-1/\xi$, i.e., for any $t > 0$,

$$\Lambda(tA) = t^{-1/\xi} \Lambda(A), \quad A \in \mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C}), \quad (1.19)$$

for some positive ξ called the tail index.

Mapping theorem

For the remainder of this section, X is a regularly varying stochastic process and Λ_n refers to the sequence of measures $n \Pr(a_n^{-1} X \in \cdot)$ defined in (1.18). As the notion of exceedance for functional peaks-over-threshold analysis is not unique, a means to switch between definitions is required. For this reason, we introduce the mapping theorem to link the different representations.

Theorem 1.5 (Mapping theorem, Lindskog et al. (2014)) *Let \mathcal{F} and \mathcal{F}' be measurable and complete metric spaces with cones \mathcal{C} and \mathcal{C}' respectively and let h be a measurable mapping*

$$h: \mathcal{F}_+ \setminus \mathcal{C} \rightarrow \mathcal{F}'_+ \setminus \mathcal{C}'$$

such that $h^{-1}(A') = \{x \in \mathcal{F}_+ \setminus \mathcal{C} : f(x) \in A'\}$ is bounded away from \mathcal{C} for any $A' \in \mathcal{B}(\mathcal{F}'_+ \setminus \mathcal{C}') \cap h(\mathcal{F}_+ \setminus \mathcal{C})$ bounded away from \mathcal{C}' . Then the mapping $\tilde{h}: M_{\mathcal{F}_+ \setminus \mathcal{C}} \rightarrow M_{\mathcal{F}'_+ \setminus \mathcal{C}'}$ defined by

$$\tilde{h}(\Lambda) = \Lambda \circ h^{-1}$$

is continuous at $\Lambda \in M_{\mathcal{F}_+ \setminus \mathcal{C}}$ provided $\Lambda(D_h) = 0$, where D_h is the set of discontinuity points of h .

For the study of extremes, we will be interested in the exclusion from \mathcal{F}_+ of a specific family of cones that is tied to the definition of exceedances. First, we define a *risk func-*

Chapter 1. Asymptotic tail distributions: theory and models

tional $r : \mathcal{F}_+ \rightarrow [0, \infty)$ as a continuous functional satisfying a homogeneity property, i.e., a functional for which there exists $\kappa > 0$ such that

$$r(ax) = a^\kappa r(x), \quad x \in \mathcal{F}_+, \quad a > 0.$$

As $r(\cdot)$ could be replaced by $r(\cdot)^{1/\kappa}$ without loss of generality, below we assume that $\kappa = 1$. For any risk functional r , the set $\mathcal{C}_r = \{x \in \mathcal{F}_+ : r(x) = 0\}$ is a closed cone of \mathcal{F}_+ . We say that an r -*exceedance* is an event of the form $\{r(X) \geq u_n\}$, where the sequence of thresholds $u_n > 0$ is chosen such that $\Pr\{r(X) \geq u_n\} \rightarrow 0$ as $n \rightarrow \infty$.

Now we consider the mapping $h_r : \mathcal{F}_+ \setminus \mathcal{C} \rightarrow \mathcal{F}_+ \setminus \mathcal{C}'$, with $\mathcal{C}' = \mathcal{C} \cup \mathcal{C}_r$, defined as

$$h_r(x) = \begin{cases} x, & r(x) > 0, \\ 0, & r(x) = 0. \end{cases} \quad (1.20)$$

Applying Theorem 2.1 in Lindskog et al. (2014) and Theorem 1.5 with the mapping (1.20) ensures the convergence of $\{\Lambda_n\}_1^\infty$ restricted to the space $\{x \in \mathcal{F}_+ \setminus \mathcal{C} : r(x) > 0\}$.

Corollary 1.1 *Suppose $\Lambda_n \rightarrow \Lambda$ in $M(\mathcal{F}_+ \setminus \mathcal{C})$ as $n \rightarrow \infty$. Then*

$$\Lambda_n \circ h_r^{-1} \rightarrow \Lambda \circ h_r^{-1}, \quad n \rightarrow \infty, \quad (1.21)$$

in $M\{\mathcal{F}_+ \setminus (\mathcal{C} \cup \mathcal{C}_r)\}$.

Corollary 1.1 implies that if $X \in \text{RV}(\mathcal{F}_+ \setminus \mathcal{C}, a_n, \Lambda)$ then the stochastic process X is also regularly varying on $\mathcal{F}_+ \setminus (\mathcal{C} \cup \mathcal{C}_r)$ with limit measure $\Lambda \circ h_r^{-1}$, and the converse is true only if $\Lambda\{\mathcal{C}_r \setminus \mathcal{C}\} = 0$ or $\mathcal{C} = \mathcal{C}_r$. The mapping theorem and Corollary 1.1 are key components in describing the limit distribution of r -exceedances.

Pseudo-polar decomposition

Alternative representations of the limiting measure Λ are obtained from pseudo-polar transformations. For a norm $\|\cdot\|_{\text{ang}}$ on \mathcal{F}_+ , called the angular norm, and a risk functional r , a pseudo-polar transformation h_r^{pp} is a map $\mathcal{F}_+ \setminus \mathcal{C} \rightarrow [0, \infty) \times \mathcal{S}_r$ such that

$$h_r^{\text{pp}}(x) = \left(r' = r(x), w = \frac{x}{\|x\|_{\text{ang}}} \right), \quad x \in \mathcal{F}_+ \setminus \mathcal{C}, \quad (1.22)$$

where \mathcal{S}_r is the unit sphere $\{x \in \mathcal{F}_+ \setminus \mathcal{C} : r(x) > 0, \|x\|_{\text{ang}} = 1\}$. If $\{x \in \mathcal{F}_+ \setminus \mathcal{C} : r(x) = 0\} = \emptyset$, then h_r^{pp} is a homeomorphism with inverse

$$(h_r^{\text{pp}})^{-1}(r', w) = r' \times \frac{w}{r(w)}.$$

1.4. Multivariate and functional limits of threshold exceedances

Theorem 1.6 combines the family of pseudo-polar mappings and the mapping theorem to factorize the limiting measure Λ .

Theorem 1.6 *Suppose $X \in \text{RV}(\mathcal{F}_+ \setminus \{0\}, a_n, \Lambda)$ and let r be a risk functional. Then there exists $\xi > 0$ such that*

$$n\Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[r', \infty), \mathcal{W}\} \right] \xrightarrow{\hat{w}} \Lambda_\xi \{[r', \infty)\} \times \sigma_r(\mathcal{W}), \quad n \rightarrow \infty, \quad r' > 0, \quad \mathcal{W} \subset \mathcal{S}_r,$$

in $M(\mathcal{F}_+ \setminus \mathcal{C}_r)$, where

$$\Lambda_\xi \{[r', \infty)\} = (r')^{-1/\xi} \Lambda(A_r),$$

with $A_r = \{x \in \mathcal{F}_+ \setminus \mathcal{C} : r(x) \geq 1\}$ and σ_r is the probability measure on $\mathcal{B}(\mathcal{S}_r)$,

$$\sigma_r(\cdot) = \frac{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1, x/\|x\|_{\text{ang}} \in (\cdot)\}}{\Lambda(A_r)}.$$

The converse holds if there exist a family of risk functionals r_l ($l = 1, \dots, L$) with $L \geq 1$ such that for every l , $X \in \text{RV}(\mathcal{F}_+ \setminus \mathcal{C}_{r_l}, a_n, \Lambda)$ and $\bigcap_{l=1}^L \mathcal{C}_{r_l} = \{0\}$.

Theorem 1.6 is used in Section 1.4.2 to describe the asymptotic distribution of r -exceedances. For specific risk functionals such as $\sup_{s \in S} x(s)$ or $\int_S x(s) ds$ for which $\mathcal{C}_r = \{0\}$, the pseudo-polar transformation is a homeomorphism and the convergence of the factorized version of the measures is equivalent to the regular variation of the stochastic process X on $\mathcal{F}_+ \setminus \{0\}$. Alternatively, let $S_l \subset S$ ($l = 1, \dots, L$) satisfy $\bigcup_{l=1}^L S_l = S$, and define the functions

$$r_{S_l}(x) = \int_S \mathbf{1}_{\{s \in S_l\}} x(s) ds, \quad l = 1, \dots, L.$$

In this case, we see that $\{x \in \mathcal{F}_+ : r_{S_l}(x) = 0\} \neq \{0\}$ individually but the family r_{S_l} satisfies $\bigcap_{l=1}^L \mathcal{C}_{r_l} = \{0\}$, and then the conditions for equivalence in Theorem 1.6 are met, ensuring regular variation of the stochastic process X on $\mathcal{F}_+ \setminus \{0\}$.

Theorem 1.6 implies that there is a pseudo-polar decomposition of Λ for any valid risk functional, and for this reason we link pseudo-polar representations in Corollary 1.2.

Corollary 1.2 *For an angular norm $\|\cdot\|_{\text{ang}}$ and 1-homogeneous risk functionals r_1 and r_2 with $\mathcal{C}_{r_1} = \mathcal{C}_{r_2}$, the angular probability measures σ_{r_1} and σ_{r_2} are linked by*

$$\sigma_{r_1}(dw) = \frac{\Lambda(A_{r_2})}{\Lambda(A_{r_1})} \left\{ \frac{r_1(w)}{r_2(w)} \right\}^{1/\xi} \sigma_{r_2}(dw), \quad dw \in \mathcal{S}_r, \quad (1.23)$$

where $A_{r_i} = \{x \in \mathcal{F}_+ \setminus \mathcal{C}_{r_i} \mid r_i(x) \geq 1\}$, $i = 1, 2$.

Chapter 1. Asymptotic tail distributions: theory and models

Equation (2.3) can thus be used to obtain the probability measure σ_{r_2} when σ_{r_1} is known, as for certain combinations of norm and risk functional the mathematical expression for σ_r is much simpler.

Link with multivariate regular variation

All the previous definitions and results also hold for finite dimensions, i.e., for L -dimensional random vectors, by replacing \hat{w} -convergence by vague convergence (Resnick, 2007, Section 3.3.5) on $M_{\mathbb{R}_+^L \setminus \mathcal{C}^L}$, the class of Borel measures on $\mathcal{B}(\mathbb{R}_+^L \setminus \mathcal{C}^L)$ endowed with the $\|\cdot\|_\infty$ norm, where \mathcal{C}^L denotes a cone in \mathbb{R}_+^L (Opitz, 2013b).

The mapping theorem also makes it possible to describe the relation between the multivariate theory and functional regular variation. Let s_1, \dots, s_L be $L > 1$ locations in S , and consider the map $h_{\text{proj}} : \mathcal{F}_+ \setminus \mathcal{C} \rightarrow \mathbb{R}_+^L \setminus \mathcal{C}^L$ defined as

$$h_{\text{proj}}(x) = \{x(s_1), \dots, x(s_L)\},$$

where $\mathcal{C}^L = h_{\text{proj}}(\mathcal{C})$. We now use the mapping h_{proj} to prove that multivariate regular variation is embedded in the functional theory.

Corollary 1.3 *If $X \in \text{RV}(\mathcal{F}_+ \setminus \{0\}, a_n, \Lambda)$, then*

$$\Lambda_n \circ h_{\text{proj}}^{-1} \rightarrow \Lambda \circ h_{\text{proj}}^{-1}, \quad n \rightarrow \infty,$$

in $M\{\mathbb{R}_+^L \setminus \mathcal{C}^L\}$.

Corollary 1.3 shows how applications, which are by nature finite-dimensional, are linked to the functional regular variation model. Indeed, suppose that a physical process X , for instance rainfall or temperature, is produced by a continuous stochastic process over a region of interest. In practice, we observe the process at only a finite number of locations. Supposing functional regular variation for X implies, through Corollary 1.3, that the vector of sampled locations is also regularly varying with the same limiting measure as the full process. For parametric models, this means that the parameters estimated using station measurements equal those of the functional model.

However, multivariate regular variation does not in general imply functional regular variation and the condition for equivalence is still an open question. Following Lindskog et al. (2014, Theorem 4.1), Theorem 1.7 gives a necessary condition for equivalence when replacing \mathcal{F}_+ by \mathbb{R}_+^∞ .

1.4. Multivariate and functional limits of threshold exceedances

Theorem 1.7 *Suppose $X \in \text{RV}(\mathbb{R}_+^\infty \setminus \mathcal{C}, a_n, \Lambda)$ and for every $L \geq 1$, the closed set $\mathcal{C} \subset \mathcal{F}_+$ is such that $h_{\text{proj},L}(\mathcal{C})$ is closed in \mathbb{R}_+^L and*

$$\{x(s_1), \dots, x(s_L)\} \in h_{\text{proj},L}(\mathcal{C}) \Rightarrow \{x(s_1), \dots, x(s_L), 0_\infty\} \in \mathcal{C}, \quad (1.24)$$

where \Rightarrow means implication. Then $\Lambda_n \rightarrow \Lambda$ in $M(\mathbb{R}_+^\infty \setminus \mathcal{C})$ if and only if for all $L \geq 1$ such that $\mathbb{R}_+^L \setminus h_{\text{proj},L}(\mathcal{C}) \neq \emptyset$,

$$\Lambda_n \circ h_{\text{proj},L}^{-1} \rightarrow \Lambda \circ h_{\text{proj},L}^{-1}, \quad n \rightarrow \infty,$$

in $M\{\mathbb{R}_+^L \setminus \mathcal{C}^L\}$.

For the family of cones $\mathcal{C}_r^L = \{x \in \mathbb{R}_+^L : r(x) > 0\}$ defined in Section 1.4.1, condition (1.24) is ensured by the homogeneity of the risk functional, and thus Theorem 1.7 holds.

1.4.2 Limit distribution of functional r -exceedances

In this section, $r : \mathcal{F}_+ \rightarrow [0, +\infty)$ is a risk functional as defined in Section 1.4.1, which we consider, without loss of generality, to be 1-homogeneous, $\|\cdot\|_{\text{ang}}$ is a norm on \mathcal{F}_+ , and X is a regularly varying process on $\mathcal{F}_+ \setminus \{0\}$ with limiting measure Λ and tail index $\xi > 0$. In practice, the choice of the angular norm $\|\cdot\|_{\text{ang}}$ has no impact and is usually made for convenience, but choosing a risk functional r allows a focus on particular types of extreme event.

Risk functionals as a characterization of risk

In Dombry and Ribatet (2015), r is called a ‘cost functional’ and Opitz (2013b) named it a ‘radial aggregation function’, but we prefer ‘risk functional’ as it better reflects the fact r measures the severity of the risk under study.

Threshold exceedances were originally studied with the functional $r(x) = \sup_{s \in S} \{x(s)\}$ by Rootzén and Tajvidi (2006) in a multivariate setting and by Ferreira and de Haan (2014) for continuous processes. In this case, events for which there is a threshold exceedance at least one location are considered extreme. Alternatively, Coles and Tawn (1996) modelled areal rainfall based on the functional $\int_S X(s) ds$, a model that can be generalized with the family r_{S_i} ($i = 1, \dots, L$) defined in equation (1.4.1) to describe the exceedances of cumulative rainfall over $L > 1$ catchments of a river basin; see Chapter 3 where we describe the multivariate limit tail distribution of aggregated data. More generally, risk functionals such as $\int_S X^2(s) ds$ for wind energy inside a climatic system (Powell and Reinhold, 2007), $\min_{s \in S'} X(s)/u(s)$ for exceedances over

Chapter 1. Asymptotic tail distributions: theory and models

dams, $X(s_0)$ for risks impacting a specific location s_0 , and so forth, can be relevant, depending on the application.

r -Pareto processes as asymptotic distribution of r -exceedances

We now focus on in the limiting distribution of the r -exceedances of a regularly varying process, i.e., we wish to describe the limiting behaviour of

$$\Pr \left\{ X \in A \mid r \left(\frac{X}{u} \right) > 1 \right\}, \quad A \subset \mathcal{F}_+ \setminus \mathcal{C}_r,$$

as the threshold function u tends to infinity, i.e., $u(s) \rightarrow \infty$ for all $s \in S$.

Theorem 1.8 *Let X be a regularly varying stochastic process on $\mathcal{F}_+ \setminus \{0\}$ with limiting measure Λ and tail index ξ . Then*

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[r', \infty), \mathcal{W}\} \mid r \left(\frac{X}{a_n} \right) \geq 1 \right] = (r')^{-1/\xi} \sigma_r(\mathcal{W}), \quad r' \geq 1, \mathcal{W} \subset S_r,$$

where $S_r = \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : \|x\|_{\text{ang}} = 1\}$.

Theorem 1.8 motivates the definition of the family of r -Pareto processes, which characterizes the limiting distribution of r -exceedances.

Definition 1.2 *Let r be a 1-homogeneous risk functional. An r -Pareto process P with tail index $\xi > 0$ is a stochastic process on $\{x \in \mathcal{F}_+ : r(x) \geq 1\}$ such that*

$$\Pr \left[P \in (h_r^{\text{pp}})^{-1} \{[r', \infty), \mathcal{W}\} \right] = (r')^{-1/\xi} \times \sigma_r(\mathcal{W}), \quad r' \geq 1, \mathcal{W} \subset S_r, \quad (1.25)$$

where σ_r is a probability measure on S_r and h_r^{pp} denotes the pseudo-polar decomposition associated to the risk functional r .

Thus, as a direct consequence of Theorem 1.8, r -Pareto processes are defined to be the unique possible limit for r -exceedances of regularly varying stochastic processes. This means that for any $X \in \text{RV}(\mathcal{F}_+ \setminus \{0\}, a_n, \Lambda)$ and a sufficiently high threshold function $u > 0$, we can approximate the distribution of X by

$$\Pr \left[\frac{X}{u} \in (h_r^{\text{pp}})^{-1} \{[r', \infty), \mathcal{W}\} \mid r \left(\frac{X}{u} \right) > 1 \right] \approx (r')^{-1/\xi} \times \sigma_r(\mathcal{W}). \quad (1.26)$$

1.4. Multivariate and functional limits of threshold exceedances

Construction of the r -Pareto processes

Let $\xi > 0$ and σ_r be the tail index and the probability measure of an r -Pareto process P . The pseudo-polar decomposition defined in Section 1.4.1 gives the construction principle

$$P = R \frac{W}{r(W)}, \quad (1.27)$$

where R is a univariate Pareto variable with tail parameter $1/\xi$ and unit scale, and W is a stochastic process on \mathcal{S}_r with probability measure σ_r .

An important property of Pareto processes is peaks-over-threshold stability, i.e., for any real number $u > 0$

$$\Pr\{u^{-1}P \in \cdot \mid r(P) \geq u\} = \Pr\{P \in \cdot\}, \quad (1.28)$$

which means that the distribution of r -exceedances over a threshold $u > 0$ is stable with respect to rescaling. Peaks-over-threshold stability allows one to extrapolate the extremal behaviour of a regularly varying stochastic process X to intensities that may not have been observed yet. Equation (1.28) is a direct consequence of the homogeneity of order -1 satisfied by the limiting measure Λ .

Multivariate density function

In practice, σ_r is rarely available but finite-dimensional versions of Λ are fairly common. For this reason, the Cartesian representation of the multivariate density function of the r -Pareto process

$$f^r(x) = \frac{\lambda(x)}{\Lambda\{A_r(1)\}}, \quad x \in A_r(1), \quad (1.29)$$

where $A_r(1) = \{x \in \mathcal{F}_+ : r(x) \geq 1\}$, and

$$\Lambda\{A_r(1)\} = \int_{A_r(1)} \lambda(x) dx,$$

is often preferred. Using equation (1.29), with the help of a change of variables, closed forms for σ_r have been derived for special choices of risk functionals such as the L_1 -norm (Coles and Tawn, 1991).

Marginal properties

Consider a location $s_0 \in S$ and a sufficiently high threshold $u_0 > 0$ such that $\{x \in \mathcal{F}_+ : x(s_0) > u_0\} \subset \{x \in \mathcal{F}_+ : r(x) \geq 1\}$. Using the Cartesian representation of the spectral measure in Theorem 1.6, we obtain

$$\Pr\{P(s_0) > r'\} = \left(\frac{r'}{u_0}\right)^{-1/\xi} \frac{\Lambda\{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1, x(s_0) \geq u_0\}}{\Lambda\{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1\}}, \quad r' \geq u_0, \quad (1.30)$$

which means that above the threshold u_0 , the r -Pareto process has Pareto margins with tail index ξ . The proof of (1.30) can be found in Appendix A.

From a practical point of view, equation (1.30) has two consequences: as supposed at the beginning of this section, the dataset of observations should lie in the Fréchet domain of attraction and have a common tail index. For environmental applications, having a common tail index is a reasonable assumption if there is no mixture in the physical process studied, but data with finite upper bound (Weibull) or exponential tail decay (Gumbel) are commonly encountered. If one or both of these properties is not satisfied, a two-step procedure, as described in Section 1.3.2 for block maxima, is used: the data X are transformed to X^* , whose margins are standardized to unit Fréchet or unit Pareto. However, the risk functional r is now applied to X^* , so the r -exceedances are defined on the Fréchet scale and thus any physical interpretation of the risk is compromised. For example, exceedances of spatially accumulated rainfall X usually do not correspond in practice to exceedances of spatial accumulation applied to the transformed process X^* . In Chapter 4 we discuss how and under which conditions we can modify the r -Pareto process to have generalized Pareto margins in order to cover the three possible regimes of tail decay and to keep the definition of the risk for the original process.

Statistical inference

Let $X_n \in \mathbb{R}_+^L$ ($n = 1, \dots, N$) be realizations of a regularly varying stochastic process sampled at locations $s_1, \dots, s_L \in S$. To fit an r -Pareto process using the sample X_n ($n = 1, \dots, N$), we propose a likelihood inference procedure based on equation (1.26): the distribution of r -exceedances of X above a sufficiently high threshold vector $u > 0$ is approximated by the density function f^r in (1.29). In practice, this means maximizing the log-likelihood

$$\mathcal{L}_{\text{Thres}}(\theta) = \sum_{m \in K_u} \log f_\theta^r \left(\frac{x^m}{u} \right),$$

1.4. Multivariate and functional limits of threshold exceedances

where K_u is the index set of the r -exceedances, i.e., $K_u = \{m \in (1, \dots, n) : r(x_m/u) \geq 1\}$, and $\theta \in \Theta$ denotes the parameter vector of a parametric limit measure Λ_θ . For such an inference procedure, two elements need to be specified: a risk functional and a threshold vector. Under the assumption that $X \in \text{RV}(\mathcal{F}_+ \setminus \{0\}, a_n, \Lambda)$, Theorem 1.6 suggests that these choices should not affect model estimates, but this is not entirely true, because the events selected depend on the risk functional r , the choice of which enables the detection of mixtures in the extremes and can improve sub-asymptotic behaviour by fitting the model using only those observations closest to the chosen type of extreme event. For example, we might expect the extremal dependence of intense local rainfall events to differ from that of heavy large-scale precipitation, even in the same geographical region.

Choosing a threshold vector is not trivial, because the components of u cannot be taken arbitrarily high, and there is no unified methodology for this task. Tools for univariate threshold selection have been developed (e.g. Hill, 1975; Davison and Smith, 1990; Northrop and Coleman, 2014) and are mainly based on the detection of a stability region in some graphical diagnostic; see Scarrott and MacDonald (2012) for an extensive review of existing procedures. For instance, components of u can be chosen as local empirical quantiles whose levels belongs to a region of $[0, 1]$ where the estimated shape parameter of a generalized Pareto distribution is stable. For multivariate threshold selection, the literature is unfortunately fairly restricted: Wadsworth (2016) presents a methodology based on the independent-increments structure of maximum likelihood estimators, while Wan and Davis (2018) look for a stability region within the regular variation framework. But so far, univariate methods applied locally have dominated applications in the environmental sciences.

An alternative to the log-likelihood function $\mathcal{L}_{\text{Thres}}$ assumes that the number of excesses is Poisson distributed, and has the benefit of reducing the sensitivity of the estimator to the choice of the threshold u . In this case, starting from equation (1.18), the approximation

$$\Pr \left\{ \frac{X(s_1)}{a_n(s_1)} > u_1, \dots, \frac{X(s_L)}{a_n(s_L)} > u_L \right\} \approx \frac{1}{n} \Lambda_\theta(u_1, \dots, u_L),$$

is used to estimate the scaling function a_n jointly with the r -Pareto process parameters θ , which yields the log-likelihood

$$\mathcal{L}_{\text{PP}}(\theta, a_n) = -\Lambda_\theta \left\{ \frac{u_1}{a_n(s_1)}, \dots, \frac{u_L}{a_n(s_L)} \right\} + \sum_{m \in K_u} \lambda_\theta \left(\frac{x^m}{a_n} \right). \quad (1.31)$$

In this case, the threshold vector u does not appear in the density function λ_θ but only in the distribution of excesses, and thus the impact of the choice of the threshold on

parameter estimates is limited. On the other hand, a parametric model needs to be specified for a_n ; see Section 3.5.2 for an example. Practical experience showed that the threshold approach is convenient to obtain preliminary point estimates of the parameters that can be used as starting values to initialize the maximization of the Poisson process likelihood.

In Chapter 4, further models for the distribution of excesses are considered and linked together. For a thorough discussion on likelihood based methods for statistical inference of extremal processes, we refer the reader to Chapter 2.

1.5 Characterization of bivariate tail dependence

In spatial statistics, the correlation between two locations of a stochastic process is often used as a simple summary to study the dependence structure. However, for an extremal process a covariance may not exist. Hence, alternative measures of bivariate dependence have been introduced. Such measures do not fully specify the dependence structure of the process, but remain very useful for data exploration and model validation. They allow a user to distinguish between the two regimes of tail dependence, namely asymptotic dependence and asymptotic independence.

1.5.1 Asymptotic independence

Let X be a stochastic process over S with common marginal distribution function F for all $s \in S$ and let $x^* = \sup\{x \in \mathbb{R}_+ : F(x) < 1\}$ be the upper bound of the support of X . Asymptotic independence arises when for any locations $s_1, s_2 \in S$,

$$\Pr\{X(s_2) \geq x | X(s_1) \geq x\} \rightarrow 0, \quad x \rightarrow x^*. \quad (1.32)$$

While equation (1.32) defines the regime of asymptotic independence, it does not measure the speed of convergence toward this limit. To this end, Ledford and Tawn (1996) introduce the coefficient of tail dependence: if a stochastic process X satisfying (1.32) has been standardized to X^* with unit Fréchet marginals, then for any $s_1, s_2 \in S$ there exist $0 < \eta_{s_1, s_2} \leq 1$ and a slowly-varying function L_{s_1, s_2} , such that

$$\Pr\{X^*(s_2) \geq x, X^*(s_1) \geq x\} = \Pr[\min\{X^*(s_2), X^*(s_1)\} \geq x] \approx L_{s_1, s_2}(x)x^{-1/\eta_{s_1, s_2}}; \quad (1.33)$$

see Section 1.2.1 for a definition of slowly-varying functions. The parameter η_{s_1, s_2} , called the coefficient of tail dependence and now denoted η for simplicity, measures the speed of convergence toward the asymptotic independence regime: the rate

1.5. Characterization of bivariate tail dependence

increases as $\eta \rightarrow 0$. The limit case $\eta = 1$ corresponds to asymptotic dependence, which will be discussed in Section 1.5.2. In practice, η can be estimated by computing the univariate summary $Y = \min\{X^*(s_2), X^*(s_1)\}$ of the standardized observations and then fitting a generalized Pareto distribution to Y^* . Equation (1.33) is easily linked to (1.32), as for sufficiently large $x > 0$,

$$\Pr\{X^*(s_2) \geq y \mid X^*(s_1) \geq x\} = \frac{\Pr\{X^*(s_2) \geq x, X^*(s_1) \geq x\}}{\Pr\{X^*(s_1) \geq x\}} \approx L(x)x^{1-1/\eta};$$

it is easy to check that for $0 < \eta < 1$, the probability tends to 0 as x tends to infinity. Finally, the tail coefficient also allows us to distinguish three asymptotic independence regimes:

- for $1/2 < \eta < 1$, $X(s_1)$ and $X(s_2)$ are positively associated. This is the case for a Gaussian process with $\text{cov}\{X(s_1), X(s_1)\} > 0$, for instance;
- for $\eta = 1/2$, $X(s_1)$ and $X(s_2)$ are exactly independent in the limit, while typically having weak dependence at sub-asymptotic levels;
- for $0 < \eta < 1/2$, $X(s_1)$ and $X(s_2)$ are negatively associated, as is the case for a Gaussian process with $\text{cov}\{X(s_1), X(s_1)\} < 0$.

The coefficient $\bar{\chi}$ introduced by Coles et al. (1999) is an alternative measure of the speed of convergence toward asymptotic independence : for a bivariate random vector $X = (X_1, X_2)$ with distribution function F_1 and F_2 , define the function

$$\bar{\chi}(u) = \frac{2 \log \Pr\{F_1(X_1) > u\}}{\log \Pr\{F_1(X_1) > u, F_2(X_2) > u\}} - 1, \quad 0 < u < 1. \quad (1.34)$$

The coefficient $\bar{\chi}$ is then defined as the limit of $\bar{\chi}(u)$ as $u \rightarrow 1$. The function $\bar{\chi}(\cdot)$ and its limit $\bar{\chi}$ take values in $[-1, 1]$, but $\bar{\chi}(\cdot)$ satisfies the stricter lower bound $2 \log(1 - u) / \log\{\max(1 - 2u, 0)\} - 1 \leq \bar{\chi}$ (Beirlant et al., 2004, p. 344). Also, $\bar{\chi}$ can be linked with the coefficient of tail dependence η through the relation

$$\bar{\chi} = 2\eta - 1.$$

Thus the three regimes described above for η are equivalent to $\bar{\chi} > 0$, $\bar{\chi} = 0$ and $\bar{\chi} < 0$ with asymptotic dependence appearing as the limit case $\bar{\chi} = 1$. The coefficient $\bar{\chi}$, while more difficult to interpret from its definition than η , is simply equal to the correlation $\rho \in [-1, 1]$ when X has a bivariate Gaussian distribution, which makes it sometimes more attractive. Finally, $\bar{\chi}$ can be estimated using a graphical diagnostic where estimates of $\hat{\chi}(u)$, obtained by replacing the quantities in (1.34) by empirical

Chapter 1. Asymptotic tail distributions: theory and models

estimates of the probabilities and distribution functions, are plotted for different values of $u \in [0, 1]$.

As mentioned above, the Gaussian process is a good example of asymptotic independence, which implies that with an increasing marginal intensity, the strength of the dependence decreases. In other words, the spatial extent of extreme events generated by an asymptotically independent process shrinks with the growth of the return period until it degenerates to a point in the limit model.

1.5.2 Asymptotic dependence

When the limit in equation (1.32) is strictly positive, the process X is asymptotically dependent and we write

$$\Pr \{X(s_2) \geq x | X(s_1) \geq x\} \rightarrow \pi(s_1, s_2), \quad x \rightarrow x^*. \quad (1.35)$$

The quantity $\pi(s_1, s_2)$ is called the coefficient of extremal dependence (Beirlant et al., 2004, Section 9.5.1), and for obvious reasons takes values in $[0, 1]$. If $\pi(s_1, s_2) = 1$, the process is exactly dependent, i.e. $X(s_1) = X(s_2)$, while for $\pi(s_1, s_2) = 0$, we retrieve asymptotic independence. If we suppose that X is an r -Pareto process, as described in Section 1.4, then there exists $u_{\min} > 0$ such that

$$\pi(s_1, s_2) = \Pr \{X(s_2) \geq u | X(s_1) \geq u\}, \quad u \geq u_{\min}.$$

The coefficient of extremal dependence was introduced by Ledford and Tawn (1996) through the coefficient χ , defined for any stochastic process X as

$$\chi(s_1, s_2) = \lim_{u \rightarrow 1} \Pr [F_1\{X(s_2)\} \geq u | F_1\{X(s_1)\} \geq u],$$

where F_1 and F_2 refer to the marginal distribution functions at locations s_1 and s_2 respectively; thus for a process with common distribution function F over S , $\pi = \chi$.

When the process X is stationary over S , the coefficient of extremal dependence can be viewed as a function of the distance $h = s_2 - s_1$; see Section 1.6.1 for a formal definition of stationarity. In this case, $\pi(h)$ is called the extremogram (Davis et al., 2013b) and a simple natural estimator of it is obtained by replacing the probabilities in (1.35) by empirical estimators, i.e.,

$$\hat{\pi}(h) = \frac{\sum_{t=1}^n 1\{X(s_2) \geq u, X(s_1) \geq u\}}{\sum_{t=1}^n 1\{X(s_1) \geq u\}}, \quad (1.36)$$

1.5. Characterization of bivariate tail dependence

where $1\{\cdot\}$ is the indicator function. In Davis and Mikosch (2009), the extremogram was presented for stationary time series; it was generalized to spatio-temporal processes by Buhl and Klüppelberg (2018).

Let X now refer to a stochastic process in the max-domain of attraction of a simple max-stable process Z with limiting measure Λ . Another widely used measure of extremal dependence for X sampled at locations $s_1, s_2 \in S$ is the extremal coefficient

$$\theta(s_1, s_2) = \Lambda [x \in C_+(s) : \max\{x(s_1), x(s_2)\} \geq 1] = 2 - \pi(s_1, s_2).$$

The strength of the extremal coefficient is its interpretability. Indeed the pairwise distribution function of Z at s_1 and s_2 can be written as

$$\Pr\{Z(s_1) \leq x, Z(s_2) \leq x\} = \exp \left\{ - \left(\frac{1}{x} \right)^{\theta(s_1, s_2)} \right\}, \quad x > 0; \quad (1.37)$$

for $\theta(s_1, s_2) = 1$, we retrieve the distribution function of a perfectly dependent bivariate vector with Fréchet marginals, and for $\theta(s_1, s_2) = 2$ the distribution function equals the product of two unit Fréchet distribution functions, corresponding to exact independence. An alternative definition of the extremal coefficient is

$$\theta(s_1, s_2) = 2 \lim_{u \rightarrow \infty} u^{-1} \Pr [\max\{X(s_1), X(s_2)\} \geq u], \quad (1.38)$$

for which equality holds for finite values of $u > 0$ if X is a r -Pareto process with unit tail index. Similarly to the extremogram, if we further suppose that X is stationary, the extremal coefficient can be considered as a function of the distance h between s_1 and s_2 . For general X with common distribution function over S , a simple estimator similar to (1.36) can be derived from (1.38). Schlather and Tawn (2003) propose an alternative maximum likelihood estimator based on (1.37), yielding

$$\hat{\theta}(s_1, s_2) = \frac{\text{Card}(\{i : \max\{x^i(s_1), x^i(s_2)\} > u\})}{\sum_{i=1}^n \max\{u, x^i(s_1), x^i(s_2)\}}, \quad u > 0, \quad (1.39)$$

where X^i , ($i = 1, \dots, n$) are independent replicates of X . However, this estimator does not equal 1 when $n = 1$ and $u = 0$, and so must be corrected. The extremal coefficient can be generalized to more than two locations (Smith, 1990) and to different probabilities with the same principle as for (1.38); see Chapter 3 for more details.

Last but not least, the F -madogram (Cooley et al., 2006) is a third dependence measure

for stationary extremal processes, defined by

$$\phi_{\text{mado}}(h) = \frac{1}{2} \mathbb{E}[|F_1\{Z(s_1)\} - F_2\{Z(s_2)\}|]$$

where Z is a max-stable process with marginal distributions F_1 and F_2 at locations s_1 and s_2 respectively. The F -madogram is linked to the extremal coefficient, and by extension to the extremogram, by the relationship

$$\theta(h) = \frac{1 + 2\phi_{\text{mado}}(h)}{1 - 2\phi_{\text{mado}}(h)}.$$

This quantity, while less popular than its competitors, is closer in definition to the semi-variogram $\gamma(h) = 2^{-1} \mathbb{E}[\{X(s_2) - X(s_1)\}^2]$ which is the dependence measure of reference in classical spatial statistics. For extremal processes, $\gamma(h)$ cannot be used in general as the second moments of the increments do not usually exist for heavy-tailed processes, but we see in the Section 1.6.2 that it is possible to build a class of extremal measures whose dependence is driven by a semi-variogram.

1.6 Spatio-temporal extremal processes based on Gaussian random functions

Estimation of the risk related to extreme natural hazards is usually based on catalogues of historical events, used as ‘stress tests’ to assess the resistance of human infrastructure or insurance portfolios. Stochastic weather generators are a natural tool to enlarge or create catalogues with unobserved extreme events. The spatio-temporal nature of environmental phenomena requires specific models that we describe in this section.

1.6.1 Classical geostatistics

Existing stochastic weather generators were built within the frame of classical spatio-temporal statistics and so are based on Gaussian processes (Cressie, 1993; Wackernagel, 2003; Diggle and Ribeiro, 2007; Gelfand et al., 2010; Cressie and Wikle, 2011), for which a large class of dependence models have been developed, with corresponding software (Pebesma and Graeler, 2013; Renard et al., 2017). However, with its exponential tail decay, the Gaussian distribution is likely to badly underestimate the probability of rare events and thus is not recommended for the study of extremes.

For better marginal modelling, heavy-tailed data can be transformed to have normal

1.6. Spatio-temporal extremal processes based on Gaussian random functions

distributions, for instance with a marginal transform, before applying tools from spatial analysis. Such methodology is referred to as copula modelling in statistics (Nelsen, 2006), and in geostatistics an approach, called Gaussian anamorphosis, uses a semi-parametric marginal transform based on Bernstein polynomials (Lajaunie, 1993). This procedure is attractive because it takes advantage of existing models developed for Gaussian processes, but Mikosch (2006) argues that in general the copula approach might have some undesirable effects; see McNeil et al. (2015, p. 210) for a discussion of the Gaussian copula in extreme value theory. Indeed for a threshold $u > 0$ and a bivariate vector X with Fréchet margins and dependence driven by Gaussian copula with correlation $-1 < \rho < 1$, we have (Ledford and Tawn, 1996)

$$\Pr(X_1 > u \mid X_2 > u) \sim C \times u^{-(1-\rho)/(\rho+1)} (\log u)^{-\rho/(1+\rho)}, \quad (1.40)$$

where $C > 0$. In equation (1.40), we notice that

$$\lim_{u \rightarrow \infty} \Pr(X_1 > u \mid X_2 > u) = 0.$$

Thus when the vector X represents a physical process in space, the spatial coverage of extreme events decreases as the intensity u increases. Such behaviour can be attractive in applications, for instance when trying to model rain storms, which tend to be more localized with increasing intensity, as observed for ‘Cévenol’ rain in the South of France. On the contrary, for events such as heatwaves, windstorms or cyclones, we do not expect the the spatial extent to shrink with severity, so asymptotically independent processes are not suited for their statistical analysis and alternative models are required.

As explained in Klüppelberg and Resnick (2008), heavy-tailed copulas do not suffer from this shortcoming and models derived from Gaussian processes, such as the Brown–Resnick model (Brown and Resnick, 1977), allow classical dependence functions to be used. For this reason, in this section we present the concepts and results of classical spatio-temporal statistics needed to build this class of extreme value copulas.

Basic notions on stochastic processes

Let S be a subset of \mathbb{R}^D , $D = 1, 2, \dots$, such as $[0, 1]^2$ for spatial applications or $[0, 1]^2 \times [0, \infty)$ for spatio-temporal extensions. A set of random variables $\{X(s), s \in S\}$ indexed by the parameter s and defined on a common probability space $(C(S), \mathcal{B}\{C(S)\}, \Lambda)$ is called a *random process* or a *random function* and its realizations x are named *regionalized variables* (Wackernagel, 2003, p. 42). When $D = 2$, X is also known as a random field; it is called a time series for $S = \mathbb{R}_+$. In applications, regionalized

Chapter 1. Asymptotic tail distributions: theory and models

variables are never observed throughout the whole of S but only at a finite sample s_1, \dots, s_L , of locations in a spatial setting and times for time series.

A random process X is characterized by knowledge of the distribution functions

$$F_{s_1, \dots, s_L}(x_1, \dots, x_L) = \Pr\{X(s_1) \leq x_1, \dots, X(s_L) \leq x_L\},$$

of the vector $(X(s_1), \dots, X(s_L))$ for any set $s_1, \dots, s_L \in S$ of any size $L > 0$. Consequently two random processes X^1 and X^2 are identically distributed if their distribution functions F_{s_1, \dots, s_L}^1 and F_{s_1, \dots, s_L}^2 are equal for any set $\{s_1, \dots, s_L\}$ of any size L . An important class of random functions is that of Gaussian processes, for which the distribution of the vector $(X(s_1), \dots, X(s_L))$ is multivariate Gaussian for any set $\{s_1, \dots, s_L\}$ and every L .

Stationarity is an important property which greatly simplifies models. While the corresponding intuition is associated to the notion of invariance, its mathematical formalization is more delicate and multiple definitions have been given. The first, called *strict stationarity*, requires that for any set $\{s_1, \dots, s_L\}$ and any vector $h \in \mathbb{R}^D$, called a *lag vector*, satisfying $\{s_1 + h, \dots, s_L + h\} \in S$, we have

$$F_{s_1, \dots, s_L}(x_1, \dots, x_L) = F_{s_1 + h, \dots, s_L + h}(x_1, \dots, x_L),$$

i.e., the distribution function F_{s_1, \dots, s_L} is invariant under translation of $\{s_1, \dots, s_L\}$ by a vector h .

More generally a weaker definition of stationarity, called *second-order stationarity*, requires first that the random process X has a constant mean $m(s) = m$ and second that covariance function $\text{cov}\{X(s), X(s + h)\}$ is finite for all $s \in S$ and depends only of the lag vector $h \in \mathbb{R}^D$ (Wackernagel, 2003, p.44). As a Gaussian process is entirely specified by its first two moments, strict stationarity is equivalent to the first-order stationarity for such processes.

Finally, a third type of stationarity, called *intrinsic stationarity*, requires the field of increments $X(s) - X(s_{\text{ref}})$ to be stationary of second-order for all $s_{\text{ref}} \in S$ (Wackernagel, 2003, p. 44). For a Gaussian process, second-order stationarity implies intrinsic stationarity, but the converse is not true in general. Indeed, in case of intrinsic stationarity, the variance $\text{var}\{X(s)\}$ may not exist, whereas its existence is a necessary condition for strict stationarity. In classical spatial statistics, intrinsic stationarity is rarely considered alone, and was introduced mostly for historical reasons, but for extremal processes, it is key to constructing flexible models; see Section 1.6.2.

Another important property of stationary random functions is isotropy: if the covariance function $C(h) = \text{cov}\{X(s), X(s + h)\}$ depends only of the norm $\|h\|_2$ of the lag

1.6. Spatio-temporal extremal processes based on Gaussian random functions

vector, and so not on its direction, then the process X is called *isotropic*. In practice, isotropy is rarely satisfied for environmental studies, for which it is usual to find a direction of stronger dependence, for instance in the presence of prevailing currents caused by wind or water flows. Thus isotropic models are usually unrealistic but constitute a basis for building complex anisotropic dependence functions.

Covariance function and semi-variogram

Once the mean $m \in \mathbb{R}$ is known, a stationary Gaussian process is characterized by its covariance function

$$C(h) = \text{cov}\{X(s), X(s+h)\} = E\{X(s)X(s+h)\} - m^2.$$

For simplicity of exposition we now suppose without loss of generality that $m = 0$; otherwise just replace X by $X - m$. The covariance function $C(h)$ of X is bounded by $C(0) = \text{var}\{X(s)\}$ and is nonnegative definite, i.e., for any L and any set $\{s_1, \dots, s_L\}$,

$$\sum_{i=0}^L \sum_{j=0}^L c_i c_j C(s_i - s_j) \geq 0, \quad c_1, \dots, c_L \in \mathbb{R}. \quad (1.41)$$

Positive definite functions in (1.41) are specific types of kernel functions, and thus are well-understood on \mathbb{R} ; see for instance Berg et al. (1984, Chapter 4) for a review of classical kernels. However, kernels on \mathbb{R}^D with $D > 1$ or on the sphere $\{x \in \mathbb{R}^D : \|x\|_2 = 1\}$ constitute a very active field of research. Let \mathcal{D}_D be the class of all D -dimensional correlation functions on \mathbb{R}^D , and let \mathcal{D}_∞ refer to the class of isotropic correlation functions that belong to \mathcal{D}_D for any $D > 0$. Then

$$\mathcal{D}_1 \supset \mathcal{D}_2 \supset \dots \supset \mathcal{D}_L \supset \dots \supset \mathcal{D}_\infty.$$

The set \mathcal{D}_1 coincides with the set of all real positive definite functions but may not yield a valid covariance function for $L > 1$; see Gelfand et al. (2010, p. 62) for more details. In general, the class of valid covariance functions on \mathbb{R}^D is described by Bochner's theorem.

Theorem 1.9 (Bochner (1955)) *A function $C(h) : \mathbb{R}^D \rightarrow \mathbb{R}_+$ is a covariance function if and only if it is the Fourier inverse-transform of a positive bounded symmetric measure $F(dw)$, i.e.,*

$$C(h) = \int \exp(2\pi i \langle w, h \rangle) F(dw), \quad (1.42)$$

where $\langle \cdot, \cdot \rangle$ is a scalar product and $\int F(dw) < \infty$.

Chapter 1. Asymptotic tail distributions: theory and models

As mentioned earlier in this section, for second-order stationary processes, the covariance function may not exist, so an alternative measure of dependence is required, such as the (semi-)variogram

$$\gamma(s_1 - s_2) = \frac{1}{2} \text{var}\{X(s_2) - X(s_1)\}, \quad s_1, s_2 \in S.$$

The existence of γ is ensured by the strict stationarity of the increments, and for isotropic random processes, it simplifies to $\gamma(h) = \gamma(\|h\|_2)$ and satisfies

$$\gamma(h) = \gamma(-h), \quad \gamma(h) \geq 0, \quad \gamma(0) = 0, \quad h \in [0, \infty).$$

Similarly to a covariance function, the semi-variogram satisfies mathematical properties such as conditional negative definiteness, i.e.,

$$-\sum_{i=0}^L \sum_{j=0}^L c_i c_j \gamma(s_i - s_j) \geq 0, \quad \sum_{i=0}^L c_i = 0. \quad (1.43)$$

Berg et al. (1984, Section 3.2) gives general links between positive and negative definite kernels, and the set of conditionally negative definite functions is included in the set of negative definite functions. A straightforward consequence of equation (1.43) is that for two valid variograms γ_1 and γ_2 on \mathbb{R}^D , the function $\gamma_1 + \gamma_2$ is also a valid variogram: this simple fact is used to combine variograms to build models known as ‘gigognes’ with greater flexibility. Also, Berg et al. (1984, Theorem 2.2) shows that for any semi-variogram γ on \mathbb{R}^D , the function $\exp\{-t\gamma(h)\}$ is a valid covariance function. Similarly to Bochner’s theorem, a valid semi-variogram admits a representation in the Fourier domain, i.e., there exists a quadratic form $Q(h)$ and a positive bounded symmetric measure $F(dw)$ satisfying

$$\int \frac{F(dw)}{1 + 4\pi^2 \|w\|^2} < \infty$$

such that

$$\gamma(h) = \int \frac{1 - \cos(2\pi \langle w, h \rangle)}{4\pi^2 \|w\|^2} F(dw) + Q(h). \quad (1.44)$$

Equation (1.44) suggests that for any semi-variogram γ , there is a positive scalar $A > 0$ such that

$$\gamma(h) \leq A \|h\|^2, \quad h \in \mathbb{R}^D,$$

with the inequality being strict if the corresponding intrinsic random process is not differentiable. Also, if there exists a constant $C_{\text{sill}} \in \mathbb{R}$ such that

$$\gamma(h) \leq C_{\text{sill}}, \quad h \in \mathbb{R}^D,$$

1.6. Spatio-temporal extremal processes based on Gaussian random functions

then the corresponding intrinsic random process is also strictly stationary, meaning that we can link its covariance function C and semi-variogram γ with the relation

$$\gamma(h) = C(0) - C(h), \quad h \in \mathbb{R}^D.$$

On the contrary, if γ is unbounded, then a covariance function does not exist because $\text{var}\{X(s)\} = \lim_{h \rightarrow \infty} \gamma(h)$ is infinite. In this case, if the process is observed at a location s_0 , we can write the covariance between two locations $s_1, s_2 \in S$ conditioned on the observation $X(s_0) = x_0 \in \mathbb{R}$, i.e.,

$$\text{cov}\{X(s_1) - X(s_0), X(s_2) - X(s_0) \mid X(s_0) = x_0\} = \gamma(s_1 - s_0) + \gamma(s_2 - s_0) - \gamma(s_2 - s_1).$$

Finally, if there exists $\alpha \in (0, 2]$ such that

$$\gamma(h) \sim \|h\|_2^\alpha, \quad \|h\|_2 \rightarrow 0,$$

then the corresponding Gaussian process has a fractal or Hausdorff dimension $D + 1 - \alpha/2$ (Gelfand et al., 2010, p. 23). The parameter α drives the smoothness of the process, with larger values of α giving smoother realizations. For $D = 1$, the process is differentiable only if $\alpha = 2$; see Figure 2.3 in Gelfand et al. (2010, p. 24) for illustrations.

In this section, we have seen that a semi-variogram function must satisfy a very specific set of mathematical properties, so it is not straightforward to build dependence models. Luckily, equations (1.42) and (1.44) give convenient representations that motivate a large class of parametric models.

Classical dependence models in \mathbb{R}^D

Table 1.1 reviews most of the classical semi-variogram models, and, when relevant, the corresponding covariance functions. The nugget variogram is the simplest dependence function; it models the discontinuity of the variogram at the origin and so is equivalent to white noise. Although we expect the variogram of a continuous process to be continuous at the origin, the nugget accounts for small variations such as support variation or measurement errors.

The most popular model in spatial statistics is the Matèrn semi-variogram (Matèrn, 1960), which is a valid model on \mathbb{R}^D for any $D > 0$, and whose smoothness is driven by a parameter ν : the associated Gaussian process is m times differentiable if $m < \nu$. The Matèrn covariance is widely used because it unifies a wide range of classical models. For instance for $\nu = 1/2$, we obtain the exponential covariance function $C_{\text{exp}}(h) = \sigma^2 \exp(-\|h\|/\tau)$. The nugget effect arises as a limiting case when $\nu \rightarrow 0$, and

Chapter 1. Asymptotic tail distributions: theory and models

for a Gaussian covariance, we set $\tau = (2\nu^{1/2})^{-1}$ and then let $\nu \rightarrow \infty$. Whittle (1954, 1963) proved that a Gaussian process with Matern covariance function arises as a solution of the stochastic partial differential equation

$$(\tau^{-1} - \Delta)^{(2\nu+D)/4} x(s) = W(s), \quad x \in \mathbb{R}^D, \quad (1.45)$$

where W is a spatial Gaussian white noise with unit variance and Δ is the Laplacian operator

$$\Delta = \sum_{i=1}^D \frac{\partial^2}{\partial^2 x_i}.$$

While theoretical, these results have led to recent development of high-dimensional applications in spatio-temporal statistics. Indeed, classical techniques for statistical inference are limited in size by the computational complexity required to invert the dense covariance matrix $\Sigma = [\text{cov}\{X(s_i), X(s_j)\}]_{i,j=1,\dots,L}$. To work around this limit, Lindgren et al. (2011) propose to directly model the precision matrix Σ^{-1} , which can be well approximated by solving equation (1.45) with finite-element methods. Also Σ^{-1} can be constrained to be sparse, and thus large and dense datasets sampled at hundreds of thousands of locations can be handled. The framework also offers ground to build new flexible models with varying dependence over space and time (Fuglstad et al., 2015a,b). For inference, Lindgren et al. (2011) advocate a Bayesian framework based on Laplace approximation, which they implement in the R-INLA package (Rue et al., 2017).

With a Matérn dependence function, the corresponding semi-variogram is bounded and thus it can only model random processes that are stationary of second order. Schlather and Moreva (2017) introduced a model based on Bernstein polynomials (Schilling et al., 2012), which bridges bounded and unbounded semi-variograms,

$$\gamma_B(h) = \frac{(1 + \|h/\tau\|^\alpha)^{\beta/\alpha} - 1}{2^{\beta/\alpha} - 1}, \quad h \in \mathbb{R}^D, \quad \tau > 0, \alpha \in (0, 2], \beta \leq 2.$$

For this model, the parameter α determines the smoothness of the associated Gaussian process, while β indicates long-range behaviour. Indeed, for $\beta < 0$, the variogram has upper bound $(1 - 2^{\beta/\alpha})^{-1}$ and for $\beta > 0$, it is unbounded. Roughly speaking, γ_B behaves like $\|h\|^\alpha$ at the origin and like $\|h\|^\beta$ as $\|h\| \rightarrow \infty$. This unifies many existing models: the power variogram $\|h\|^\alpha$ when $\alpha = \beta$, corresponding to a fractional Brownian random field, the generalized Cauchy variogram (Gneiting and Schlather, 2004) for $\beta < 0$, the logarithmic model (Schilling et al., 2012, p. 90) as $\beta \rightarrow 0$, and the nugget model as $\alpha \rightarrow 0$ and $\beta \leq 0$. Quoting Gelfand et al. (2010, p. 25), ‘the long-memory behaviour is usually not relevant for interpolation purposes but can be critical for inference’; this is even more true in the context of extremal processes, as we will explain

1.6. Spatio-temporal extremal processes based on Gaussian random functions

in Section 1.6.2. For its flexibility, the Bernstein model is strongly recommended for spatial analysis of extreme events.

So far, we have described isotropic dependence models only, but isotropy is rarely a reasonable assumption in spatial statistics. All the aforementioned models can be made anisotropic with the help of a deformation matrix Ω , using

$$\gamma_{\text{anisotropic}}(h) = \gamma_{\text{isotropic}}(\Omega^{1/2}h),$$

where for instance in \mathbb{R}^2 ,

$$\Omega = \begin{bmatrix} \cos \eta & \sin \eta \\ -a \sin \eta & a \cos \eta \end{bmatrix},$$

where $a > 0$ and $\eta \in (-\pi/4, \pi/4]$ and $\Omega^{1/2}$ is such that $\Omega = \Omega^{1/2}\Omega^{1/2}$. This type of anisotropy is called geometrical because it relies on a deformation of the space based on a dilation of scale a combined with a rotation of angle η . A second type of model for anisotropy, called zonal anisotropy or stratified anisotropy (Wackernagel, 2003, p. 65), was developed for applications where the empirical covariance function calculated in different directions has different upper bounds. In this case, a variogram model is fitted in one direction only, and then a geometrical anisotropic model is fitted to the overall data, yielding for instance in \mathbb{R}^2 ,

$$\gamma_{\text{zonal}}(h) = \gamma_1(h_1) + \gamma_2(\Omega^{1/2}h), \quad h = (h_1, h_2) \in \mathbb{R}^2.$$

for which the corresponding Gaussian process X is composed of the sum of two independent processes X_1 and X_2 with variograms γ_1 and γ_2 respectively.

Spatio-temporal models

We have reviewed classical isotropic dependence models along with some specific cases of anisotropy. For spatio-temporal processes, the behaviour of the temporal component is usually very different from the spatial components, and this characteristic cannot be handled by the models presented so far. Thus, specific models for space-time dependence functions have been developed.

In this section, we suppose that S is a subset of $\mathbb{R}^2 \times [0, \infty)$, and write $(s, t) \in S \times \mathbb{R}$, where s and t denote the spatial and the temporal coordinates respectively. The notation $\{X(s, t) : (s, t) \in S\}$ thus refers to a spatio-temporal random process. A covariance function C , respectively a variogram function γ , on $\mathbb{R}^2 \times \mathbb{R}$ is *separable* if there exist

Table 1.1 – Classical parametric models for isotropic dependence in \mathbb{R}^D . For all models, $\tau > 0$ is the scale parameter, and $\alpha \in (0, 2]$ and $\nu > 0$ are smoothness parameters. For bounded variograms, $\sigma^2 > 0$ is the sill. The function K_ν is a modified Bessel function of second kind and order $\nu > 0$. For the Bernstein model $\beta \leq 2$ is a long-memory parameter that determines the behaviour of the variogram as $\|h\| \rightarrow \infty$.

Model	Semi-variogram	Covariance function
Nugget	$\mathbf{1}(\ h\ > 0)\sigma^2$	$\mathbf{1}(\ h\ = 0)\sigma^2$
Spherical	$\mathbf{1}(\ h\ \leq \tau) \left\{ 1 - \sigma^2 \left(1 - \frac{3\ h\ }{2\tau} + \frac{\ h\ ^3}{2\tau^3} \right) \right\}$	$\mathbf{1}(\ h\ \leq \tau) \sigma^2 \left(1 - \frac{3\ h\ }{2\tau} + \frac{\ h\ ^3}{2\tau^3} \right)$
Power exponential	$\sigma^2 \{ 1 - \exp(-\ h/\tau\ ^\alpha) \}$	$\sigma^2 \exp(-\ h/\tau\ ^\alpha)$
Matérn	$\sigma^2 \left\{ 1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\ h\ }{\tau} \right)^\nu K_\nu \left(\frac{\ h\ }{\tau} \right) \right\}$	$\frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\ h\ }{\tau} \right)^\nu K_\nu \left(\frac{\ h\ }{\tau} \right)$
Power	$\left\ \frac{h}{\tau} \right\ ^\alpha$	∞
Bernstein	$\frac{(1 + \ h/\tau\ ^\alpha) \beta / \alpha - 1}{2\beta / \alpha - 1}$	$\begin{cases} \infty, & \beta \geq 0 \\ \frac{(1 + \ h/\tau\ ^\alpha) \beta / \alpha}{1 - 2\beta / \alpha}, & \beta < 0 \end{cases}$

1.6. Spatio-temporal extremal processes based on Gaussian random functions

covariance functions C_{space} and C_{time} , variogram γ_{space} and γ_{time} such that

$$\begin{aligned} C\{(s_1, t_1), (s_2, t_2)\} &= C_{\text{space}}(s_1, s_2)C_{\text{time}}(t_1, t_2), \\ \gamma\{(s_1, t_1), (s_2, t_2)\} &= \gamma_{\text{space}}(s_1, s_2) + \gamma_{\text{time}}(t_1, t_2). \end{aligned}$$

Separable dependence functions are easily obtained by combining the isotropic models described in Table 1.1. However in practice, separability is often too simplistic, as it does not account for space-time interactions.

Symmetry is also an important property for spatio-temporal dependence functions: a random function $\{X(s, t) : (s, t) \in S\}$ with a semi-variogram γ is fully symmetric if

$$\gamma(s, t) = \gamma(s, -t) = \gamma(-s, t) = \gamma(-s, -t), \quad (s, t) \in S \times \mathbb{R}.$$

From practical perspective, with a fully symmetric random process, it is not possible to detect whether the regionalized variable is moving forward or backward in space and time. While such an assumption can be plausible in some cases, for instance when modelling daily mean temperature, it is not a reasonable assumption for phenomena that are under the influence of prevailing currents such as windstorms, rainstorms or tides. A simple non-separable and fully symmetric covariance function is (Christakos et al., 2000)

$$C_{s-t}(s, t) = C \left\{ \left(\frac{\|s\|^2}{\tau_1} + \frac{t^2}{\tau_2} \right)^{1/2} \right\}, \quad (s, t) \in S \times \mathbb{R}, \quad \tau_1, \tau_2 > 0$$

where τ_1, τ_2 are anisotropy parameters for the space and time dimensions, and C is a valid covariance function on \mathbb{R} . Cressie and Huang (1999) propose also a wide and flexible class of non-separable models, that were used by Huser and Davison (2014) to model extreme rainfall.

Due to their complexity, few non-fully symmetric models exist, however for modelling extreme windstorms in Europe, for which prevailing wind comes from the Atlantic Ocean, such models are required. A physically motivated construction (Cox and Isham, 1988) relies on the introduction of a wind vector $V \in \mathbb{R}^2$, such that for any valid covariance function C on \mathbb{R}^2 ,

$$C_V(s, t) = C(\|s - Vt\|), \quad (s, t) \in \mathbb{R}^2 \times [0, \infty).$$

The wind vector can be either randomly drawn from a bivariate distribution, or fixed, in which case the model is known as the *frozen field model*. In Chapter 5, we use this model to develop a stochastic weather generator for extreme windstorms.

1.6.2 Extremal processes based on Gaussian random functions

Section 1.6.1 presented basic notions of spatio-temporal statistics necessary to build heavy-tailed stochastic processes for extreme environmental events. We now describe extremal processes that use covariance or semi-variogram functions to characterize their dependence. Looking at Theorem 1.4 and the definition (1.27) of r -Pareto processes, it is clear that modelling the dependence of an extremal process is equivalent to specifying the distribution of the stochastic process W .

In the max-stable case, the spectral functions, denoted W_{\max} , take values in the set of continuous functions \mathcal{F}_+ over $S \subset \mathbb{R}^D$, and satisfy $E\{W_{\max}(s)\} = 1$ for all $s \in S$ and $E\{\sup_{s \in S} W_{\max}(s)\} < \infty$, while for the r -Pareto process (1.27), the angular component, denoted W_{pot} , refers to any process on the set of continuous functions over S with a unit norm, and is linked to W_{\max} through the relation

$$W_{\text{pot}} = \left[\frac{W_{\max}}{\|W_{\max}\|_{\text{ang}}} \right]^{1/\xi}. \quad (1.46)$$

Max-stable models were derived first so here presentation of extremal processes is given in terms of W_{\max} , but equation (1.46) allows us to switch representations.

Smith model

The first spatial model, introduced by Smith (1990) and called the Gaussian extreme value process, was physically inspired; the angular process was designed to represent a deterministic storm shape randomly shifted in space, yielding

$$W_{\max}(s) = \phi(s - U, C), \quad s \in S$$

where $\phi(s, C)$ is a Gaussian density on \mathbb{R}^D with mean zero and covariance function C , and U is uniformly distributed on S . For this model, the bivariate exponent measure (1.12) of the corresponding max-stable process Z at locations s_1 and s_2 is

$$\Lambda_{\text{Smith}} \{A_{\infty}^2(u)\} = \frac{1}{u_1} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{u_2}{u_1} \right) + \frac{1}{u_2} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{u_1}{u_2} \right), \quad u_1, u_2 > 0,$$

where $A_{\infty}^2(u) = \{x \in \mathbb{R}^2 : \max(x_1/u_1, x_2/u_2) \geq 1\}$, Φ is the standard Gaussian distribution function, and $a_h^2 = (s_1 - s_2)\Sigma^{-1}(s_1 - s_2)$ with $\Sigma_{ij} = C(s_i, s_j)$, ($i, j = 1, 2$). The Smith model yields extremal coefficient $\theta(h) = 2\Phi(a_h/2) < 2$, so the process is asymptotically dependent but satisfies

$$\theta(h) \rightarrow 2, \quad h \rightarrow \infty,$$

1.6. Spatio-temporal extremal processes based on Gaussian random functions

which implies that the dependence weakens as the distance between the sites s_1 and s_2 grows, and reaches independence as $h \rightarrow \infty$. The speed of weakening is determined by the covariance function C which can favour specific directions by using the anisotropic dependence functions described in Section 1.6.1.

Owing to its computational tractability and its relative simplicity, the Smith model was the first which was used in applications. Padoan et al. (2010) fit a Smith model to extreme rainfall in the U.S., but the Gaussian kernel makes the process too smooth to be realistic. For historical reasons, this extremal process is a standard in academic examples. In Smith (1990), the multivariate t density function are proposed as an alternative to the Gaussian kernel, but though more general, the corresponding model suffers from a lack of flexibility. More general models have been developed, which we now present.

Brown–Resnick model

A more realistic alternative to the Smith process is the Brown–Resnick model (Brown and Resnick, 1977): let X be a zero-mean Gaussian process with stationary increments and semi-variogram γ , and let $\sigma^2(s) = \text{var}\{X(s)\}$ for any $s \in S$. Then the limiting process of component-wise maxima defined in Theorem 1.4 with the spectral functions

$$W_{\max}(s) = \exp\{X(s) - \sigma^2(s)/2\}, \quad s \in S,$$

is a stationary Brown–Resnick process with standard Fréchet margins, whose distribution depends only on γ (Kablichko et al., 2009). In this case, the L -dimensional exponent measure of the corresponding max-stable process at locations s_1, \dots, s_L is (Huser and Davison, 2013)

$$\Lambda_{\text{BR}}\{A_\infty(u)\} = \sum_{l=1}^L \frac{1}{u_l} \Phi\{\eta_l(u), R_l\}, \quad (1.47)$$

where $A_\infty(u) = \{x \in \mathbb{R}^L : \max(x/u) \geq 1\}$, $\gamma_{l,i}$ denotes $\gamma(s_l, s_i)$ ($s_l, s_i \in S$), $\Phi(\cdot, R_l)$ is the multivariate normal distribution function with zero mean and covariance matrix R_l having (i, j) entry $(\gamma_{l,i} + \gamma_{l,j} - \gamma_{l,i,j})/\{2(\gamma_{l,i}\gamma_{l,j})^{1/2}\}$, with $i, j \in \{1, \dots, l-1, l+1, \dots, L\}$, and η_l is the $(L-1)$ -dimensional vector with i^{th} component $\eta_{l,i} = (\gamma_{l,i}/2)^{1/2} + \log(u_i/u_l)/(2\gamma_{l,i})^{1/2}$.

In Section 1.4.2 it was shown that the multivariate density function of the r -Pareto process is closely related to the intensity function λ , which can be found by taking partial derivatives of $\Lambda_{\text{BR}}\{A_\infty(u)\}$ with respect to u_1, \dots, u_L . For the Brown–Resnick

Chapter 1. Asymptotic tail distributions: theory and models

process, a closed form for λ_{BR} is (Engelke et al., 2015)

$$\lambda_{\text{BR}}(x) = \frac{|\Sigma|^{-1/2}}{x_1^2 x_2 \cdots x_L (2\pi)^{(L-1)/2}} \exp\left(-\frac{1}{2} \tilde{x}^T \Sigma^{-1} \tilde{x}\right), \quad x \in \mathbb{R}_+^L, \quad (1.48)$$

where \tilde{x} is the $(L-1)$ -dimensional vector with components $\{\log(x_i/x_1) + \gamma_{i,1} : i = 2, \dots, L\}$ and Σ is the $(L-1) \times (L-1)$ matrix with elements $\{\gamma_{i,1} + \gamma_{j,1} - \gamma_{i,j}\}_{i,j \in \{2, \dots, L\}}$. Wadsworth and Tawn (2014) derive an alternative symmetric expression of (1.48), which can be found in Appendix C and which will be useful in Chapter 2, but we prefer equation (1.48) which is more readily interpreted.

The Brown–Resnick model has bivariate extremal coefficient

$$\theta(h) = 2\Phi\left[\left\{\frac{\gamma(h)}{2}\right\}^{1/2}\right], \quad h = s_2 - s_1;$$

$\theta(h) \rightarrow 2$ as $h \rightarrow \infty$ only if $\gamma(h) \rightarrow \infty$. Thus, asymptotic independence for infinite distance between sites is possible only if the underlying Gaussian process is intrinsically stationary, whereas if the semi-variogram is bounded, i.e., $\gamma(h) \rightarrow \sigma^2$ as $h \rightarrow \infty$, then the extremal coefficient has the upper bound $2\{1 - \Phi(\sigma/2)\}$. The long range behaviour of the variogram determines extremal dependence at far distances, while, as explained in Section 1.6.1, its behaviour at the origin determines the smoothness of the process. Figure 1.1 illustrates the importance of these mathematical properties of the variogram and their impact on the appearance of a r -Pareto process.

For classical models the long-range dependence and the smoothness of the process are driven by only one parameter, for instance the shape parameter α in the power variogram, and will most likely fail to produce realistic simulations in practice. For this reasons, flexible models such as the Bernstein variogram are required so that the smoothness and long-range behaviour are governed by more than one parameter. Oesting et al. (2017b) propose the ‘gigogne’ variogram model

$$\gamma(h) = \gamma_{\text{matern}}(h) + \frac{\|h\|^2}{(1 + \|h\|^2)^\alpha}, \quad h \in \mathbb{R}^2,$$

where γ_{matern} drives the behaviour of the variogram at the origin while the second term determines the long range behaviour with the parameter α . Modelling spatio-temporal extremes requires very precise properties for variogram models and thus opens up a field of research for flexible parametric models that can capture both the smoothness and the middle to long distance behaviour of the process.

1.6. Spatio-temporal extremal processes based on Gaussian random functions

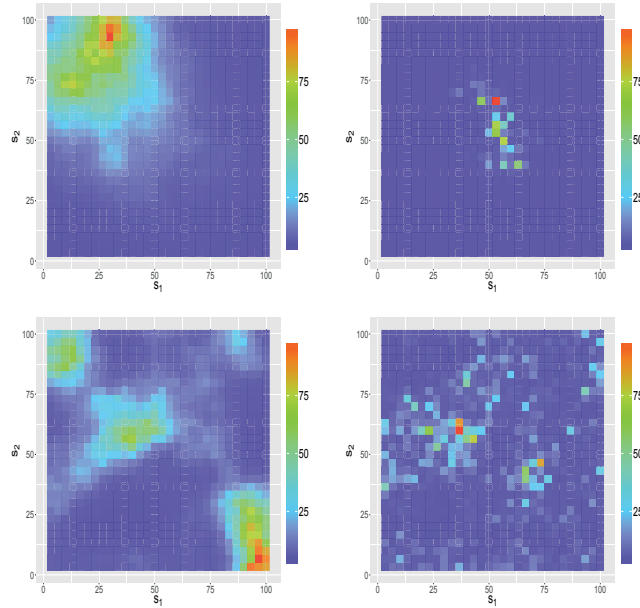


Figure 1.1 – Simulated r -Pareto process with risk functional $r(x) = \max_{s \in S} x(s)$ and different semi-variogram models. The intensity is fixed to a 100-year return level for each simulation: Top: $\gamma(h) = \|h/\lambda\|^\alpha$ with $\alpha = 1.8$, $\lambda = 30$ (left) and $\alpha = 0.5$, $\lambda = 1$ (right) ; Bottom: $\gamma(h) = 2\{1 - \exp(-\|h/\lambda\|^\alpha)\}$ with $\alpha = 1.8$, $\lambda = 30$ (left) and $\alpha = 0.5$, $\lambda = 1$ (right).

Schlather model

Schlather (2002) proposed to construct the angular process W_{\max} by taking the positive part of a stationary Gaussian process X with zero mean, unit variance and covariance function C , i.e.,

$$W_{\max}(s) = (2\pi)^{1/2} \max\{X(s), 0\}, \quad s \in S. \quad (1.49)$$

In this case, similarly to a Gaussian random field, the smoothness of the resulting max-stable process is driven by the behaviour of C at the origin. The bivariate exponent measure of the corresponding max-stable process Z at locations s_1 and s_2 is

$$\Lambda_{\text{Schlather}}\{A_\infty^2(u)\} = \frac{1}{2} \left(\frac{1}{u_1} + \frac{1}{u_2} \right) \left(1 + \left[1 - 2 \frac{\{1 + C(s_1 - s_2)\} u_1 u_2}{(u_1 + u_2)^2} \right] \right), \quad u_1, u_2 > 0,$$

where $A_\infty^2(u) = \{x \in \mathbb{R}^2 : \max(x_1/u_1, x_2/u_2) \geq 1\}$, which yields the extremal coefficient $\theta(h) = 1 + [\{1 - C(h)\}/2]^{1/2}$, with upper bound 1.838 when $S \subset \mathbb{R}^2$ (Davison and Gholamrezaee, 2012). The corresponding model is thus always asymptotically dependent and, from a practical point of view, the probability of observing an exceedance at location s_2 knowing that there is an excess at location s_1 decreases to a minimum of

0.162. This particularity of the Schlather model might not be realistic when studying physical processes over large regions, where we expect the probability to decrease to zero as the distance grows arbitrarily large. Schlather (2002) proposed to alleviate this by introducing random sets that ensure that extremes at sites sufficiently far apart are close to independence. This approach was successfully applied by Huser and Davison (2014) to model extreme rainfall in Switzerland, but statistical inference on the random set is difficult and thus is limited to problems of small size.

Extremal- t model

The extremal- t model (Opitz, 2013a), which appears as limit of component-wise maxima for all elliptical random functions, is characterized by the angular process

$$W_{\max}(s) = \pi^{1/2} 2^{-(\nu-2)/2} \max\{X(s), 0\}^\nu, \quad s \in S,$$

where $\nu > 0$ and X is a stationary Gaussian process X with unit variance and correlation function C . We see that for $\nu = 1$ we retrieve (1.49), which makes the extremal- t model a direct generalization of the Schlather model, and if we consider the limit case where we let $\nu \rightarrow \infty$, we obtain a Brown–Resnick model (Nikoloulopoulos et al., 2009). The bivariate exponent measure of the corresponding max-stable process Z at locations s_1 and s_2 is (Ribatet and Sedki, 2013)

$$\Lambda_t\{A_\infty^2(u)\} = \frac{1}{u_1} t_{\nu+1} \left\{ -\frac{C(h)}{b} + \frac{1}{b} \left(\frac{u_2}{u_1} \right)^{1/\nu} \right\} + \frac{1}{u_2} t_{\nu+1} \left\{ -\frac{C(h)}{b} + \frac{1}{b} \left(\frac{u_1}{u_2} \right)^{1/\nu} \right\}, \quad u_1, u_2 > 0,$$

where $A_\infty^2(u) = \{x \in \mathbb{R}^2 : \max(x_1/u_1, x_2/u_2) \geq 1\}$, $t_{\nu+1}$ is the standard distribution function of a t distribution with $\nu + 1$ degree of freedom and $b^2 = \{1 + C(h)^2\}/(\nu + 1)$. This yields the extremal coefficient

$$\theta(h) = 2t_{\nu+1} \left[(\nu + 1)^{1/2} \left\{ \frac{1 - C(h)}{1 + C(h)} \right\}^{1/2} \right], \quad (1.50)$$

from which we deduce that $\theta(h) \leq 2t_{\nu+1} \{(\nu + 1)^{1/2}\}$, so the resulting max-stable process is asymptotically dependent, a limitation similar to that of the Schlather model for applications in large regions if ν is too small.

A closed form for the L -dimensional exponent measure at locations s_1, \dots, s_L has been derived by Nikoloulopoulos et al. (2009) :

$$\Lambda_t\{A_\infty^L(u)\} = \sum_{l=1}^L u_l^{-1} t_{\nu+1} \left\{ \frac{u_{-l}}{u_l} - \Sigma_{-l,l}, (\nu + 1)^{-1} \left(\Sigma_{-l,-l} - \Sigma_{-l,l} \Sigma_{-l,l}^T \right) \right\},$$

where $A_\infty^L(u) = \{x \in \mathbb{R}^L : \max(x/u) \geq 1\}$, $\Sigma = \{C(s_l - s_k)\}_{l,k=1,\dots,L}$ is the correlation matrix of the underlying Gaussian process X and $t_{\nu+1}(\cdot, \Sigma)$ is the distribution function of a zero mean multivariate t distribution with $\nu + 1$ degrees of freedom and covariance matrix Σ . Thibaud and Opitz (2015) derived the corresponding intensity function,

$$\lambda_t(x) = \frac{\nu^{1-L} \pi^{(1-L)/2}}{|\Sigma|^{-1/2}} \frac{\Gamma\{(\nu+L)/2\}}{\Gamma\{(\nu+1)/2\}} \left(\prod_{l=1}^L |x_l| \right)^{1/\nu-1} \{T_{1/\nu}(x)^T \Sigma^{-1} T_{1/\nu}(x)\}^{-(\nu+D)/2}, \quad x \in \mathbb{R}^L, \quad (1.51)$$

where $T_{1/\nu}(x) = \text{sign}(x)|x|^{1/\nu}$, and used it to develop a censored likelihood inference procedure, similar to those presented in Chapter 2.

1.7 Discussion

This chapter has introduced notions required to study extreme environmental events. We first described the univariate asymptotic distributions of block maxima and threshold exceedances. For applications in the environment univariate quantities are limited as they do not capture the spatio-temporal nature of physical processes. Thus, the max-stable process is presented as functional generalization of the generalized extreme value distribution. We reviewed the main properties of this family of processes, which have been successfully applied in environmental studies.

For risk estimation and mitigation, regulators and (re-)insurers might often prefer models for single extreme events. For this reason, we presented the r -Pareto process which generalizes peaks-over-thresholds analysis to functions. The proofs of the original construction by Dombry and Ribatet (2015) were revisited and we emphasized their limitations from the perspective of applications. Indeed, the r -Pareto process is limited to the Fréchet domain of attraction, while its univariate counterpart, the generalized Pareto distribution, covers the three possible regimes of tail decay. Chapter 4 generalizes the r -Pareto process to become the functional equivalent of the generalized Pareto distribution. Also, in this framework we defined the risk through the aggregation of the original process by a risk functional. In practice, the joint distribution of more than one aggregations might be of interest, for instance when modelling the cumulative rainfall over several catchments of a river. In Chapter 3, we derive the multivariate asymptotic tail distribution of a family of risk functionals and apply this result to estimate the local tail behaviour of extreme temperature in the South of France from aggregated data.

Catalogues of extreme historical events are commonly used for risk estimation of natural hazards. One way to enlarge these catalogues is stochastic weather generators,

Chapter 1. Asymptotic tail distributions: theory and models

for which classical models rely on Gaussian processes and therefore fail to accurately reproduce extreme events. However, a class of extremal processes whose dependence is driven by Gaussian random fields has been developed, and so we presented the basic results of spatial statistics with classical models of dependence functions. Then, we introduced extremal processes, highlighting their strengths and limitations with an emphasis on the requirements for the semi-variogram function to build realistic stochastic weather generators: flexible dependence models that can capture both local and mid- to long-range extremal behaviour need to be further developed.

For environmental applications, we must be able to handle large datasets, such as climate model outputs, and thus statistical inference in high-dimensions, around a few thousand, must be possible. Chapter 2 reviews existing techniques for fitting an extremal process, but these methods are usually limited, either theoretically or computationally, to a few dozen locations. For this reason, we develop alternative methodologies that can handle large datasets.

2 High-dimensional peaks-over-threshold inference

This chapter is a postprint version of the article ‘High-dimensional peaks-over-threshold inference’ written with Anthony Davison and published in *Biometrika* (de Fondeville and Davison, 2018). Some parts of this chapter may be redundant with other sections of the thesis, but we decided to keep the chapter self-standing for clarity. Only light modifications were performed for consistency with the rest of the thesis.

2.1 Introduction

Recent contributions to extreme value theory describe models capable of handling spatio-temporal phenomena (e.g., Kabluchko et al., 2009) and provide a flexible framework for modelling rare events, but their complexity makes inference difficult, if not intractable, for high-dimensional data. For instance, the number of terms in the block maximum likelihood for a Brown–Resnick process grows with dimension D like the Bell numbers (Huser and Davison, 2013), so computationally cheaper methods like composite likelihood (Padoan et al., 2010) or the inclusion of partition information (Stephenson and Tawn, 2005; Thibaud et al., 2016) have been advocated. The first is slow and, though more efficient, the second is liable to bias if the partition is incorrect (Wadsworth, 2015).

An attractive alternative to the use of block maxima is peaks-over-threshold analysis, which includes more information by focusing on single extreme events. In the multivariate case, specific definitions of exceedances have been used (e.g., Rootzén and Tajvidi, 2006; Ferreira and de Haan, 2014; Engelke et al., 2015), which can be unified within the framework of r -Pareto processes (Dombry and Ribatet, 2015). For this approach, a full likelihood is often available in closed form, thus increasing the maximum number of variables that can be jointly modelled from a handful to a few dozen, but biased estimation may occur if non-extreme components are included.

Censored likelihood, proposed in this context by Wadsworth and Tawn (2014), is more robust with regard to non-extreme observations, but it involves multivariate normal distribution functions, which can be computationally expensive. Nevertheless, inference is feasible for $D \approx 30$.

Nonparametric alternatives to full likelihood inference developed using the tail dependence coefficient (Davis and Mikosch, 2009; Davis et al., 2013a) or the stable tail dependence function (Einmahl et al., 2016a) rely on pairwise estimators and allow peaks-over-threshold inference for $D \approx 100$, but are potentially inefficient and may be limited by combinatorial considerations.

Applications of max-stable processes (e.g., Asadi et al., 2015) or Pareto processes (Thibaud and Opitz, 2015) have focused on small regions and have used at most a few dozen locations with particular types of exceedance, but exploitation of much larger gridded datasets, along with more complex definitions of risk, is needed for a better understanding of extreme events and to reduce model uncertainties. The goals of this paper are to highlight the advantages of functional peaks-over-threshold modelling using r -Pareto processes, to show the feasibility of high-dimensional inference for the Brown–Resnick model with hundreds of locations, and to compare the robustness of different procedures with regard to finite thresholds. We develop an estimation method based on the gradient score (Hyvärinen, 2005) for a general notion of exceedances, for which the computation of multivariate normal probabilities is not needed and computational complexity is driven by matrix inversion, as with classical Gaussian likelihood inference. This method focuses on single extreme events and a general notion of exceedance, modelled by Pareto processes, instead of the max-stable approach.

2.2 Modelling exceedances over a high threshold

2.2.1 Functional regular variation

Let S be a compact metric space, such as $[0, 1]^2$ for spatial applications. We write $\mathcal{F}_+ = C\{S, [0, \infty)\}$ for the closed subset of the Banach space of continuous functions $x : S \rightarrow \mathbb{R}$ endowed with the uniform norm $\|x\| = \sup_{s \in S} |x(s)|$, \mathcal{F}_0 for \mathcal{F} with the singleton $\{0\}$ excluded, and $\mathcal{B}(\Xi)$ for the Borel σ -algebra associated to a metric space Ξ . Let $M_{\mathcal{F}_0}$ denote the class of Borel measures on $\mathcal{B}(\mathcal{F}_0)$, and say that a set $\mathcal{A} \in \mathcal{B}(\mathcal{F}_0)$ is bounded away from $\{0\}$ if $d(\mathcal{A}, \{0\}) = \inf_{x \in \mathcal{A}} \|x\| > 0$. A sequence of measures $\{\Lambda_n\} \subset M_{\mathcal{F}_0}$ is said to converge to a limit $\Lambda \in M_{\mathcal{F}_0}$ (Hult and Lindskog, 2005), if $\lim_{n \rightarrow \infty} \Lambda_n(\mathcal{A}) = \Lambda(\mathcal{A})$, for all $\mathcal{A} \in \mathcal{B}(\mathcal{F}_0)$ bounded away from $\{0\}$ with $\Lambda(\partial\mathcal{A}) = 0$, where $\partial\mathcal{A}$ denotes the boundary

2.2. Modelling exceedances over a high threshold

of \mathcal{A} . For equivalent definitions of this so-called \hat{w} -convergence, see Lindskog et al. (2014, Theorem 2.1).

Regular variation provides a flexible mathematical setting in which to characterize the tail behaviour of random processes in terms of \hat{w} -convergence of measures. A stochastic process X with sample paths in \mathcal{F}_0 is regularly varying (Hult and Lindskog, 2005) if there exists a sequence of positive real numbers a_1, a_2, \dots with $\lim_{n \rightarrow \infty} a_n = \infty$, and a measure $\Lambda \in M_{\mathcal{F}_0}$ such that

$$n \Pr(a_n^{-1} X \in \cdot) \rightarrow \Lambda(\cdot), \quad n \rightarrow \infty; \quad (2.1)$$

then we write $X \in \text{RV}(\mathcal{F}_0, a_n, \Lambda)$. For a normalized process X^* , obtained for instance by standardizing the margins of X to unit Fréchet (Coles and Tawn, 1991, Section 5) or unit Pareto (Klüppelberg and Resnick, 2008), regular variation is equivalent to the convergence of the renormalised pointwise maximum $n^{-1} \max_{i=1}^n X_i^*$ of independent replicates of X^* to a non-degenerate process Z^* , with unit Fréchet margins and exponent measure Λ^* (de Haan and Lin, 2001). The process Z^* is called simple max-stable, and X^* is said to lie in the max-domain of attraction of Z^* .

Regular variation also impacts the properties of exceedances over high thresholds. For any nonnegative measurable functional $r : \mathcal{F}_0 \rightarrow [0, \infty)$ and stochastic process $\{X(s)\}_{s \in \mathcal{S}}$, an r -exceedance is defined to be an event $\{r(X) > u_n\}$ where the threshold u_n is such that $\Pr\{r(X) > u_n\} \rightarrow 0$ as $n \rightarrow \infty$. We further require that r is homogeneous, i.e., there exists $\kappa > 0$ such that $r(ax) = a^\kappa r(x)$, for $a > 0$ and $x \in \mathcal{F}_0$. As $r(\cdot)$ could be replaced by $r(\cdot)^{1/\kappa}$ without loss of generality, below we assume that $\kappa = 1$. Dombry and Ribatet (2015) called r a cost functional and in his 2013 Université de Montpellier II PhD thesis Thomas Opitz called it a radial aggregation function, but we prefer the term risk functional because r determines the type of extreme event whose risk is to be studied.

A natural formulation of subsequent results on r -exceedances uses a pseudo-polar decomposition. For a norm $\|\cdot\|_{\text{ang}}$ on \mathcal{F}_0 , called the angular norm, and a risk functional r , a pseudo-polar transformation T_r is a map such that

$$T_r : \mathcal{F}_0 \rightarrow [0, \infty) \times \mathcal{S}_{\text{ang}}, \quad T_r(x) = \left\{ r' = r(x), w = \frac{x}{\|x\|_{\text{ang}}} \right\},$$

where \mathcal{S}_{ang} is the unit sphere $\{x \in \mathcal{F}_0 : \|x\|_{\text{ang}} = 1\}$. If r is continuous and T is restricted to $\{x \in \mathcal{F}_0 : r(x) > 0\}$, then T is a homeomorphism with inverse $T_r^{-1}(r', w) = r' \times w / r(w)$.

Theorem 2.1 in Lindskog et al. (2014) provides an equivalent pseudo-polar formula-

Chapter 2. High-dimensional peaks-over-threshold inference

tion of equation (2.1). For any $X \in \text{RV}(\mathcal{F}_0, a_n, \Lambda)$ and any uniformly continuous risk functional r not vanishing Λ -almost everywhere, there exist $\xi > 0$ and a measure σ_r on $\mathcal{B}(\mathcal{S}_{\text{ang}})$ such that

$$n \Pr \left[T_r^{-1} \left\{ a_n^{-1} r(X), \frac{X}{\|X\|_{\text{ang}}} \right\} \in \cdot \right] \rightarrow \Lambda \circ T_r^{-1}(\cdot) = \Lambda_\xi \times \sigma_r(\cdot), \quad n \rightarrow \infty, \quad (2.2)$$

where $\Lambda_\xi[r', \infty) = (r')^{-1/\xi}$ and $\sigma_r(\cdot) = \Lambda \{x \in \mathcal{F}_0 : r(x) \geq 1, x/\|x\|_{\text{ang}} \in (\cdot)\}$ is called the angular measure. The converse holds if $\{x \in \mathcal{F}_0 : r(x) = 0\} = \emptyset$ (Lindskog et al., 2014, Corollary 4.4).

The functional $r(x) = \sup_{s \in S} x(s)$, used by Rootzén and Tajvidi (2006) in a multivariate setting and by Ferreira and de Haan (2014) for continuous processes, implies that realisations of $X(s)$ exceeding the threshold at any location $s \in S$ are labelled extreme, but this functional can only be used in applications where $X(s)$ is observed throughout S . Thus it may be preferable to use functions such as $\max_{s \in S'} X(s)$ or $\max_{s \in S'} X(s)/u(s)$, where $S' \subset S$ is a finite set of gauged sites. Other risk functionals include $\int_S X(s) ds$ for the study of areal rainfall (Coles and Tawn, 1996), $\min_{s \in S'} X(s)/u(s)$, or $X(s_0)$ for risks impacting a specific location s_0 . Although the choice of risk functional allows a focus on particular types of extreme event, the choice of the angular norm $\|\cdot\|_{\text{ang}}$ has no impact and is usually made for convenience.

Finally, for a common angular norm $\|\cdot\|_{\text{ang}}$, the angular measures of two risk functionals r_1 and r_2 that are strictly positive Λ -almost everywhere are linked by the expression

$$\sigma_{r_1}(dw) = \left\{ \frac{r_1(w)}{r_2(w)} \right\}^{1/\xi} \sigma_{r_2}(dw), \quad dw \in \mathcal{B}(\mathcal{S}_{\text{ang}}). \quad (2.3)$$

Equation (2.3) is useful for simulation and when we are interested in r_2 -exceedances but inference has been performed based on r_1 . All the previous definitions and results hold for finite dimensions, i.e., for D -dimensional random vectors, by replacing \hat{w} -convergence by vague convergence (Resnick, 2007, Section 3.3.5) on $M_{\mathbb{R}_+^D \setminus \{0\}}$, the class of Borel measures on $\mathcal{B}(\mathbb{R}_+^D \setminus \{0\})$ endowed with the $\|\cdot\|$ norm; see the PhD thesis of Thomas Opitz mentioned above.

2.2.2 r -Pareto processes

In this section, r denotes a functional that is nonnegative and homogeneous with $\alpha = 1$. The r -Pareto processes (Dombry and Ribatet, 2015) are important for modelling

2.2. Modelling exceedances over a high threshold

exceedances, and may be constructed as

$$P = U \frac{W}{r(W)}, \quad (2.4)$$

where U is a univariate Pareto random variable with $\Pr(U > r') = (r')^{-1/\xi}$ ($r' \geq 1$) and W is a random process with sample paths in $\mathcal{S}_{\text{ang}}^r = \{x \in \mathcal{F}_0 : r(x) \geq 1, \|x\|_{\text{ang}} = 1\}$ and probability measure σ_{ang} . The process P is then called an r -Pareto process with tail index $\xi > 0$ and angular measure σ_{ang} ; in order to distinguish different Pareto processes, below we use the notation P_{ξ, σ_r}^r for P .

An important property of r -Pareto processes is threshold-invariance: for all $\mathcal{A} \in \mathcal{B}(\{x \in \mathcal{F}_0 : r(x) \geq 1\})$ and all $u \geq 1$ such that $\Pr\{r(P) \geq u\} > 0$,

$$\Pr\{u^{-1}P \in \mathcal{A} \mid r(P) \geq u\} = \Pr(P \in \mathcal{A}). \quad (2.5)$$

Furthermore, for $X \in \text{RV}(\mathcal{F}_0, a_n, \Lambda)$ with index $\xi > 0$ and for a risk functional r that is continuous at the origin and does not vanish Λ -almost everywhere, the distribution of the r -exceedances converges weakly to that of a r -Pareto process, i.e.,

$$\Pr\{u^{-1}X \in \cdot \mid r(X) \geq u\} \rightarrow \Pr\left(P_{\xi, \sigma_r}^r \in \cdot\right), \quad u \rightarrow \infty, \quad (2.6)$$

with tail index ξ and probability measure σ_r defined in (2.2) (Dombry and Ribatet, 2015, Theorem 2). When working with a normalized process X^* , the exponent measure Λ^* of the limiting max-stable process Z^* and the measure $\Lambda^1 \times \sigma_r$ of the Pareto process are equal up to a coordinate transform, as suggested by equation (2.2).

For two different risk functionals r_1 and r_2 and angular measures σ_{r_1} and σ_{r_2} for which there exists $\Lambda \in M_{\mathcal{F}_0}$ such that

$$\Lambda \circ T_{r_1}^{-1}(\cdot) = \Lambda \circ T_{r_2}^{-1}(\cdot),$$

the associated Pareto processes $P_{\xi, \sigma_{r_1}}^{r_1}$ and $P_{\xi, \sigma_{r_2}}^{r_2}$ are defined on different subsets of \mathcal{F}_0 , but, as suggested by equation (2.3), if there exists a threshold u_{\min} such that

$$\{x \in \mathcal{F}_0 : r_1(x) \geq u_{\min}\} \subset \{x \in \mathcal{F}_0 : r_2(x) \geq 1\},$$

then

$$\Pr\left\{\frac{P_{\xi, \sigma_{r_1}}^{r_1}}{u} \in \cdot \mid r_2\left(P_{\xi, \sigma_{r_1}}^{r_1}\right) \geq u\right\} = \Pr\left(P_{\xi, \sigma_{r_2}}^{r_2} \in \cdot\right), \quad u \geq u_{\min}. \quad (2.7)$$

Simulation of r -Pareto processes is feasible only for a few risk functionals, such as $r_1(x) = \|x\|_1$, but equation (2.7) can be used to obtain samples of one process from

Chapter 2. High-dimensional peaks-over-threshold inference

those of another: for independent replicates x^1, \dots, x^N from $P_{\xi, \sigma_{r_1}}^{r_1}$, $\{y^n = x^n / u_{\min} : r_2(y^n) \geq 1\}$ is a sample from $P_{\xi, \sigma_{r_2}}^{r_2}$.

Finally, let σ_r be a probability measure on $\mathcal{S}_{\text{ang}}^r$, and define the process

$$M(s) = \max_{n \in \mathbb{N}} U^n \frac{W^n(s)}{r(W^n)}, \quad s \in S, \quad (2.8)$$

where $\{U^n : n \in \mathbb{N}\}$ is a Poisson process on $(0, \infty)$ with intensity $u^{-2} du$ and W^1, W^2, \dots are replicates of a process W with probability measure σ_r . Then M is a max-stable process with exponent measure $\Lambda\{\mathcal{A}_{\max}(x)\} = \Lambda^1 \times \sigma_r\{\mathcal{A}_{\max}(x)\}$, where $\mathcal{A}_{\max}(x) = \{y \in \mathcal{F}_0 : \sup_{s \in S} y(s)/x(s) \geq 1\}$. Thus equation (2.8) connects r -Pareto processes and their max-stable counterparts.

2.2.3 Extreme value processes associated to log-Gaussian random functions

We focus on a class of r -Pareto processes based on log-Gaussian stochastic processes, whose max-stable counterparts are Brown–Resnick processes. This class is particularly useful, not only for its flexibility but also because it is based on Gaussian models widely used in applications. Chiles and Delfiner (2012, p. 84–108) review these classical models.

Let Z be a zero-mean Gaussian process with stationary increments, i.e., the semi-variogram $\gamma(s, s') = E\{[Z(s) - Z(s')]^2\}/2$ ($s, s' \in S$) depends only on the difference $s - s'$ (Chiles and Delfiner, 2012, p. 30), and let $\sigma^2(s) = \text{var}\{Z(s)\}$. If Z^1, Z^2, \dots are independent replicates of Z and $\{U^n : n \in \mathbb{N}\}$ is a Poisson process on $(0, \infty)$ with intensity $u^{-2} du$, independent of the Z^n , then

$$M(s) = \max_{n \in \mathbb{N}} U^n \exp\{Z^n(s) - \sigma^2(s)/2\}, \quad s \in S, \quad (2.9)$$

is a stationary Brown–Resnick process with standard Fréchet margins, whose dependence is driven only by γ (Kablichko et al., 2009); such processes are max-stable. Let γ_θ denote a parametrized semi-variogram whose parameter θ lies in a compact set Θ , and let σ_θ^2 denote the corresponding variance function.

Let s_1, \dots, s_L be locations of interest in S . In the rest of the paper, x denotes an element of \mathbb{R}_+^L with components $x_l \equiv x(s_l)$ ($l = 1, \dots, L$). The finite-dimensional exponent

measure $\Lambda_\theta(\cdot)$ of a simple Brown–Resnick process with L variables is

$$\Lambda_\theta \{ \mathcal{A}_{\max}(x) \} = \mathbb{E} \left[\max_{l=1, \dots, L} \left\{ \frac{Z(s_l) - \sigma_\theta^2(s_l)/2}{x_l} \right\} \right], \quad (2.10)$$

where $\Lambda_\theta(\cdot)$ is the finite-dimensional projection of the measure defined in (2.1). Then we can write (Huser and Davison, 2013)

$$\Lambda_\theta \{ \mathcal{A}_{\max}(x) \} = \sum_{l=1}^L \frac{1}{x_l} \Phi \{ \eta_l(x), R_l \}, \quad (2.11)$$

where η_l is the $(L-1)$ -dimensional vector with i^{th} component $\eta_{l,i} = (\gamma_{l,i}/2)^{1/2} + \log(x_i/x_l)/(2\gamma_{l,i})^{1/2}$, $\gamma_{l,i}$ denotes $\gamma(s_l, s_i)$ ($s_l, s_i \in S$), and $\Phi(\cdot, R_l)$ is the multivariate normal distribution function with zero mean and covariance matrix R_l having (i, j) entry $(\gamma_{l,i} + \gamma_{l,j} - \gamma_{l,i,j})/2(\gamma_{l,i}\gamma_{l,j})^{1/2}$, where $i, j \in \{1, \dots, l-1, l+1, \dots, L\}$.

The r -Pareto processes associated to log-Gaussian random functions are closely related to the intensity function λ_θ corresponding to the measure Λ_θ , which can be found by taking partial derivatives of $\Lambda_\theta(x)$ with respect to x_1, \dots, x_L , yielding (Engelke et al., 2015)

$$\lambda_\theta(x) = \frac{|\Sigma_\theta|^{-1/2}}{x_1^2 x_2 \cdots x_L (2\pi)^{(L-1)/2}} \exp \left(-\frac{1}{2} \tilde{x}^T \Sigma_\theta^{-1} \tilde{x} \right), \quad x \in \mathbb{R}_+^L, \quad (2.12)$$

where \tilde{x} is the $(L-1)$ -dimensional vector with components $\{\log(x_i/x_1) + \gamma_{i,1} : i = 2, \dots, L\}$ and Σ_θ is the $(L-1) \times (L-1)$ matrix with elements $\{\gamma_{i,1} + \gamma_{j,1} - \gamma_{i,j}\}_{i,j \in \{2, \dots, L\}}$. Wadsworth and Tawn (2014) derive an alternative symmetric expression for (2.12) that will be useful in Section 2.3.3, but Equation (2.12) is more readily interpreted. Similar expressions exist for extremal- t processes (Thibaud and Opitz, 2015).

2.3 Inference for r -Pareto processes

2.3.1 Generalities

In this section, x^1, \dots, x^N are independent replicates of an L -dimensional r -Pareto random vector P with tail index $\xi = 1$ and y^1, \dots, y^N are independent replicates of a regularly-varying L -dimensional random vector Y^* with normalized margins.

As in the univariate setting, statistical inference based on block maxima and the max-stable framework discards information by focusing on maxima instead of single events. These models are difficult to fit not only due to the small number of replicates, but

Chapter 2. High-dimensional peaks-over-threshold inference

also because the likelihood is usually too complex to compute in high dimensions (Castruccio et al., 2016). For the Brown–Resnick process, the full likelihood cannot be computed for L greater than around ten (Huser and Davison, 2013), except in special cases. When the occurrence times of maxima are available, inference is typically possible up to $L \approx 30$ (Stephenson and Tawn, 2005; Thibaud et al., 2016).

A useful alternative is composite likelihood inference (Padoan et al., 2010; Varin et al., 2011) based on subsets of observations of sizes smaller than L , which trades off a gain in computational efficiency against a loss of statistical efficiency. The number of possible subsets increases very rapidly with L , and their selection can be vexed, though some statistical efficiency can be retrieved by taking higher-dimensional subsets. Castruccio et al. (2016) found higher-order composite likelihoods to be more robust than the spectral estimator, but in realistic cases these methods are limited to fairly small dimensions.

Estimation based on threshold exceedances and the Pareto process has the advantages that individual events are used, the likelihood function is usually simpler, and the choice of the risk functional can tailor the definition of an exceedance to the application. Equation (2.2) suggests that the choice of risk functional should not affect the estimates, but this is not entirely true, because the threshold cannot be taken arbitrarily high and the events selected depend on the risk functional, the choice of which enables the detection of mixtures in the extremes and can improve sub-asymptotic behaviour by fitting the model using only those observations closest to the chosen type of extreme event. For example, we might expect the extremal dependence of intense local rainfall events to differ from that of heavy large-scale precipitation, even in the same geographical region.

The probability density function of a Pareto process for r -exceedances over the threshold vector $u \in \mathbb{R}_+^L$ can be found by rescaling the intensity function λ_θ by $\Lambda_\theta\{\mathcal{A}_r(u)\}$, yielding

$$\lambda_{\theta,u}^r(x) = \frac{\lambda_\theta(x)}{\Lambda_\theta\{\mathcal{A}_r(u)\}}, \quad x \in \mathcal{A}_r(u), \quad (2.13)$$

where

$$\Lambda_\theta\{\mathcal{A}_r(u)\} = \int_{\mathcal{A}_r(u)} \lambda_\theta(x) dx, \quad (2.14)$$

and $\mathcal{A}_r(u)$ is the exceedance region $\{x \in \mathbb{R}_+^L : r(x/u) \geq 1\}$. Equation (2.13) yields the loglikelihood

$$\ell(\theta; x^1, \dots, x^N) = \sum_{n=1}^N \mathbf{1} \left\{ r \left(\frac{x^n}{u} \right) \geq 1 \right\} \log \left[\frac{\lambda_\theta(x^n)}{\Lambda_\theta\{\mathcal{A}_r(u)\}} \right], \quad (2.15)$$

where division of vectors is component-wise and 1 denotes the indicator function. Maximization of ℓ gives an estimator $\hat{\theta}_r(x^1, \dots, x^N)$ that is consistent, asymptotically normal and efficient under mild conditions.

Numerical evaluation of the L -dimensional integral $\Lambda_\theta\{\mathcal{A}_r(u)\}$ is generally intractable for large L , though it simplifies for certain risk functionals; an example is $r(x) = \max_l x_l$, for which the integral is a sum of multivariate probability functions; see Equation (2.11). Similarly, $\Lambda_\theta\{\mathcal{A}_r(u)\}$ does not depend upon θ when $r(x) = L^{-1} \sum_d x_d$ (Coles and Tawn, 1991); we call the corresponding version of (2.15) the spectral loglikelihood and its maximiser the spectral estimator.

In practice observations cannot be assumed to be exactly Pareto distributed; it is usually more plausible that they lie in the domain of attraction of some extremal process. As a consequence of Theorem 3.1 of de Haan and Resnick (1993), asymptotic properties of $\hat{\theta}_r(x^1, \dots, x^N)$ hold for $\hat{\theta}_r(y^1, \dots, y^N)$ as $N \rightarrow \infty$ and $u \rightarrow \infty$ with the number of exceedances $N_u = o(N) \rightarrow \infty$; see Section 2.3.3. However, the threshold u is finite and thus low components of $y^n \in \mathcal{A}_r(u)$ may lead to biased estimation. As it is due to model mis-specification, this bias is unavoidable, and moreover, it grows with L , so these methods can perform poorly, especially if the extremal dependence is weak, as it is then more likely that at least one component of y^n will be small (Engelke et al., 2015; Thibaud and Opitz, 2015; Huser et al., 2016). The bias can be reduced by a censored likelihood proposed in the multivariate setting by Joe et al. (1992) and used for the Brown–Resnick model by Wadsworth and Tawn (2014) and for the extremal- t process by Thibaud and Opitz (2015). This method works well in practice but typically requires the computation of multivariate normal and t probabilities, which can be challenging in realistic cases if standard code is used. Some modest changes to the code to perform quasi-Monte Carlo maximum likelihood estimation with hundreds of locations are described in Section 2.3.2.

For spatio-temporal applications, inference for r -Pareto processes must be performed using data from thousands of locations, and in Section 2.3.3 we discuss an approach that applies to a wide range of risk functionals and is computationally fast, statistically efficient and robust with regard to finite thresholds.

2.3.2 Efficient censored likelihood inference

Censored likelihood estimation for extreme value process associated to log-Gaussian random functions was developed by Wadsworth and Tawn (2014) and is based on equation (2.15) with $\max_l x_l / u_l$ as risk functional and where any component lying below the threshold vector $(u_1, \dots, u_L) > 0$ is treated as censored. The corresponding

Chapter 2. High-dimensional peaks-over-threshold inference

estimator has a higher variance but a lower bias than the spectral estimator. The censored likelihood density function for the Brown–Resnick process is (Asadi et al., 2015)

$$\lambda_{\theta,u}^{\text{cens}}(x) = \frac{1}{\Lambda_{\theta}\{\mathcal{A}_{\max}(u)\}} \frac{1}{x_1^2 x_2 \cdots x_k} \phi_{k-1}(\tilde{x}_{2:k}; \Sigma_{2:k}) \Phi_{L-k}\{\mu_{\text{cens}}(x_{1:k}), \Sigma_{\text{cens}}(x_{1:k})\}, \quad (2.16)$$

where $x \in \mathcal{A}_{\max}(u)$, k components exceed their thresholds, $\tilde{x}_{2:k}$ and $\Sigma_{2:k}$ are subsets of the variables \tilde{x} and Σ_{θ} in equation (2.12), and ϕ_{k-1} and Φ_{L-k} are the multivariate Gaussian density and distribution functions with mean zero. The argument and covariance matrix for Φ_{L-k} are

$$\begin{aligned} \mu_{\text{cens}}(x_{1:k}) &= \{\log(u_j/x_j) + \gamma_{j,1}\}_{j=k+1,\dots,L} - \Sigma_{(k+1):L,2:k} \Sigma_{2:k,2:k}^{-1} \tilde{x}_{2:k}, \\ \Sigma_{\text{cens}}(x_{1:k}) &= \Sigma_{(k+1):L,(k+1):L} - \Sigma_{(k+1):L,2:k} \Sigma_{2:k,2:k}^{-1} \Sigma_{2:k,(k+1):L}. \end{aligned}$$

Wadsworth and Tawn (2014) derived similar expressions. The estimator

$$\hat{\theta}_{\text{cens}}(y^1, \dots, y^N) = \arg \max_{\theta \in \Theta} \sum_{n=1, \dots, N} 1 \left\{ \max_l \left(\frac{y_l^n}{u_l} \right) \geq 1 \right\} \log \lambda_{\theta,u}^{\text{cens}}(y^n), \quad (2.17)$$

is also consistent and asymptotically normal as $u \rightarrow \infty$, $N \rightarrow \infty$, $N_u \rightarrow \infty$ with $N_u = o(N)$. For finite thresholds, $\hat{\theta}_{\text{cens}}$ has been found to be more robust to the presence of low components than the spectral estimator (Engelke et al., 2015; Huser et al., 2016), but it is awkward due to the potentially large number of multivariate normal integrals involved, thus far limiting its application to $L \lesssim 30$ (Wadsworth and Tawn, 2014; Thibaud et al., 2016).

When maximizing the right-hand side of equation (2.17), the normalizing constant $\Lambda_{\theta}\{\mathcal{A}_{\max}(u)\}$ described in equation (2.10) and the multivariate normal distribution functions require the computation of multidimensional integrals. Theorem 7 of Geyer (1994) suggests that we approximate $\hat{\theta}_{\text{cens}}$ by maximizing

$$\ell_{\text{cens}}^p(\theta) = \sum_{n=1}^N 1 \left\{ \max \left(\frac{x^n}{u} \right) \geq 1 \right\} \left[\log \left\{ \frac{\phi_{k-1}(\tilde{x}_{2:k}; \Sigma_{2:k})}{(x_1^n)^2 x_2^n \cdots x_k^n} \right\} + \log \frac{\Phi_{L-k}^p\{\mu_{\text{cens}}(x_{1:k}^n), \Sigma_{\text{cens}}(x_{1:k}^n)\}}{\Lambda_{\theta}^p\{\mathcal{A}_{\max}(u)\}} \right],$$

where Φ_{L-k}^p and Λ_{θ}^p are Monte Carlo estimates of the corresponding integrals based on p simulated samples, yielding a maximizer $\hat{\theta}_{\text{cens}}^p$ that converges almost surely to $\hat{\theta}_{\text{cens}}$ as $p \rightarrow \infty$.

Classical Monte Carlo estimation for multivariate integrals yields a probabilistic error

bound that is $O(\omega p^{-1/2})$, where $\omega = \omega(\phi)$ is the square root of the variance of the integrand ϕ . Quasi-Monte Carlo methods can achieve higher rates of convergence and thus improve computational efficiency while preserving the consistency of $\hat{\theta}_{\text{cens}}^p$. For estimation of multivariate normal distribution functions, Genz and Bretz (2009, Section 4.2.2) advocate the use of randomly-shifted deterministic lattice rules, which can achieve a convergence rate of order $O(p^{-2+\epsilon})$ for some $\epsilon > 0$. Lattice rules rely on regular sampling of the hypercube $[0, 1]^L$, taking

$$z_q = |2 \times \text{frac}(qv' + \Delta) - 1|, \quad q = 1, \dots, p, \quad (2.18)$$

where $\text{frac}(z)$ denotes the component-wise fractional part of $z \in \mathbb{R}^L$, p is a prime number of samples in the hypercube $[0, 1]^L$, $v' \in \{1, \dots, p\}^L$ is a carefully-chosen generating vector and $\Delta \in [0, 1]^L$ is a uniform random shift. Fast construction rules exist to find an optimal v' for given numbers of dimensions L and samples p (Nuyens and Cools, 2004). The existence of generating vectors achieving a nearly optimal convergence rate, with integration error independent of the dimension, has been proved and methods for their construction exist (Dick and Pillichshammer, 2010).

Our implementation of this approach applied to equation (2.17) and coupled with parallel computing is tractable for L of the order of a few hundred; see Appendix B.1. Our algorithm extends to the extremal- t model by writing multivariate t probabilities in terms of the multivariate normal distribution function; see Genz and Bretz (2009) for more details.

2.3.3 Score matching

Classical likelihood inference requires either evaluation or simplification of the scaling constant $\Lambda_{\theta}\{\mathcal{A}_r(u)\}$, whose complexity increases with the number of dimensions. Hence we seek alternatives that do not require its computation.

Let \mathcal{S} be a sample space such as \mathbb{R}_+^L , \mathcal{P} a convex class of probability measures on \mathcal{S} , and X a random variable with distribution $F \in \mathcal{S}$. A proper scoring rule (Gneiting and Raftery, 2007) is a functional $\delta : \mathcal{P} \times \mathcal{S} \rightarrow \mathbb{R}$ such that

$$\Delta_{\delta}(G, F) = E_X \{\delta(G, X)\} - E_X \{\delta(F, X)\} \geq 0, \quad G \in \mathcal{P}. \quad (2.19)$$

The scoring rule is said to be strictly proper if $\Delta_{\delta}(G, F) = 0$ if and only if $G = F$, and, under this hypothesis, Δ_{δ} defines a divergence measure on \mathcal{P} (Thorarinsdottir et al., 2013).

Let δ denote a strictly proper scoring rule, let $\{F_{\theta} : \theta \in \Theta\} \subset \mathcal{S}$ be a parametric family of

Chapter 2. High-dimensional peaks-over-threshold inference

distributions, and let X^1, \dots, X^N be independent observations from F_{θ_0} . The first term of the divergence $\Delta_\delta(F_\theta, F_{\theta_0})$ can be estimated by

$$\delta(\theta) = \frac{1}{N} \sum_{i=1}^N \delta(F_\theta, X^i),$$

minimization of which defines an unbiased and asymptotically normal estimator of θ_0 (Dawid et al., 2016, Theorem 2) under suitable regularity conditions; see Huber (1967), Barndorff-Nielsen and Cox (1994, Section 9.2), or Molenberghs and Verbeke (2005, Section 9.2.2). Consequently, for a risk functional r , the estimator

$$\hat{\theta}_{\delta,u}^r(X^1, \dots, X^N) = \arg \min_{\theta \in \Theta} \sum_{n=1}^N \mathbf{1} \left\{ r \left(\frac{X^n}{u} \right) > 1 \right\} \delta \left(\lambda_{\theta,u}^r, \frac{X^n}{u} \right), \quad (2.20)$$

is also consistent and asymptotically normal. Owing to de Haan and Resnick (1993, Propositions 3.1, 3.2), these asymptotic properties can be generalized to samples from a regularly-varying random vector with normalized marginals; see Appendix B.6 for the proof.

Proposition 2.1 *Let Y^1, \dots, Y^N be independent replicates of a regularly-varying random vector Y^* with normalized marginals and limiting measure Λ_{θ_0} and let δ be a strictly proper scoring rule. Under suitable regularity conditions, if $N \rightarrow \infty$ and $N_u \rightarrow \infty$ in such a way that $N_u/N \rightarrow 0$ as $N \rightarrow \infty$, then*

$$N_u^{1/2} \left\{ \hat{\theta}_{\delta, N/N_u}^r(Y^1, \dots, Y^N) - \theta_0 \right\} \rightarrow \mathcal{N} \left\{ 0, K^{-1} J (K^{-1})^T \right\}$$

in distribution, where

$$J = \mathbb{E} \left\{ \frac{\partial \delta(\theta)}{\partial \theta} \frac{\partial \delta(\theta)}{\partial \theta^T} \right\} \Bigg|_{\theta=\theta_0}, \quad K = \mathbb{E} \left\{ \frac{\partial^2 \delta(\theta)}{\partial \theta \partial \theta^T} \right\} \Bigg|_{\theta=\theta_0}. \quad (2.21)$$

Estimates of the Godambe information matrix $KJ^{-1}K$ can be used for inference, and the scoring-rule ratio statistic

$$V^\delta = 2 \left\{ \frac{\partial \delta(\theta_0)}{\partial \theta} - \frac{\partial \delta(\hat{\theta}_{\delta, N/N_u}^r)}{\partial \theta} \right\},$$

properly calibrated, can be used to compare nested models (Dawid et al., 2016, Section 4.1).

The log-likelihood is the proper scoring rule associated to the Kullback–Leibler di-

vergence. Although efficient, it is not robust with respect to model misspecifications, which is problematic for fitting asymptotic models like Pareto processes, and a closed form for the normalizing coefficient $\Lambda_\theta\{\mathcal{A}_r(u)\}$ defined in equation (2.14) is available only in special cases. The gradient scoring rule (Hyvärinen, 2005) uses only the derivative $\nabla_x \log \lambda_{\theta,u}^r$, and thus does not require computation of $\Lambda_\theta\{\mathcal{A}_r(u)\}$. Hyvärinen (2007) adapted this scoring rule for strictly positive variables, and we propose to extend it to any domain of the form $\mathcal{A}_r(u) = \{x \in \mathbb{R}_+^L : r(x/u) \geq 1\}$, using the divergence measure

$$\Delta_{\text{grad}}(\theta, \theta_0) = \int_{\mathcal{A}_r(u)} \|\nabla_x \log \lambda_\theta(x) \otimes w(x) - \nabla_x \log \lambda_{\theta_0}(x) \otimes w(x)\|_2^2 \lambda_{\theta_0}(x) dx, \quad (2.22)$$

where λ_θ is differentiable for all $\theta \in \Theta$ on $\mathcal{A}_r(u) \setminus \partial\mathcal{A}_r(u)$, $\partial\mathcal{A}$ denotes the boundary of \mathcal{A} , ∇_x is the gradient operator, $w : \mathcal{A}_r(u) \rightarrow \mathbb{R}_+^L$ is a positive weight function, and \otimes denotes the Hadamard product. If w is differentiable on $\mathcal{A}_r(u)$, and satisfies certain boundary conditions discussed in Appendix B.2, then the scoring rule

$$\delta_w(\lambda_\theta, x) = \sum_{l=1}^L \left(2w_l(x) \frac{\partial w_l(x)}{\partial x_l} \frac{\partial \log \lambda_\theta(x)}{\partial x_l} + w_l(x)^2 \left[\frac{\partial^2 \log \lambda_\theta(x)}{\partial x_l^2} + \frac{1}{2} \left\{ \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \right\}^2 \right] \right) \quad (2.23)$$

for $x \in \mathcal{A}_r(u)$ is strictly proper. The gradient score for a log-Gaussian Pareto process satisfies the regularity conditions for Theorem 2 in Dawid et al. (2016), so the resulting estimator $\hat{\theta}_w$ is asymptotically normal.

For the Brown–Resnick model, two possible weight functions are

$$w_l^1(x) = x_l [1 - e^{-\{r(x/u)-1\}}], \quad w_l^2(x) = \{1 - e^{-3(x_l-u_l)/u_l}\} [1 - e^{-\{r(x/u)-1\}}], \quad l = 1, \dots, L, \quad (2.24)$$

where r is a risk functional differentiable on \mathbb{R}_+^L and the threshold vector u lies in \mathbb{R}_+^L . The weights w^1 are derived from Hyvärinen (2007), whereas w^2 is designed to approximate the effect of censoring by down-weighting components of x^n near u . These weighting functions are well-suited for extremes: a vector $x \in \mathcal{A}_r(u)$ is penalized if $r(x/u)$ is close to 1, and low components of x induce low weights for the associated partial derivatives. For these reasons, inference using δ_w with the weighting functions in equation (2.24) should be more robust to low components than is the spectral estimator. The estimator $\hat{\theta}_w$ is much cheaper to compute than $\hat{\theta}_{\text{cens}}$ and can be obtained for any risk functional differentiable on \mathbb{R}_+^L . Expressions for the gradient score for the Brown–Resnick model can be found in Appendix B.3, and the performances of these inference procedures are compared in Section 2.4.

The gradient score can be applied to any extremal model with a multivariate density function whose logarithm is twice differentiable away from the boundaries of its

support, and if these display discontinuities on this support then a carefully chosen weighting function w ensures the existence and the consistency of the score. Indeed, similar expressions can be derived for the extremal- t model, though choices for the weight functions are more restricted: w^2 satisfies the boundary conditions, but w^1 does not ensure that the score is proper.

2.4 Simulation study

2.4.1 Exact simulation

The inference procedures and simulation algorithms described herein are wrapped in an R package, `mvPot` (de Fondeville, 2016) available on CRAN.

We first illustrate the feasibility of high-dimensional inference by simulating r -Pareto processes associated to log-normal random functions at L locations in $S = [0, 100]^2$. Details of the algorithm can be found in Appendix B.7.

We use an isotropic power semi-variogram, $\gamma(s, s') = (\|s - s'\|/\tau)^\alpha / 2$, shape parameters $\alpha = 0.5, 1, 1.3, 1.8$, and scale parameter $\tau = 2.5$, chosen such that the dependence models defined on S cover strong to weak extremal dependence. For this simulation, the dependence model with $\alpha = 1.8$ requires us to work on the log-scale to avoid rounding errors. For each simulation, $N = 10,000$ r -Pareto processes, with $r(x) = L^{-1} \sum_l x_l$, were simulated on regular 10×10 , 20×10 and 20×15 grids. The grid size was restricted to at most 300 locations for ease of comparison with the second simulation study. For the gradient score, we use $r(x) = L^{-1} \sum_l x(s_l)$. The components of the threshold vector u are taken equal to the empirical 0.99 quantile of $r(x^1), \dots, r(x^N)$, giving $N_u = 100$. For censored likelihood inference, we use the approach described in Appendix B.1.2 with $\bar{p} = 10$ and threshold u equal to the empirical 0.99 quantile of $\max_l x_l^1, \dots, \max_l x_l^N$, so that conditions for equation (2.7) are satisfied. One hundred replicates are used in each case.

Table 2.1 gives the relative root mean square error for estimation based on the censored loglikelihood and the gradient score with weights w^1 and w^2 , compared to that based on the spectral estimator. For all the methods and parameter combinations, bias is negligible and performance is mainly driven by the variance. As expected, efficiency is lower than 100% because when simulating and fitting from the true model, the spectral estimator performs best. The gradient score and censored likelihood estimators deteriorate as the extremal dependence weakens and the number of low components in the simulated vectors increases. The gradient score outperforms the censored likelihood except when censoring is low, i.e., when $\alpha = 0.5$. The performance of

Table 2.1 – Relative root mean square error (%) for comparison of estimates based on censored loglikelihood (left) and the gradient score with weights w^1 (middle) and w^2 (right) relative to spectral estimates, for the parameters α and $\tau = 2.5$. Efficiency of 100% would correspond to the spectral estimator, and smaller values to less efficient estimators. Inference is performed using the top 1% of 10,000 Pareto processes with semi-variogram $\gamma(s, s') = (\|s - s'\|/\tau)^\alpha / 2$ simulated on regular 10×10 , 20×10 and 20×15 grids.

Grid size	Shape α			
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.8$
10×10	53/46/44	10/32/33	4.7/39/39	1.0/51/52
20×10	67/51/52	10/25/24	5.4/34/35	1.0/54/55
20×15	67/47/47	11/30/31	4.1/25/25	1.4/49/49

Grid size	Scale τ			
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.8$
10×10	52/58/57	19/60/59	10/63/66	1.7/53/53
20×10	41/80/79	17/70/70	9.2/71/70	3.3/52/51
20×15	38/68/69	17/82/81	7.1/62/61	3.9/51/52

censored likelihood estimation deteriorates when L increases, suggesting that the gradient score will be preferable in high dimensions. These results does, however, not reflect real case studies, since the data are simulated from the model fitted, whereas in practice the model is used as a high-threshold approximation to the data distribution.

The optimization of the likelihood based on the spectral density and gradient score functions takes only a dozen seconds even for the finest grid. The same random starting point is used for each optimization to ensure fair comparison. Estimation using the censored approach takes several minutes and slows greatly as the dimension increases; see Appendix B.4.

2.4.2 Domain of attraction

As in practice the asymptotic regime is never reached, we now compare the robustness of each inference procedure for finite thresholds. The Brown–Resnick process belongs to its own max-domain of attraction, so its peaks-over-threshold distribution converges to a generalized Pareto process with log-Gaussian random function. We repeat the simulation study of Section 2.4.1 with 10,000 Brown–Resnick processes and

Chapter 2. High-dimensional peaks-over-threshold inference

Table 2.2 – As Table 2.1 but with inference based on the top 1% of 10,000 simulated Brown–Resnick processes. “NC” means that optimization does not converge.

Grid size	Shape α			
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.8$
10×10	154/111/81	473/183/108	196/170/105	NC
20×10	172/122/95	413/150/114	309/181/137	144/168/122
20×15	142/119/99	369/133/110	314/170/140	163/173/137

Grid size	Scale τ			
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.3$	$\alpha = 1.8$
10×10	107/127/116	263/38/35	109/231/452	NC
20×10	105/133/119	206/94/80	315/66/53	105/336/261
20×15	104/138/126	173/102/90	290/92/46	103/211/144

the same parameter values. This simulation uses the algorithm of Dombry et al. (2016) and is computationally expensive, so we used only 300 variables. It took around three hours using 16 cores to generate $N = 10,000$ samples on the finest grid.

Table 2.2 shows the results. As expected when the model is misspecified, the root relative mean square error is mainly driven by bias, which increases with the shape α and the dimension L . Spectral estimation is overall outperformed by both other methods. For $\alpha = 0.5$, the three methods show fairly similar overall performance, with the censored likelihood better capturing the shape parameter, whereas the gradient score does better for the scale. The moderate extremal dependence cases, with $\alpha = 1$ and 1.3, are dominated by the censored likelihood, whereas for weak extremal dependence, $\alpha = 1.8$, the gradient score performs best, because too much information is lost by censoring. For the 100-point grid, the optimization procedures do not converge when the extremal dependence is too weak. The choice of the weighting function w affects the robustness of the gradient score. Computation times are similar to those in 2.4.1.

Quantile-quantile plots show that the score-matching estimators are very close to normally distributed, but censored likelihood estimates can deviate somewhat from normality due to the quasi-Monte Carlo approximation; this can be remedied by increasing the value of p .

To summarise: for strong extremal dependence, the three types of estimator are roughly equivalent. For moderate extremal dependence, we recommend using the

censored likelihood if the number of variables permits, this is $L \lesssim 500$ with our computational capabilities, though if extremal independence is reached at far distances and the grid is dense, the gradient score is an excellent substitute. Owing to its robustness and lack of dimensionality limitations, the gradient score appears to be the best choice for gridded applications with fine resolution. Empirical work suggests that it can be robustified by careful design of the weight function.

2.5 Extreme rainfall over Florida

2.5.1 General

We fit an r -Pareto process based on the Brown–Resnick model to radar measurements of rainfall taken every 15 minutes during the wet season, June–September, from 1999 to 2004 on a regular 2 km grid in a 120 km \times 120 km region of east Florida; see Figure 2.2. There are 3,600 spatial observations in each radar image, and 70,272 images in all. The region was chosen to repeat the application of Buhl and Kluppelberg (2016), but in a spatial setting only; a spatio-temporal model is outside the scope of the present paper. Buhl and Kluppelberg (2016) analysed daily maxima for 10 km \times 10 km squares, but we use non-aggregated data to fit a non-separable parametric model for spatial extremal dependence, using single extreme events instead of daily maxima.

The marginal distributions for each grid cell were first locally transformed to unit Pareto using their empirical distribution functions. For general application, where we wish to extrapolate the distribution above observed intensities, a model for the marginal distributions of exceedances is needed, but since our goal here is to illustrate the feasibility of dependence model estimation on dense grids, we regard marginal modelling as outside the scope of this study.

2.5.2 Multivariate extremal dependence model

The spatial model of Buhl and Kluppelberg (2016) is fully separable, i.e., it is a sum of two separate semi-variograms. This has the advantage that inference for each direction can be performed separately, but it cannot capture anisotropy that does not follow the axis of the grid, i.e., is not in the South-North or East-West directions. Furthermore their pairwise likelihood approach focuses on short-distance pairs, and so might mis-estimate dependence at longer distances. To better capture possible

Chapter 2. High-dimensional peaks-over-threshold inference

anisotropy, we use the non-separable semi-variogram

$$\gamma(s_i, s_j) = \left\| \frac{\Omega(s_i - s_j)}{\tau} \right\|^\alpha, \quad s_i, s_j \in [0, 120]^2, \quad i, j \in \{1, \dots, 3600\}, \quad 0 < \alpha \leq 2, \tau > 0, \quad (2.25)$$

and anisotropy matrix

$$\Omega = \begin{bmatrix} \cos \eta & -\sin \eta \\ a \sin \eta & a \cos \eta \end{bmatrix}, \quad \eta \in \left(-\frac{\pi}{4}, \frac{\pi}{4}\right], \quad a > 0. \quad (2.26)$$

The semi-variogram γ achieves asymptotic extremal independence as the distance between sites tends to infinity, i.e., the pairwise extremal index increases to 2 as $\|s - s'\| \rightarrow \infty$.

To apply peaks-over-threshold methodology, we must define exceedances by choosing risk functionals. We focus on two types of extremes: local very intense rainfall at any point of the region, and high cumulative rainfall over the whole region. We therefore take the risk functionals

$$r_{\max}(X^*) = \left\{ \sum_{l=1}^{3600} X^*(s_l)^{20} \right\}^{1/20}, \quad r_{\text{sum}}(X^*) = \left\{ \sum_{l=1}^{3600} X^*(s_l)^{\xi_0} \right\}^{1/\xi_0}.$$

The function r_{\max} is a differentiable approximation to $\max_d X(s_d)$, which cannot be used with the gradient score because of its non-differentiability. Censored likelihood is computationally out of reach with so many locations. Directly summing normalized observations X^* makes no physical sense, so our function r_{sum} , which selects extreme events with large spatial extent, attempts to transform the data back to the original scale; we take $\xi_0 = 0.114$, which is the average of independent local estimates of a generalized Pareto distribution.

We fitted univariate generalized Pareto distributions to $r_{\text{sum}}(x^{*n})$ and $r_{\max}(x^{*n})$ ($n = 1, \dots, 70, 272$) with increasing thresholds. The estimated shape parameters are stable around the 99.9 percentile, which we used for event selection, giving 59 exceedances; just two events were found to be extreme relative to both risk functionals. This threshold may appear rather high, but it corresponds to around 10 events per year, which seems reasonable in light of the time-frame. Here we merely illustrate the feasibility of high-dimensional inference, so we treat them as independent, but in practice temporal declustering should be considered.

Optimization of the gradient score with the w^1 weighting function on a 16-core cluster took from 1 to 6 hours, depending on the initial point. Different initial points must be considered because of the possibility of local maxima. Results are shown in Table 2.3,

Table 2.3 – Parameter estimates (standard errors) for an r -Pareto process derived from log-Gaussian random functions with the semi-variogram $\gamma(s, s') = \{\|\Omega(s - s')\|/\tau\}^\alpha$ obtained by maximization of the gradient score for events corresponding to 59 highest exceedances of the risk functionals r_{sum} and r_{max} for the Florida radar rainfall data. Standard errors are obtained using a jackknife with 20 blocks.

Risk functional	α	τ	η	a
r_{sum}	0.814 (0.036)	25.63 (4.70)	-0.009 (0.458)	1.059 (0.031)
r_{max}	0.955 (0.048)	3.540 (0.67)	-0.316 (0.410)	0.940 (0.029)

where standard deviations are obtained using a jackknife procedure with 20 blocks. Both the estimated bias and variance are fairly low. For $r_{\text{sum}}(X^*)$, we obtain a model similar to that of Buhl and Kluppelberg (2016).

The estimated parameters differ appreciably for the two risk functionals, suggesting the presence of a mixture of types of extreme events. The structure for r_{max} is consistent with the database, in which the most intense events tend to be spatially concentrated. Our model suggests higher dependence for middle distances than was found by Buhl and Kluppelberg (2016), but they note that their model underestimates dependence, especially for high quantiles. The estimated smoothness parameters are very close, and neither estimate of η differs significantly from zero, as imposed by Buhl and Kluppelberg (2016). For r_{sum} , the estimated parameters show strong extremal dependence even at long distances, corresponding to exceedances of cumulated rainfall with large spatial cover. As $\hat{a} \approx 1$, there is no evidence of anisotropy.

2.5.3 Model checking and simulation

For model checking, we propose to use the conditional exceedance probability,

$$\pi_{ij} = \Pr[X^*(s_j) \geq u' \mid \{X^*(s_i) \geq u'\} \cap \{r(X^*/u) \geq 1\}] = 2 \left[1 - \Phi \left\{ \left(\frac{\gamma_{ij}}{2} \right)^{1/2} \right\} \right], \quad (2.27)$$

where $\gamma_{i,j}$ is the semi-variogram for sites s_i and s_j ($i, j = 1, \dots, 3600$), as defined in (2.11) and $u' > 0$. An empirical estimator of π_{ij} is

$$\hat{\pi}_{ij} = \frac{\sum_{n=1}^N \mathbf{1} \left[\{r(x^{*n}/u) \geq 1\} \cap \{x_i^{*n} \geq u'\} \cap \{x_j^{*n} \geq u'\} \right]}{\sum_{n=1}^N \mathbf{1} \left[\{r(x^{*n}/u) \geq 1\} \cap \{x_i^{*n} \geq u'\} \right]}, \quad (2.28)$$

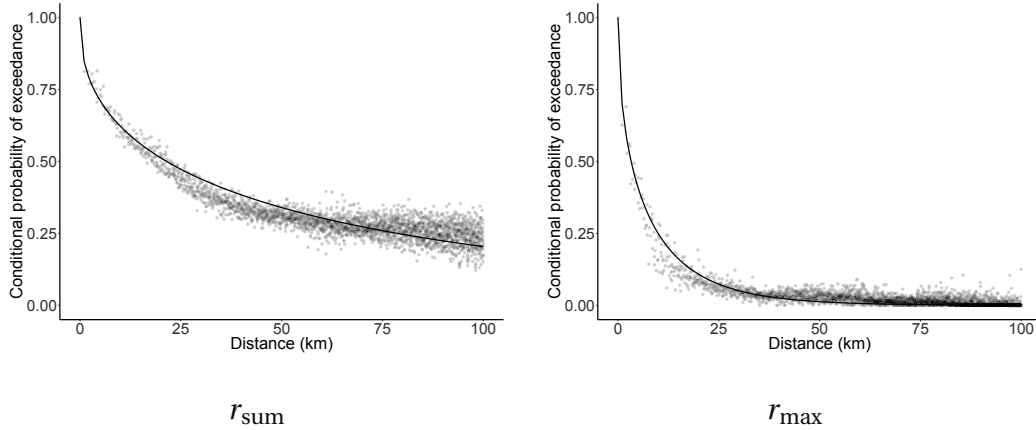


Figure 2.1 – Estimated conditional exceedance probability π_{ij} for the risk functional r_{sum} (left) and r_{max} (right) as a function of the distance between locations s_i and s_j , $i, j = 1, \dots, 3600$. The solid black line represents the model fitted using gradient score estimation.

whose asymptotic behaviour derives from Davis and Mikosch (2009). For both risk functionals, the fitted model, represented by the solid black lines in Figure 2.1, follows the cloud of estimated conditional exceedance probabilities reasonably well and captures the general trend, but fails to represent some local variation, perhaps due to a lack of flexibility of the power model.

Finally, we use the models fitted in Section 2.5.2 to simulate events with intensities equivalent to the weakest of the 59 events found by our risk functionals. Simulation is performed by generating the corresponding r -Pareto process with the fitted dependence structure, as in Section 2.4.1. Figure 2.2 compares observations from the database and representative simulations, which seem to successfully reproduce both the spatial dependence and the intensity of the selected observations. A closer examination suggests that in both cases the models produce over-smooth rainfall fields. This could be addressed by improving event selection using risk functionals r that characterize special spatial structures or physical processes. Although we fail to detect anisotropy, more complex models for dependence that allow stochasticity of the spatial patterns might be worthwhile.

These models can reproduce both spatial patterns and extreme intensity for spatially accumulated and local heavy rainfall. In both cases the fitted dependence model provides a reasonable fit and simulations seem broadly consistent with observations. However, the presence of two contrasting dependence structures highlights the complexity of extreme rainfall and suggests that a mixture model for both dependence and margins might be considered. Marginal and dependence parameters are often

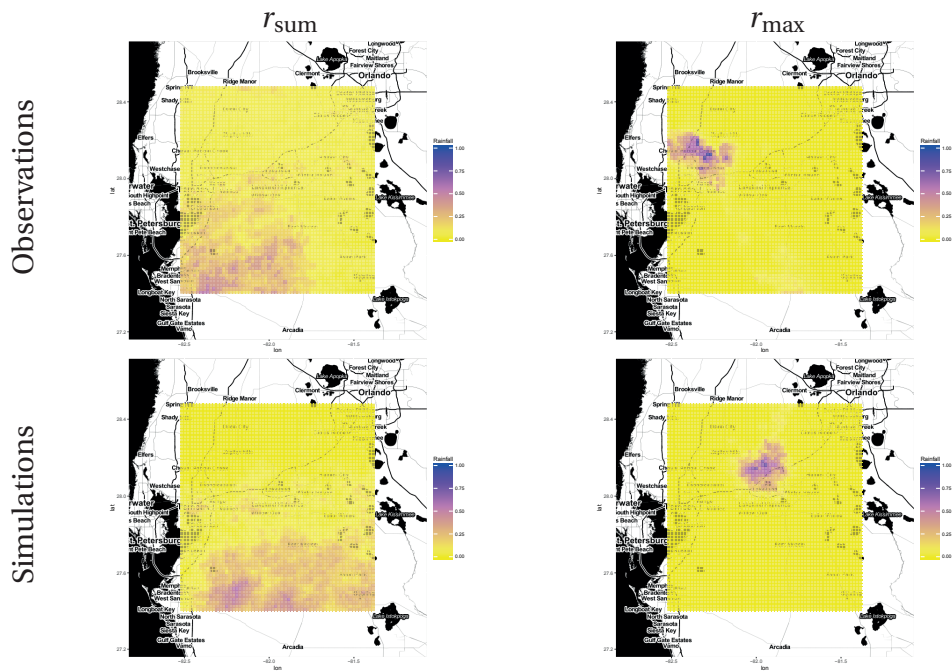


Figure 2.2 – Fifteen-minute cumulated rainfall (inches), observed (first row) and simulated (second row) for the risk functionals r_{sum} (left) and r_{max} (right) with an intensity equivalent to the 59th most intense event.

estimated separately, but with the presence of mixtures, which can be detected using different risk functionals, joint estimation is required, which is outside the scope of this paper. For this reason and because we neglected the temporal dependence, our model is merely a first step towards a spatio-temporal rainfall generator.

3 Extremal behaviour of aggregated data

This chapter is a postprint version of the article ‘Extremal Behaviour of Aggregated Data with an Application to Downscaling’ written with Sebastian Engelke and Marco Oesting. This paper has been accepted for publication in *Biometrika* (Engelke et al., 2019). Some parts of this chapter may be redundant with other sections of the thesis, but we decided to keep the chapter self-standing for clarity. Only light modifications were performed for consistency with the rest of the thesis.

3.1 Introduction

Spatial extreme value models and, especially, max-stable processes, are widely applied to assess risks in environmental science. These processes are motivated by the study of

$$M_n(s) = \max_{i=1, \dots, n} \frac{X_i(s) - b_s(n)}{a_s(n)}, \quad s \in S, \quad (3.1)$$

where X_1, \dots, X_n are independent observations of a continuous process X , modelling a phenomenon of interest such as rainfall or temperature on some region S . The scaling functions $a_s(n) > 0$ and $b_s(n) \in \mathbb{R}$, $n \in \mathbb{N}$, are both continuous in $s \in S$. Functional limits obtained from this construction as $n \rightarrow \infty$, named max-stable processes, are appealing models for spatial extremes. Their realizations, however, are composed of different single events X_i , which prohibits direct interpretation and renders efficient inference and simulation challenging (e.g., Dombry et al., 2016; Thibaud et al., 2016).

It is often more natural to study threshold exceedances, or, more precisely, the extremal behaviour of $r(X_i)$ ($i = 1, \dots, n$), where r is a functional on the space of continuous functions on S . For instance, Buishand et al. (2008) consider the daily rainfall over a certain region S , and therefore choose $r(X) = \int_S X(s) ds$. Using the same functional, Coles and Tawn (1996) relate the tail of the distribution of the integral to the tail of the

Chapter 3. Extremal behaviour of aggregated data

distribution at a single location, and Ferreira et al. (2012) formalize this idea through the so-called reduction factor. For general homogeneous functionals r , Dombry and Ribatet (2015) characterize the functional limits of threshold exceedances $u^{-1}X$ conditional on $r(X) > u$, for a high threshold u .

In this paper we follow Coles and Tawn (1996) and Ferreira et al. (2012) and investigate the tail behaviour of more general functionals r . Under certain conditions, we show that for sufficiently large n

$$\Pr \left[\frac{r(X) - r\{b_s(n)\}}{r\{a_s(n)\}} > x \right] \approx \theta^r \Pr \left\{ \frac{X(s_0) - b_{s_0}(n)}{a_{s_0}(n)} > x \right\}, \quad x \in \mathbb{R}, \quad s_0 \in S, \quad \theta^r > 0. \quad (3.2)$$

Thus, the tail of the r -functional of X behaves like the tail at an individual location times a reduction factor θ^r , which we call the r -extremal coefficient. In different contexts, the interpretation of θ^r might differ, but it summarizes the effect of spatial extremal dependence in X on the risk diversification through the functional r .

The r -extremal coefficient relates the tail of the univariate random variable $r(X)$ to the multivariate or spatial extremal dependence in X . This functional perspective has the advantage of producing return level estimates that are consistent with respect to the underlying structure of X , even when considering different aggregation functionals applied to X . Indeed, for functionals r_1, \dots, r_L , we study the multivariate tail behaviour of $\{r_1(X), \dots, r_L(X)\}$, which turns out to be in the max-domain of attraction of a multivariate max-stable distribution.

Popular models for the functional limit of the maxima M_n in (3.1) are Brown–Resnick processes, which take a role in spatial extremes similar to Gaussian processes in classical geostatistics. The reason for this is that the former are essentially the only such limits when X is a stationary Gaussian process and an additional rescaling is allowed (Kablichko et al., 2009). This connection can be exploited to perform efficient inference (Wadsworth and Tawn, 2014; Engelke et al., 2015; Thibaud and Opitz, 2015) and simulation (Dombry et al., 2013, 2016; Oesting et al., 2017a) for Brown–Resnick processes based on densities and sampling algorithms of Gaussian random vectors. In our framework, this link to Gaussian distributions allows us to use results from the geostatistical literature on data aggregation (e.g., Wackernagel, 2003) to obtain explicit expressions for θ^r and the extremal dependence in $\{r_1(X), \dots, r_L(X)\}$ if the limiting process Z in (3.1) is Brown–Resnick with Gumbel margins.

An important consequence is that our findings allow us to recover the tail distribution of X based only on information from the aggregated vector. This is similar to inferring the extremal dependence of X based only on extremal coefficients (Schlather and

Tawn, 2003). In meteorology, for instance, large-scale climate models provide only data over grid cells, but practical questions require risk assessment at point locations such as cities. Techniques to perform this transition from large to small scales are summarized under the notion of downscaling. In the second part of the paper we propose a statistical downscaling method to infer the tail behaviour of the underlying stochastic process X in a spatially consistent way based on aggregated data. Relevant outputs are the exceedance probabilities at point locations and simulations of spatial extreme events of X , both unconditionally and conditionally on the observed aggregated extremes. We apply this procedure to coarse-scale gridded temperature data in the south of France from the e-obs data set (Haylock et al., 2008). The fitted model provides fine-resolution simulations of the warmest day during the 2003 heatwave, conditionally on the observed grid values.

3.2 Limit results for extremes of aggregated data

3.2.1 Background on extremes

Let S be a compact subset of a complete separable metric space, and let $C(S)$ denote the space of real-valued continuous functions on S equipped with the supremum norm $\|\cdot\|_\infty$, defined by $\|x\|_\infty = \sup_{s \in S} |x(s)|$, and the corresponding Borel σ -algebra $C(S)$.

We consider a continuous stochastic process $\{X(s) : s \in S\}$, which we assume to be in the max-domain of attraction of a max-stable process with common extreme value index $\xi \in \mathbb{R}$. More precisely, for independent copies X_1, \dots, X_n of X , there exist functions $a_s : (0, \infty) \rightarrow (0, \infty)$ and $b_s : (0, \infty) \rightarrow \mathbb{R}$, both continuous in $s \in S$, such that as $n \rightarrow \infty$, the process M_n of componentwise maxima defined in (3.1) converges in distribution on the space $C(S)$, i.e.,

$$\mathcal{L} \left\{ \max_{i=1, \dots, n} \frac{X_i(s) - b_s(n)}{a_s(n)}, s \in S \right\} \longrightarrow \begin{cases} \mathcal{L}\{\text{sgn}(\xi)Z(s)^\xi, s \in S\}, & \xi \neq 0, \\ \mathcal{L}\{\log Z(s), s \in S\}, & \xi = 0, \end{cases} \quad (3.3)$$

where $\mathcal{L}(\eta)$ denotes the law of a process η . By definition, the process Z in the limit is max-stable, and it is simple in the sense that it has unit Fréchet margins (de Haan and Ferreira, 2006, Chapter 9). Moreover, for any $s \in S$, the margin $X(s)$ is in the max-domain of attraction of an extreme value distribution

$$G_\xi(x) = \begin{cases} \exp[-\{\text{sgn}(\xi)x\}^{-1/\xi}], & \xi \neq 0, \\ \exp\{-\exp(-x)\}, & \xi = 0, \end{cases} \quad (3.4)$$

Chapter 3. Extremal behaviour of aggregated data

for all $x \in \mathbb{R}$ with $x\xi \geq 0$. The different distributions are called $(1/\xi)$ -Fréchet for $\xi > 0$, Gumbel for $\xi = 0$ and $(-1/\xi)$ -Weibull for $\xi < 0$. The assumption of a spatially constant ξ in (3.3) is usually reasonable in applications and common in the literature since it is required to obtain meaningful theoretical results: when ξ is allowed to vary over space, the asymptotic distribution of $r(X)$ is driven only by the location(s) in S with the largest tail index.

According to its spectral representation (cf., de Haan, 1984; Giné et al., 1990; Penrose, 1992),

$$Z(s) = \max_{i \in \mathbb{N}} U_i W_i(s), \quad s \in S, \quad (3.5)$$

where $\{U_i : i \in \mathbb{N}\}$ are the points of a Poisson point process on $(0, \infty)$ with intensity measure $u^{-2} du$ and the spectral functions W_i ($i \in \mathbb{N}$) are independent copies of some non-negative, continuous process $\{W(s), s \in S\}$ with $E\{W(s)\} = 1$ for all $s \in S$.

Below we assume that in the Fréchet case $X(s)$ possesses a finite lower endpoint $x_*(s) > -\infty$ ($s \in S$), and, due to the continuity of X , the infimum $\inf_{s \in S} X(s)$ is bounded from below on the compact domain S .

Example 3.1 *Let S be a compact subset of \mathbb{R}^D , $D \geq 1$ and let $\{G(s) : s \in S\}$ be a centred Gaussian process with variogram $\gamma(s, s') = \text{var}\{G(s) - G(s')\}$. A Brown–Resnick process is the max-stable process Z in (3.5) where the spectral functions follow the distribution of*

$$W(s) = \exp [G(s) - \text{var}\{G(s)\}/2], \quad s \in S.$$

The distribution of Z only depends on the variogram γ , and for $s_1, \dots, s_m \in S$, the finite dimensional distribution of $\{Z(s_1), \dots, Z(s_m)\}$ is called the Hüsler–Reiss distribution (Hüsler and Reiss, 1989) with parameter matrix $\Gamma = \{\gamma(s_j, s_k)\}_{j,k=1,\dots,m}$; more details can be found in Appendix C.2.1 and in Brown and Resnick (1977), Kabluchko et al. (2009) and Kabluchko (2011).

3.2.2 Univariate limiting distributions of aggregated data

We first derive the univariate asymptotic distribution of aggregated data. Following Ferreira et al. (2012), we assume that the normalizing functions $a_s(t)$ can be decomposed asymptotically into positive functions $A(s)$ and $a(t)$ in the sense that

$$\sup_{s \in S} \left| \frac{a_s(t)}{a(t)} - A(s) \right| = \begin{cases} o(1), & \xi \neq 0, \\ 0, & \xi = 0, \end{cases} \quad (3.6)$$

3.2. Limit results for extremes of aggregated data

where $o(1) \rightarrow 0$ as $t \rightarrow \infty$. For data aggregation, we consider a positive homogeneous functional $r : C(S) \rightarrow \mathbb{R}$, i.e., r satisfies $r(ax) = ar(x)$ for all $a > 0$, $x \in C(S)$. We further assume that r is uniformly continuous and monotone, i.e., $r(x) \leq r(x')$ if $x(s) \leq x'(s)$ for all $s \in S$.

The following theorem is a particular case of Theorem 3.2. Alternatively it can be proved similarly to Ferreira et al. (2012, Theorem 2.1).

Theorem 3.1 *Let r be a positive homogeneous, monotone and uniformly continuous functional on $C(S)$, which for $\xi \leq 0$ is assumed to be linear. If (3.3) and (3.6) hold, then*

$$\lim_{t \rightarrow \infty} t \Pr \left[\frac{r(X) - r\{b_s(t)\}}{a(t)r(A)} > x \right] = \begin{cases} \theta_\xi^r |x|^{-1/\xi}, & \xi \neq 0, \\ \theta_0^r \exp(-x), & \xi = 0, \end{cases} \quad x \in \mathbb{R}, x\xi \geq 0, \quad (3.7)$$

where for $\xi \neq 0$ and $\xi = 0$

$$\theta_\xi^r = E \left[\left\{ \frac{r(AW^\xi)}{r(A)} \right\}^{1/\xi} \right], \quad \theta_0^r = E \left[\exp \left\{ \frac{r(A \log W)}{r(A)} \right\} \right], \quad (3.8)$$

respectively. For $\xi > 0$ we may always choose $b_s(t) \equiv 0$, and for $\xi < 0$ we can use $b_s(t) = x^*(s)$, where $x^*(s)$ is the upper endpoint of the distribution of $X(s)$ ($s \in S$).

Remark 3.1 *Theorem 3.1 is formulated for threshold exceedances, but can be reformulated to describe the limiting behaviour of $\max_{i=1}^n r(X_i)$, for independent copies X_1, \dots, X_n of X .*

Remark 3.2 *For $\xi \leq 0$, the functions W^ξ and $\log W$ may take the value $-\infty$ if W is not strictly positive. The terms in (3.8) then contain expressions of the type $r(x)$ for continuous functions $x : S \rightarrow \mathbb{R} \cup \{-\infty\}$, which we interpret as $r(x) = \inf_{x' > x} r(x')$. If this value is $-\infty$, the expression inside the expectations in (3.8) is 0, i.e., θ_ξ^r is not necessarily positive.*

We call the quantity θ_ξ^r the r -extremal coefficient, as it describes the change of the upper tail of the r -aggregated data compared to the tail of the univariate marginal data. Our definition of θ_ξ^r in Theorem 3.1 contains a normalization by $r(A)$, making it invariant under multiplication of r by a constant and thus simplifying interpretation. Indeed,

$$\theta_\xi^r = \lim_{u \rightarrow \infty} \frac{\Pr \{r(X)/r(A) > u\}}{\Pr \{X(s_0)/A(s_0) > u\}}, \quad \xi > 0, s_0 \in S.$$

Chapter 3. Extremal behaviour of aggregated data

In general, θ_ξ^r summarizes the effect of the spatial extremal dependence in X on the diversification of risk through the functional r . Both the dependence and the marginal tail index ξ affect the coefficient θ_ξ^r , which we stress in Theorem 3.1 and henceforth through the index ξ .

The concept of the r -extremal coefficient extends and unifies various notions in extreme value statistics and applied sciences such as extremal coefficients, diversification factors in portfolios and areal reduction factors. We present these and other examples for illustration, always assuming that X satisfies the conditions of Theorem 3.1.

Example 3.2 *The important case where $S \subset \mathbb{R}^2$ is a compact region and*

$$r(x) = |S|^{-1} \int_S x(s) ds$$

was first studied in Coles and Tawn (1996) and Buishand et al. (2008) in the framework of total areal rainfall, and then formalized by Ferreira et al. (2012). In this case of a spatial average, the coefficient $\theta_\xi^r = \theta_\xi^{avg}$ is popular in environmental science, where it is called the areal reduction factor. Hydrologists use it to convert quantiles of point rainfall to quantiles of total rainfall over a river catchment. If the spectral functions W are almost surely strictly positive then this coefficient satisfies $0 < \theta_\xi^{avg} \leq 1$ for $\xi \leq 1$, and $\theta_\xi^{avg} \geq 1$ for $\xi \geq 1$ (Ferreira et al., 2012, Prop. 2.2), so average rainfall is less extreme than point rainfall if the marginal distributions have finite expectation, as typically encountered in practice, and more extreme if they have infinite expectation.

Example 3.3 *If $S = \{s_1, \dots, s_m\}$ is a finite set and $r(x) = \sum_{i=1}^m c_i x(s_i)$ is a weighted sum with fixed $c_1, \dots, c_m \geq 0$, then Zhou (2010) and Mainik and Embrechts (2013) computed the corresponding coefficient θ_ξ^r for $\xi > 0$. In this setup, $X(s_i)$ ($i = 1, \dots, m$) are interpreted as dependent, heavy-tailed risk factors, and θ_ξ^r represents the diversification in the portfolio $P = \sum_{i=1}^m c_i X(s_i)$. More precisely, the value-at-risk of P for high levels $\alpha \rightarrow 1$ can be expressed as the value-at-risk of a single factor times a constant that involves the r -extremal coefficient θ_ξ^r . Theorem 3.1 yields an analogous result also for light-tailed risk factors.*

Example 3.4 *Another well-known example is the case of $S = \{s_1, \dots, s_m\}$ being a finite set and $r(x) = \max_{i=1}^m x(s_i)$. If $A(s_1) = \dots = A(s_m)$, then $\theta_\xi^r = E \{\max_{i=1}^m W(s_i)\}$ corresponds to the classical extremal coefficient (Schlather and Tawn, 2003), a number between 1 and m that is usually interpreted as the number of asymptotically independent random variables among $X(s_1), \dots, X(s_m)$. A similar interpretation applies if S*

3.2. Limit results for extremes of aggregated data

is an arbitrary compact subset, and $\theta_\xi^r = E\{\max_{s \in S} W(s)\}$ is a spatial extension of the classical extremal coefficient.

Example 3.5 As a last example, we consider energy functionals of $r(x) = \{\int_S x^2(s) ds\}^{1/2}$, for $x \geq 0$, which appear in various applications in physics. In the case of X being a wind field, $r^2(X)$ represents the integrated kinetic energy over a region S , which is an indicator for the potential damage caused by the corresponding storm event (e.g., Powell and Reinhold, 2007).

The expressions in (3.8) for the r -extremal coefficient are expected values of functions of the spectral process W . The distribution of the latter is known for most popular models, and it includes truncated Gaussian processes (Schlather, 2002; Opitz, 2013a) and log-Gaussian processes (Brown and Resnick, 1977; Kabluchko et al., 2009), for instance. Numerical evaluation of θ_ξ^r is thus readily implemented through simulation of W . In the important case of $\xi = 0$ and W corresponding to a log-Gaussian process, we obtain a closed-form expression for θ_0^{avg} .

Example 3.6 Suppose that $\xi = 0$ and Z is a Brown–Resnick process on a compact set $S \subset \mathbb{R}^D$, as introduced in Example 3.1. The extremal coefficient of the spatial average then is

$$\log \theta_0^{avg} = -\frac{\int_S \int_S A(s)A(t)\gamma(s,t)dsdt}{4\{\int_S A(s)ds\}^2}. \quad (3.9)$$

Let $D = 1$ and $S = [0, T]$, $T > 0$, and let $A \equiv 1$ be constant over S . For the popular power variogram model $\gamma(s, t) = |(s - t)/\tau|^\alpha$ with $\alpha \in (0, 2]$, $\tau > 0$, we obtain

$$\theta_0^{avg} = \exp\left\{-\frac{T^\alpha}{2\tau^\alpha(\alpha+1)(\alpha+2)}\right\}.$$

This expression tends to zero if the length of the domain $T \rightarrow \infty$, meaning that the distribution of the average eventually has a much lighter tail than the marginal distributions. This strong diversification effect can be explained by the fact that the Brown–Resnick process with power variogram is mixing (cf., Kabluchko and Schlather, 2010).

3.2.3 Multivariate limiting distributions of aggregated data

In the previous section we derived the univariate tail distribution of data aggregated through a functional r . In applications we often observe data through several different functionals, e.g., the integrals over not necessarily disjoint areas. The consistency of return level estimates discussed in the introduction has even more important

Chapter 3. Extremal behaviour of aggregated data

implications when different risk functionals are applied to the data. The univariate tail of each aggregation could be estimated separately, but the dependence between the tails would not be captured. We therefore consider arbitrary positive homogeneous, uniformly continuous functionals $r_1, \dots, r_L : C(S) \rightarrow \mathbb{R}$, and we aim to describe the multivariate tail behaviour of the vector $(r_1(X), \dots, r_L(X))$. The proof of the following theorem is given in Appendix C.1.

Theorem 3.2 *Let r_1, \dots, r_L be a positive homogeneous, monotone and uniformly continuous functionals on $C(S)$, which for $\xi \leq 0$ are assumed to be linear. If (3.3) and (3.6) hold, then for $\xi \neq 0$ and all $x_i \in \mathbb{R}$ with $\xi x_i > 0$ ($i = 1, \dots, L$)*

$$\lim_{t \rightarrow \infty} t \Pr \left(\bigcup_{j=1}^L \left[\frac{r_j(X) - r_j\{b.(t)\}}{a(t)r_j(A)} > x_j \right] \right) = \mathbb{E} \left[\bigvee_{j=1}^L \left\{ \frac{r_j(AW^\xi)}{|x_j|r_j(A)} \right\}^{1/\xi} \right], \quad (3.10)$$

and for $\xi = 0$ and $x_1, \dots, x_L \in \mathbb{R}$,

$$\lim_{t \rightarrow \infty} t \Pr \left(\bigcup_{j=1}^L \left[\frac{r_j(X) - r_j\{b.(t)\}}{a(t)r_j(A)} > x_j \right] \right) = \mathbb{E} \left[\bigvee_{j=1}^L \exp \left\{ -x_j + \frac{r_j(A \log W)}{r_j(A)} \right\} \right]. \quad (3.11)$$

Theorem 3.2 states that the vector $\{r_1(X), \dots, r_L(X)\}$ of aggregations is in the max-domain of attraction of the multivariate max-stable distribution with exponent measure given by the right-hand side of (3.10) or (3.11), respectively. For the j th margin, for $\xi \neq 0$, the scale of the Weibull or Fréchet distribution is $(\theta_\xi^{r_j})^\xi$, and for $\xi = 0$ the location parameter of the Gumbel distribution is $\log \theta_0^{r_j}$. This recovers the univariate results in Theorem 3.1. For details on multivariate domains of attraction and exponent measures, see Resnick (1987, Chapter 5). In general this max-stable distribution is not available in closed form, but for the purpose of evaluating risk regions for $\{r_1(X), \dots, r_L(X)\}$, it can be approximated by Monte Carlo methods. In the following important special case, we can compute the multivariate distribution explicitly.

Example 3.7 *Consider the same framework as in Example 3.6, namely $S \subset \mathbb{R}^D$ compact, $\xi = 0$ and X in the max-domain of attraction of Brown–Resnick process with spectral functions W . Suppose that for all $j = 1, \dots, L$, the functional r_j is the spatial average over the compact region $A_j \subset S$, respectively. Since W is log-Gaussian in this case, the random vector*

$$\{r_1(A \log W)/r_1(A), \dots, r_L(A \log W)/r_L(A)\} \quad (3.12)$$

is multivariate Gaussian, and its variogram matrix $\Gamma \in \mathbb{R}^{L \times L}$ can be computed explicitly; see Appendix C.2.2. The exponent measure in (3.11) therefore corresponds to a L -variate Hüsler–Reiss distribution with dependence matrix Γ whose j th margin has a Gumbel distribution with location parameter $\log \theta_0^{r_j}$ given in (3.9).

3.3 Statistical Inference

3.3.1 Setting

Suppose we observe independent data X_1, \dots, X_n ($n \in \mathbb{N}$) of the process $X = \{X(s) : s \in S\}$, but only through the aggregation functionals r_j satisfying the conditions of Theorem 3.2. The observations are therefore L -dimensional and of the form

$$\{r_1(X_i), \dots, r_L(X_i)\}, \quad i = 1, \dots, n.$$

We aim to use Theorems 3.1 and 3.2 to infer the extremal behaviour of the whole process from the observed aggregated data. This requires estimation of both the marginal tail behaviour and the extremal dependence of X .

We suppose that the process X is in the functional max-domain of attraction of a max-stable process Z as in (3.3) with marginal distributions of $Z(s)$ of the form (3.4) for all $s \in S$. A natural and fairly general assumption is that the marginal distributions of X belong to a location-scale family, i.e., for some distribution function F and continuous $A : S \rightarrow (0, \infty)$, $B : S \rightarrow \mathbb{R}$,

$$\Pr\{X(s) \leq x\} = F \left\{ \frac{x - B(s)}{A(s)} \right\}, \quad s \in S.$$

Since $X(s)$ lies in the max-domain of attraction of $Z(s)$, the distribution of $M_n(s)$ must converge to G_ξ as $n \rightarrow \infty$. In particular, F must satisfy $\lim_{t \rightarrow \infty} F^t\{a(t)x + b(t)\} = G_\xi(x)$ for all $x \in \mathbb{R}$ with $\xi x \geq 0$ and appropriate functions $a : (0, \infty) \rightarrow (0, \infty)$ and $b : (0, \infty) \rightarrow \mathbb{R}$. This implies that the normalizing functions a_s and b_s of $X(s)$ can be chosen as

$$a_s(t) = A(s)a(t), \quad b_s(t) = B(s) + A(s)b(t), \quad t \in \mathbb{R}. \quad (3.13)$$

Moreover, if $\xi \neq 0$, without loss of generality, we may assume $b(t) \equiv 0$ does not depend on t by the arguments in the proof of Theorem 3.2. For $\xi > 0$, we may further assume that $B(s) \equiv 0$, as a shift in location does not affect the asymptotic behaviour of the process, while, for $\xi < 0$, $B(s)$ can be assumed to be the possibly unknown upper end point $x^*(s)$ of the distribution of $X(s)$.

We impose a parametric structure on the marginal scale and location parameters, i.e., the unknown functions A and B , and the extremal dependence of X , which is given by the exponent measure of Z . For the marginal distributions, we assume that A and B belong to parametric families of functions $\{A_{\vartheta_A}, \vartheta_A \in \Theta_A\}$ and $\{B_{\vartheta_B}, \vartheta_B \in \Theta_B\}$ where Θ_A and Θ_B are subsets of \mathbb{R}^{k_A} and \mathbb{R}^{k_B} . For the dependence, we suppose that the probability measure \mathbb{P}^{spec} induced by the spectral function W of the limiting max-stable process Z belongs to a parametric class $\{\mathbb{P}_{\vartheta_W}^{\text{spec}}, \vartheta_W \in \Theta_W\}$ with $\vartheta_W \subset \mathbb{R}^{k_W}$. Further, the extreme value index $\xi \in \mathbb{R}$ and the joint normalization constants $a(t) \in (0, \infty)$ and $b(t) \in \mathbb{R}$ must be estimated for some large t . We present two ways to estimate the parameter vector $\vartheta = \{\xi, a(t), b(t), \vartheta_A, \vartheta_B, \vartheta_W\}$ based on the marginal and multivariate tail behaviour of $\{r_1(X), \dots, r_L(X)\}$ in Theorems 3.1 and 3.2, respectively.

3.3.2 Fitting based on marginal estimates

As a first approach, we approximate the tail of the distribution of $r_j(X)$ separately for each $j = 1, \dots, L$. From an equivalent formulation of (3.7) for maxima over blocks with sufficiently large length t , we obtain

$$\mathbb{E}\{r_j(X) \leq x\}^t \approx \begin{cases} \exp\left\{-\left|\frac{x - \mu_{j,t}}{\sigma_{j,t}}\right|^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{x - \mu_{j,t}}{\sigma_{j,t}}\right)\right\}, & \xi = 0, \end{cases} \quad (3.14)$$

where the location parameters $\mu_{j,t}$ and the scale parameters $\sigma_{j,t}$ ($j = 1, \dots, L$), are given by

$$\mu_{j,t} = \begin{cases} r_j(B_{\vartheta_B}), & \xi \neq 0, \\ r_j(A_{\vartheta_A}) \left\{b(t) + a(t) \log \theta_0^{r_j}\right\} + r_j(B_{\vartheta_B}), & \xi = 0, \end{cases} \quad (3.15)$$

$$\sigma_{j,t} = \begin{cases} (\theta_\xi^{r_j})^\xi a(t) r_j(A_{\vartheta_A}), & \xi \neq 0, \\ a(t) r_j(A_{\vartheta_A}), & \xi = 0, \end{cases} \quad (3.16)$$

where $\theta_\xi^{r_j}$ is defined in (3.8) and depends on ϑ . Analogously, the exceedance probability of some value $x \in \mathbb{R}$ that is larger than the $(1 - 1/t)$ -quantile of the distribution of $r_j(X)$ is

$$\mathbb{E}\{r_j(X) > x\} \approx \begin{cases} t^{-1} \left\{\text{sgn}(\xi) \frac{x - \mu_{j,t}}{\sigma_{j,t}}\right\}^{-1/\xi}, & \xi \neq 0, \\ t^{-1} \exp\left(-\frac{x - \mu_{j,t}}{\sigma_{j,t}}\right), & \xi = 0. \end{cases} \quad (3.17)$$

While the asymptotic behaviour of $\mu_{j,t}$ and $\sigma_{j,t}$ as $t \rightarrow \infty$ is uniquely determined by (3.17), additional assumptions on $A(s)$ and $B(s)$, such as $r_1(A) = 1$ and $r_1(B) = 0$, are necessary to ensure the identifiability of a, b, A, B and $\theta_\xi^{r_j}$ from (3.15) and (3.16).

For large t , estimates of the three parameters $\xi, \mu_{j,t}$ and $\sigma_{j,t}$ can be obtained using standard techniques of univariate extreme value statistics by assuming equality in (3.14) or (3.17). For instance, with u_j being a suitably high marginal threshold and $\mathcal{I} = \{i = 1, \dots, n : r_j(X_i) > u_j\}$, (3.17) provides the censored log-likelihood for $\xi, \mu = \mu_{j,t}$ and $\sigma = \sigma_{j,t}$ by

$$\begin{aligned} & \log \ell_j^{\text{cens}}(\xi, \mu, \sigma) \\ &= \begin{cases} (n - |\mathcal{I}|) \log \left[1 - t^{-1} \left\{ \text{sgn}(\xi) \frac{u_j - \mu}{\sigma} \right\}^{-1/\xi} \right] \\ \quad - |\mathcal{I}| \log(t|\xi|\sigma) - (1 + \xi^{-1}) \sum_{i \in \mathcal{I}} \log \left\{ \text{sgn}(\xi) \frac{r_j(X_i) - \mu}{\sigma} \right\}, & \xi \neq 0, \\ (n - |\mathcal{I}|) \log \left\{ 1 - t^{-1} \exp \left(-\frac{u_j - \mu}{\sigma} \right) \right\} - |\mathcal{I}| \log(t\sigma) - \sum_{i \in \mathcal{I}} \frac{r_j(X_i) - \mu}{\sigma}, & \xi = 0. \end{cases} \end{aligned}$$

We obtain the estimate $\hat{\vartheta}_{\text{IndCens}}$ as the maximizer of the independence log-likelihood $\log L^{\text{IndCens}}(\vartheta) = \sum_{j=1}^L \ell_j^{\text{cens}}(\xi, \mu_{j,t}, \sigma_{j,t})$ (Chandler and Bate, 2007).

3.3.3 Censored likelihood for the joint tail behaviour

Alternatively, we can estimate ϑ making use of the multivariate tail behaviour of the whole vector $\{r_1(X), \dots, r_L(X)\}$. For simplicity, we present formulae for $\xi = 0$ only, but similar formulae can be obtained for $\xi \neq 0$. For $x_1, \dots, x_L \in \mathbb{R}$ and sufficiently large $t > 0$, by Theorem 3.2,

$$\mathbb{E} \left[\bigcup_{j=1}^L \left\{ \frac{r_j(X) - \mu_{j,t}}{\sigma_{j,t}} > x_j \right\} \right] \approx t^{-1} V_\vartheta(x_1, \dots, x_L),$$

where

$$V_\vartheta(x_1, \dots, x_L) = \mathbb{E} \left[\max_{j=1}^L \exp \left\{ -x_j - \log \theta_0^{r_j} + \frac{r_j(A_{\vartheta_A} \log W)}{r_j(A_{\vartheta_A})} \right\} \right], \quad (3.18)$$

is the exponent measure of a max-stable vector with standard Gumbel margins. Thus, ϑ can be estimated by a censored likelihood approach. Define a vector $u = (u_1, \dots, u_L)$ whose j th element $u_j \in \mathbb{R}$ is a suitably high marginal threshold for $r_j(X)$, such as its empirical $(1 - 1/t)$ -quantile, and let $\mathcal{K}_i = \{j = 1, \dots, L : r_j(X_i) > u_j\}$. Denoting the

Chapter 3. Extremal behaviour of aggregated data

normalized thresholds and data by $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_L)$ and $Y_i = (Y_{i1}, \dots, Y_{iL})$ with

$$\tilde{u}_j = \frac{u_j - \mu_{j,t}}{\sigma_{j,t}}, \quad Y_{ij} = \frac{r_j(X_i) - \mu_{j,t}}{\sigma_{j,t}}, \quad j = 1, \dots, L,$$

respectively, we let $\hat{\vartheta}_{\text{cens}}$ be the maximizer of the log-likelihood

$$(n - |\mathcal{I}|) \log \{1 - t^{-1} V_{\vartheta}(\tilde{u})\} + \sum_{i \in \mathcal{I}} \log \left[\left\{ \prod_{j \in \mathcal{K}_i} \frac{1}{a(t) r_j(A_{\vartheta_A})} \right\} (-1) t^{-1} V_{\vartheta, \mathcal{K}_i}(Y_i) \right], \quad (3.19)$$

where $\mathcal{I} = \{i = 1, \dots, n : r_j(X_i) > u_j \text{ with } j = 1, \dots, L\}$ and $V_{\vartheta, \mathcal{K}_i}$ are the partial derivatives of V_{ϑ} in directions \mathcal{K}_i . By the homogeneity of V_{ϑ} , it can be seen that the likelihood (3.19) asymptotically does not depend on the choice of t , but only on the u_1, \dots, u_L . This likelihood corresponds to multivariate threshold exceedances and their approximation by Pareto processes (Thibaud and Opitz, 2015). The censoring of the exponent measure V_{ϑ} reduces possible bias for observations below the marginal threshold that might not yet have converged to the limit model; see Wadsworth and Tawn (2014).

Using the censored likelihood requires knowledge of the distribution of r_j applied to $\log W$. This limits this multivariate approach to the special though important case of the Brown–Resnick model where the aggregations are spatial averages and the marginals are in the Gumbel domain of attraction, that is, $\xi = 0$. For this case, in a simulation study described in Appendix C.3, we compare the inference procedures described above. The censored likelihood approach is significantly more efficient since it uses the full information on extremal dependence. In the other cases, namely $\xi < 0$ and $\xi > 0$, the simulation study shows that the independence likelihood procedure provides accurate estimates of the model parameters, including the shape parameter ξ if it is treated as unknown.

3.4 Simulation of Extreme Events

Environmental risk assessment is often based on rare event simulation of scenarios with long return periods. Two kinds of simulations are typically required: unconditional simulations of a given or fitted model capturing the spatial extent and the variability of possible extreme events; and simulations at points of interest conditional on a particular event that was only observed at different locations or scales. Conditional and unconditional simulations have for instance been studied for max-stable processes (Dombry et al., 2013, 2016) and for threshold exceedances (Thibaud and Opitz, 2015; de Fondeville and Davison, 2018).

In this section, we discuss how the multivariate result in Theorem 3.2 allows us to perform these two kinds of simulations for extreme events of the process X . We assume that the process X satisfies the assumptions of Theorem 3.2 for known normalizing functions a_s and b_s with representation (3.13), extreme value index $\xi \in \mathbb{R}$, and known distribution of the spectral process W . For simplicity, we again restrict to the case $\xi = 0$, but the procedure can be adapted for $\xi \neq 0$.

For simulation of X at locations $s_1, \dots, s_K \in S$, we artificially augment the vector of functionals to $\{r_1(X), \dots, r_L(X), r_{L+1}(X), \dots, r_{L+K}(X)\}$, where $r_{L+k}(X) = X(s_k)$ is the point evaluation at location s_k ($k = 1, \dots, K$). We apply Theorem 3.2 to this augmented vector to obtain

$$\lim_{t \rightarrow \infty} t \mathbb{E} \left[\bigcup_{j=1}^L \left\{ \frac{r_j(X) - \mu_{j,t}}{\sigma_{j,t}} > x_j \right\} \right] = \mathbb{E} \left\{ \prod_{j=1}^{L+K} \exp(-x_j + \log \Psi_j) \right\},$$

where $(\Psi_1, \dots, \Psi_{L+K})$ is a random vector with distribution P given by

$$\Psi_j = \exp \left\{ \frac{r_j(A \log W)}{r_j(A)} - \log \theta_0^{r_j} \right\}, \quad j = 1, \dots, L+K,$$

and $\mu_{j,t}$ and $\sigma_{j,t}$ ($j = 1, \dots, L+K; t > 0$) are defined in (3.15) and (3.16). In other words, $\{r_1(X), \dots, r_{L+K}(X)\}$ is in the max-domain of attraction of a max-stable distribution with standard Gumbel margins and spectral vector $(\Psi_1, \dots, \Psi_{L+K})$.

For conditional and unconditional simulation of an extreme event we consider the case of only one aggregation functional, i.e., $L = 1$, which is assumed to be large. This functional might itself be an aggregation of other functionals, which makes this setting rather general. Reformulating Theorem 3.2 in terms of threshold exceedances, we obtain the convergence in distribution

$$\mathcal{L} \left[\left\{ \frac{r_j(X) - \mu_{j,t}}{\sigma_{j,t}} \right\}_{j=1}^{L+K} \mid \frac{r_1(X) - \mu_{1,t}}{\sigma_{1,t}} > 0 \right] \rightarrow \mathcal{L}(U + \log \Psi^{(1)}), \quad t \rightarrow \infty, \quad (3.20)$$

where U is a standard exponential random variable and, independently of U , $\Psi^{(1)}$ is a $(L+K)$ -dimensional random vector satisfying $\Psi_1^{(1)} = 1$ almost surely. The distribution P_1 of $\Psi^{(1)}$ is obtained from P via a measure transform (Dombry and Ribatet, 2015; Dombry et al., 2016) and in many cases it can be simulated by rejection sampling (de Fondeville and Davison, 2018).

While unconditional simulation requires X to be extreme in the sense that $\{r_1(X) - \mu_{1,t}\}/\sigma_{1,t} > 0$ for large t , for conditional simulation, the large value y_1 of $r_1(X)$ is explicitly given. Assuming equality in (3.20), this condition determines the value

Chapter 3. Extremal behaviour of aggregated data

$u = (y_1 - \mu_{1,t})/\sigma_{1,t} > 0$ of the exponential random variable U since $\log \Psi_1^{(1)} = 0$ almost surely. We can perform unconditional and conditional simulation of the vector $\{X(s_1), \dots, X(s_K)\}$ in the following way.

(i) Sample a realization u of standard exponential random variable for an unconditional simulation. For a conditional simulation given $r_1(X) = y_1$, set $u = (y_1 - \mu_{1,t})/\sigma_{1,t}$.

(ii) Simulate a realization $(\psi_{L+1}, \dots, \psi_{L+K})$ of the distribution of $(\Psi_{L+1}^{(1)}, \dots, \Psi_{L+K}^{(1)})$.

(iii) Return $x = \{x(s_1), \dots, x(s_K)\}$ with

$$x(s_k) = a_s(t)(u + \log \psi_{L+k}) + b_s(t), \quad k = 1, \dots, K.$$

Equation (3.20) can also be used for conditional simulation when $L > 1$, i.e., if the values y_1, \dots, y_L for several functionals $r_1(X), \dots, r_L(X)$ are given. In this case, only the second step of the above procedure has to be modified: instead of an unconditional simulation of $(\Psi_{L+1}^{(1)}, \dots, \Psi_{L+K}^{(1)})$, a conditional simulation given $\Psi_j^{(1)} = (y_j - \mu_{j,t})/\sigma_{j,t} - u$ ($j = 2, \dots, L$) has to be performed. To this end, the conditional distribution of the transformed measure P_1 needs to be tractable, which is true in few cases only. For our running example of a limiting Brown–Resnick process, the following makes this explicit.

Example 3.8 *As in Example 3.7, let $S \subset \mathbb{R}^D$ be compact, let $\xi = 0$ and let X be in the max-domain of attraction of a Brown–Resnick process. The aggregation functionals r_j are spatial averages over compact regions $S_j \subset S$ ($j = 1, \dots, L$) or point evaluations $r_{L+k}(X) = X(s_k)$ at locations $s_k \in S$ ($k = 1, \dots, K$). The vector $\{r_1(X), \dots, r_{L+K}(X)\}$ then satisfies the assumptions of Theorem 3.2, and it is in the max-domain of attraction of a multivariate Hüsler–Reiss distribution with dependence matrix*

$$\Gamma = \begin{pmatrix} \{\Gamma_{jk}\}_{j,k} & \{\Gamma_{jq}\}_{j,q} \\ \{\Gamma_{pk}\}_{p,k} & \{\Gamma_{pq}\}_{p,q} \end{pmatrix}, \quad \begin{cases} j, k = 1, \dots, L, \\ p, q = L+1, \dots, L+K. \end{cases}$$

The entries of the four sub-matrices and the explicit form of the exponent measure are given in Appendix C.2.2. In this case, the above algorithms essentially reduce to conditional and unconditional simulation of Gaussian processes.

3.5 Application: downscaling extremes

3.5.1 Statistical downscaling

Environmental data can be classified into two broad categories. On the one hand, station measurements are obtained through direct observation of the physical quantity. This type of data refers to a precise location in space, but may suffer from inhomogeneities between stations due to varying record lengths and differences between measurement instruments, and, moreover, it usually has sparse spatial coverage. Gridded databases, for instance generated by climate models, on the other hand, cover a large region or even the entire globe, but at a coarse scale where data points can be considered as an aggregation of the physical variable.

Understanding the link from these gridded data to point measurements is an important area of research in environmental sciences called downscaling. Apart from dynamical downscaling procedures based on the solution of partial differential equations describing the physical processes, a large number of downscaling techniques relying on the statistical relationship between variables at different scales have been proposed. Most of these techniques focus on central characteristics of the distribution such as its mean and variance. In geostatistics, for instance, the so-called change of support problem has been extensively studied for Gaussian processes (cf., Chiles and Delfiner, 2012, and references therein). There are few examples of statistical downscaling procedures for extremes. Mannshardt-Shamseldin et al. (2010) and Kallache et al. (2011) follow an approach related to univariate extreme value theory, and Bechler et al. (2015b) and Oesting et al. (2017a) propose conditional simulation from a spatial max-stable process that has been estimated from station measurements. The downscaling method in Towe et al. (2017) for significant wave height involves several other variables but it is not multivariate in space.

Here, using the theoretical results in Section 3.2, we extend the idea of changing the support of a stochastic process X to the context of extremes, basing inference only on aggregated observations $r_1(X), \dots, r_L(X)$. These might come from gridded data sets, as in our case, supposing that the grid values represent an aggregation of the underlying physical quantity. If additional station measurements $X(s_1), \dots, X(s_K)$ are available, they can also be used. The method allows the estimation of marginal characteristics such as return levels at point locations, as well as unconditional and conditional simulations of rare events on the entire region S .

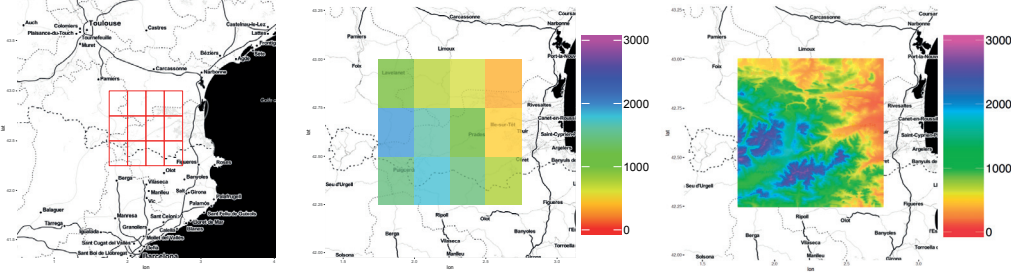


Figure 3.1 – The study region consisting of 12 grid cells in the south of France (left), mean altitude within each cell (middle) and elevation in the region.

3.5.2 Application to extreme temperature in the South of France

We apply our downscaling procedure to daily temperature maxima in Europe from the e-obs data set (Haylock et al., 2008), which covers the period from 1950 to 2016 at a 0.25° grid resolution. To avoid potential temporal non-stationarity, we restrict the study to July and August. Our study region S is a $80\text{km} \times 80\text{km}$ subset of the gridded product located in the south of France, west of Perpignan; see Figure 3.1. The region is mountainous and thus altitude is a natural covariate for our model. The underlying spatial process of temperatures is denoted by $\{X(s) : s \in S\}$, and the observations $\{r_1(X_i), \dots, r_L(X_i)\}$ on day i ($i = 1, \dots, n$) can be considered as the spatial averages over the $L = 12$ cells in S . Here, n is the number of days in the given time span of 67 years. The null hypothesis that the marginal tails of the aggregated data are in the Gumbel domain of attraction cannot be rejected, and we thus assume below that $\xi = 0$. This simplification, while dangerous in practice, as it is likely to induce a severe underestimation of the confidence intervals, is made to illustrate the full potential of our downscaling model.

Throughout we assume the same setting as in Section 3.3.1, namely that the marginal distributions of $X(s)$ belong to a location-scale family for all $s \in S$, parametrized through the functions

$$A(s) = 1, \quad B_{\theta_B}(s) = b_0 + b_1 \text{alt}(s) + b_2 \text{lon}(s) + b_3 \text{lat}(s),$$

where $\text{alt}(s)$, $\text{lon}(s)$ and $\text{lat}(s)$ denote the altitude, longitude and latitude at location $s \in S$. We further suppose that X is in the functional max-domain of attraction of a max-stable process Z belonging to a parametric family $\{Z_{\theta_W} : \theta_W \in \Theta_W\}$, for which we consider the Brown–Resnick processes introduced in Example 3.1, parametrized

3.5. Application: downscaling extremes

	a_n	b_n	b_1	b_2	b_3	α	τ	η	a
Estimate	1.90	35.53	4.51	-0.53	-0.20	0.90	6.42	-0.08	1.14
Standard deviation	0.06	0.27	0.14	0.26	0.28	0.07	0.51	0.22	0.08

Table 3.1 – Estimated parameters and standard deviations for the temperature downscaling model. Standard deviations are computed using a block jackknife with 19 blocks of size 6.

by $\vartheta_W = (\alpha, \tau, \eta, a)$ for the anisotropic power variogram

$$\gamma(s_1, s_2) = \left\| \frac{\Omega(s_1 - s_2)}{\tau} \right\|^\alpha, \quad s_1, s_2 \in S,$$

with $0 < \alpha \leq 2, \tau > 0$ and anisotropy matrix

$$\Omega = \begin{bmatrix} \cos \eta & -\sin \eta \\ a \sin \eta & a \cos \eta \end{bmatrix}, \quad \eta \in \left(-\frac{\pi}{2}; \frac{\pi}{2}\right], \quad a > 1.$$

In Section 3.3.2 and Section 3.3.3 we discussed two approaches to estimating the parameters of this model, namely independence likelihood and censored likelihood estimation for multivariate threshold exceedances. The formulae required for the implementation of these approaches were derived in Section 3.2 and Section 3.3 and in Appendix C.2.2. For censored likelihood estimation of the model parameters in (3.19), we require the partial derivatives $V_{\mathcal{K}}$ of the exponent measure V , which can be obtained as in Asadi et al. (2015, Section 4.3.2). In order to assess its effectiveness and to compare the efficiency of the two methods, we conduct a simulation study with a setup similar to this application, that can be found in Appendix C.3. The censored likelihood approach is significantly more efficient since it uses the full information on extremal dependence.

The parameters of our model for temperature extremes are therefore fitted using the censored likelihood procedure based on all observations where the empirical marginal 0.98 quantile is exceeded at at least one location. To avoid possible temporal dependence we keep only observations that are at least 5 days apart, yielding 114 events. The parameter estimates are displayed in Table 3.1, where standard deviations are obtained using a jackknife procedure with 19 blocks of size 6; censored maximum likelihood estimation is performed repeatedly with one block left out.

We assess the model fit in the diagnostic plots shown in Appendix C.4. We check the marginal distributions implied by the fitted linear model by comparing them in quantile-quantile plots to the observations. The model provides a good fit for most

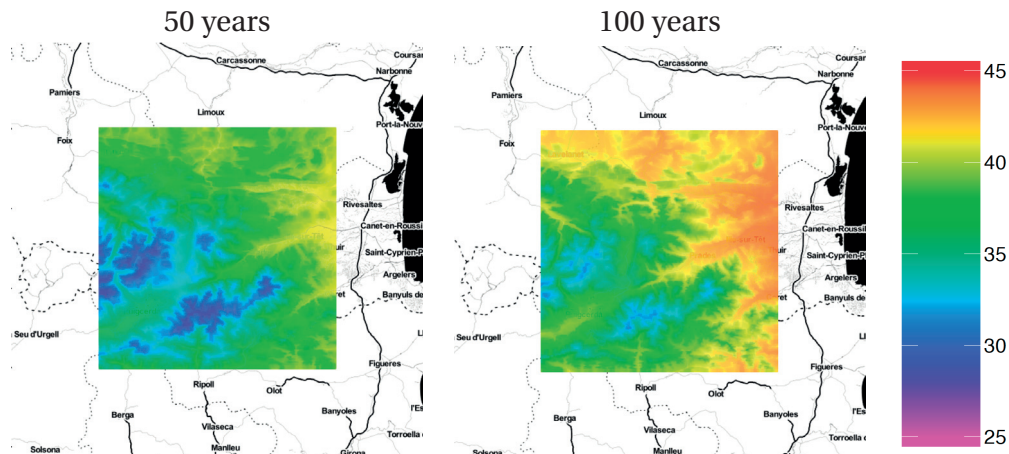


Figure 3.2 – Downscaled return levels of daily temperature maxima ($^{\circ}\text{C}$) for the 50- (left) and 100- (right) year return periods in the study region at a $25 \times 25\text{m}$ resolution.

stations and the quantiles of the fitted model generally remain in the confidence bounds obtained by parametric bootstrap. For a small number of stations, the model slightly over-estimates return levels.

Verification of the dependence structure is based on a graphical comparison of the pairwise extremogram (Davis and Mikosch, 2009) from the fitted multivariate Hüsler-Reiss model to its empirical counterpart based on the gridded observations. The extremogram values were significantly larger than zero for increasing thresholds and stable around the empirical 0.98 quantile, validating the asymptotic dependence model. The fitted variogram model successfully captures the major trend of the cloud of points. The effect of spatial anisotropy seems to be rather weak, which is also reflected in the parameter estimate for a close to 1.

The fitted marginal model allows us to obtain return level maps for point locations at arbitrarily fine resolutions. In Figure 3.2, we produced such maps for the 50 and 100 year return periods. The full fitted model of marginal distributions and dependence structure further enables us to conditionally and unconditionally generate spatial extreme events of temperature fields at both a coarse and a fine resolution grid via the simulation procedures described in Section 3.4. Figure 3.3 displays two high resolution simulations of the temperature field conditionally on the observed aggregated temperatures during the warmest day of the 2003 heatwave. The simulations show that extreme temperatures at fine resolutions can be much larger than at a coarse scale. Moreover, both simulations are constrained to have the same observed averages on the grid boxes, but they may exhibit different spatial patterns. This illustrates the variability of such a heatwave and provides practitioners with a set of possible

3.5. Application: downscaling extremes

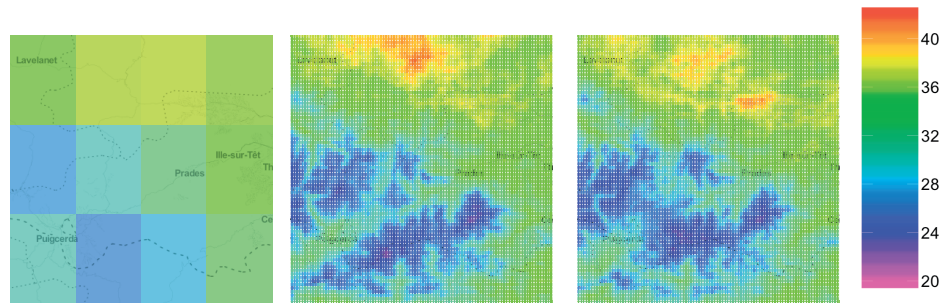


Figure 3.3 – Maximal temperature ($^{\circ}\text{C}$) on the warmest day during the 2003 heatwave. Gridded data from the e-obs database (Haylock et al., 2008) (left); conditional simulations with a 1×1 km resolution (centre and right).

scenarios that can be used for risk assessment.

4 Functional peaks-over-threshold analysis

This chapter is a preprint of a paper entitled ‘Functional peaks-over-threshold analysis and generalized r -Pareto processes’ jointly written with Anthony Davison. For this reason, the chapter is structured to be self-standing and thus for clarity of the exposition some parts may overlap with other chapters of the thesis.

4.1 Introduction

Extreme Value Theory (EVT) provides a theoretical framework to describe and model tails of statistical distributions within which estimating the frequency of past extreme events as well as to extrapolating beyond observed severities is possible. These have been extensively studied in a univariate framework (Fisher and Tippett, 1928; Gnedenko, 1943; Davison and Smith, 1990) especially for independent identically distributed replicates, and applications have been developed in fields such as finance, insurance, hydrology and telecommunications (Hosking and Wallis, 1987; Katz et al., 2002; Embrechts et al., 1997).

Due to recent extreme events, there has been a surge of interest in environmental applications, motivated by the necessity to better understand the impact of global warming. Floods, windstorms, heatwaves have a complex spatio-temporal structure that cannot be modelled using univariate extreme value theory.

Max-stable processes (de Haan and Ferreira, 2006, Section 9.2), which provide a functional extension of the generalized extreme value distribution, have successfully been used to study the extremal behaviour of monthly and annual maxima (Coles, 2001, p.47-48), but applications have been limited due to the mathematical and computational complexity of such models (Huser and Davison, 2013). Also, the study of maxima discards a fair amount of information, making detection of mixtures in tail

Chapter 4. Functional peaks-over-threshold analysis

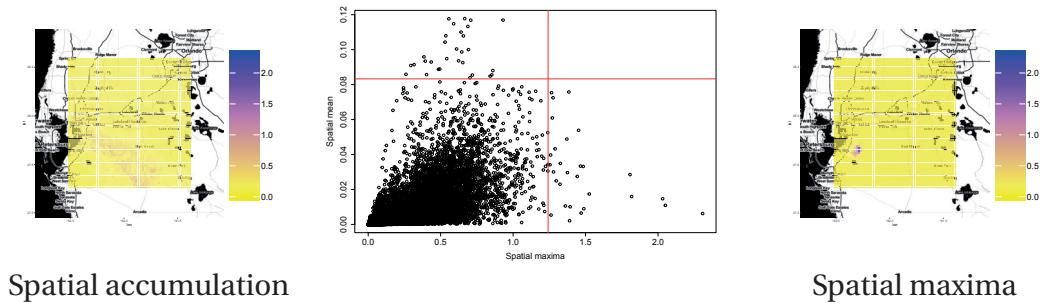


Figure 4.1 – The middle panel displays the spatial mean $\int_S X(s)ds$ against the spatial maxima $\max_{s \in S} X(s)$ of observed fifteen-minute cumulated rainfall (inches) from radar rainfall measurements over Florida from 1999 to 2004. The red lines represent the thresholds corresponding to the top 20 events. The left panel displays the most intense event for spatial accumulation, while the right displays the largest exceedances for spatial maxima.

behaviour very difficult. For example, rainfall is usually divided into two classes: convective rain, which is local and marginally very intense, and cyclonic spells generating larger spatial accumulations of water but with lower local intensities. These phenomena are driven by different independent weather conditions that may both cause severe floods and, as suggested by Figure 4.1, their tail marginal distribution and spatio-temporal structure are likely to differ. With block maxima, marginally intense events naturally dominate and thus impose a focus on convective rainfall, while disregarding potential extreme cyclonic events. For risk mitigation, studying extremes of different natures is crucial, and max-stable processes are inappropriate for modelling such complex phenomena, since taking maxima largely eliminates certain types of events.

Functional peaks-over-thresholds methods, similarly to the generalized Pareto distribution in univariate extreme value theory, define extreme events as exceedances over a threshold and enable the analyst to detect and model intricate and complex extreme events. Indeed, the framework gives a theoretical foundation to detect mixtures of tail behaviour through different definitions of exceedances tailored to the type of extreme events of interest.

In this context, reduction of multivariate datasets to univariate structural variables, such as $\max(X_1, X_2)$ or $X_1^2 + X_2^2$, on which generalized Pareto distributions are fitted (Coles and Tawn, 1994), is common to study complex multivariate extreme events. However, this approach does not give insight on the combination of events yielding an exceedance and is hindered by the fact that different univariate summaries may lead to

different tail behaviour. One way to understand these differences is to suppose that the observations are generated by an underlying mixture of generative processes, which are disentangled by computing these univariate summaries. Thus if the summary captures only one of these processes, for instance only cyclonic, it is not surprising that we obtain different tail behaviours. Functional peaks-over-threshold analysis generalizes this methodology for a better understanding of the underlying dependence structure.

Classical approaches to functional peaks-over-threshold rely on particular types of exceedances (Ferreira and de Haan, 2014) or are limited to regularly-varying stochastic processes (Hult and Lindskog, 2005; Dombry and Ribatet, 2015). The latter means in practice that the data must have an unbounded support and share the same polynomial-type tail decay. If not, observations are standardized to have, for instance, unit Pareto marginals (Klüppelberg and Resnick, 2008), so exceedances must be defined on the transformed data. An extreme event caused by a natural phenomenon such as cyclonic rain is more easily characterized on the original scale, and thus standardization limits the applicability of functional peaks-over-threshold methods. In univariate extreme value theory, the generalized Pareto distribution gives a unified framework to describe directly the tail decay of the original data, and encompasses the Weibull, Gumbel and Fréchet tail decay regimes. This paper provides a similar unified formulation for functional peaks-over-threshold analysis under the assumption that the process has the same tail decay over its domain. This restriction on the tail behaviour is necessary to define the exceedances directly on the original process, otherwise the region or location with the heaviest tail dominates the limit distribution and yields unrealistic models. In this context, we extend Dombry and Ribatet (2015) by introducing the generalized r -Pareto process, allowing more flexible excess definitions and generalized Pareto tail margins. The generalized r -Pareto process is the only limit of exceedances of a properly rescaled regularly varying process and for some specific definitions of exceedance, it can be factorized to enable simulation of events with a fixed intensity, i.e. events for which the risk measure equals a pre-determined return level.

Section 4.2 reviews classical results for univariate and functional peaks-over-threshold analysis and highlights their limitations. In Section 4.3, we derive convergence results for the three possible regimes of tail decay and show how these results simplify depending on the properties of the exceedance functional. Section 4.4 defines and characterizes the generalized r -Pareto process, for which we emphasize the relation with max-stable processes and present potential simulation algorithms. In Section 4.5, we discuss statistical inference and conclude in Section 4.6 by describing methods for model validation. The proof of the main results are relegated to Appendix D.

4.2 Modelling exceedances over a high threshold

4.2.1 Univariate model

Let X be a random variable for which there exist sequences of constants $a_n > 0$ and b_n such that

$$n \Pr\left(\frac{X - b_n}{a_n} > x\right) \rightarrow -\log G(x), \quad n \rightarrow \infty, \quad (4.1)$$

where G is a non-degenerate distribution function. Then, X is said to belong to the max-domain of attraction of G (Resnick, 1987, p. 12). For a large enough threshold $u < \inf\{x : F(x) = 1\}$, its tail distribution can be approximated by a generalized Pareto distribution (Davison and Smith, 1990), yielding

$$\Pr(X - u > x \mid X > u) \approx H_{(\xi, \sigma)}(x) = \begin{cases} (1 + \xi x/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x/\sigma), & \xi = 0, \end{cases} \quad (4.2)$$

where $\sigma = \sigma(u) > 0$ and $a_+ = \max(a, 0)$. If the shape parameter ξ is negative, then x must lie in the interval $[0, -\sigma/\xi]$, whereas x can take any positive value with positive or zero ξ . The random variable X is said to belong to the Weibull, the Gumbel or the Fréchet domains of attraction if the limiting shape parameter is respectively negative, zero or positive. The max-domain of attraction conditions are satisfied by a vast class of random variables (e.g. Beirlant et al., 2004, pp. 59, 62, 72). Davison and Smith (1990) use equation (4.2) to approximate the distribution function F of X by

$$F(x) \approx 1 - \zeta_u H_{(\xi, \sigma)}(x - u), \quad x > u, \quad (4.3)$$

where ζ_u , the probability that X exceeds the threshold u , is determined by u . The generalized Pareto distribution offers a flexible and unified model for tails of a wide class of random variable X , and is today considered as a standard approach for univariate risk estimation and extrapolation.

In its simplest form the model for univariate exceedances in equation (4.3) applies to independent and identically distributed variables, but its use has been extended to time series, non-stationary and spatial data. The modelling of exceedances can be extended to a multivariate setting (Rootzén and Tajvidi, 2006; Rootzén et al., 2018b,a) and to continuous processes (Ferreira and de Haan, 2014; Dombry and Ribatet, 2015) within the functional regular variation framework, which we now describe.

4.2.2 Functional regular variation and \mathcal{R} -Pareto processes

Let $S \subset \mathbb{R}^D$ ($D \geq 1$) be a compact metric space, such as $[0, 1]^2$ for spatial applications. We write $\mathcal{F}_+ = C\{S, [0, \infty)\}$ for the subset of the Banach space of continuous functions $x : S \rightarrow \mathbb{R}$ endowed with the uniform norm $\|x\|_\infty = \sup_{s \in S} |x(s)|$ and $\mathcal{B}(\Xi)$ for the Borel sigma-algebra associated to a metric space Ξ . A measurable closed subset \mathcal{C} of \mathcal{F}_+ is called a cone if $tx \in \mathcal{C}$ for any $x \in \mathcal{C}$ and $t > 0$. When studying extremes the cones $\mathcal{C} = \{0\}$ or $\mathcal{C} = \{x \in \mathcal{F}_+ : \inf_{s \in S} x(s) = 0\}$ are often excluded from \mathcal{F}_+ to avoid the appearance of limiting measures with infinite masses at the origin or on the coordinate axes. Let $M_{\mathcal{F}_+ \setminus \mathcal{C}}$ denote the class of Borel measures on $\mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ for a cone \mathcal{C} , and say that a set $\mathcal{A} \in \mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ is bounded away from \mathcal{C} if $d(\mathcal{A}, \mathcal{C}) = \inf_{x \in \mathcal{A}, y \in \mathcal{C}} \|x - y\| > 0$.

A sequence of measures $\{\Lambda_n\} \subset M_{\mathcal{F}_+ \setminus \mathcal{C}}$ is said to converge to a limit $\Lambda \in M_{\mathcal{F}_+ \setminus \mathcal{C}}$, written $\Lambda_n \xrightarrow{\hat{w}} \Lambda$ (Hult and Lindskog, 2005), if $\lim_{n \rightarrow \infty} \Lambda_n(\mathcal{A}) = \Lambda(\mathcal{A})$, for all $\mathcal{A} \in \mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$ bounded away from \mathcal{C} with $\Lambda(\partial\mathcal{A}) = 0$, where $\partial\mathcal{A}$ denotes the boundary of \mathcal{A} . For equivalent definitions of this so-called \hat{w} -convergence, see Lindskog et al. (2014, Theorem 2.1).

One way to generalize equation (4.1) to functions is to use the concept of functional regular variation (Hult and Lindskog, 2005): a stochastic process X with sample paths in $\mathcal{F}_+ \setminus \mathcal{C}$ is regularly varying if there exist a sequence of strictly positive continuous functions $\{a_n\}_{n=1}^\infty$ with $\lim_{n \rightarrow \infty} a_n(s) = \infty$ for each $s \in S$ and a measure $\Lambda \in M_{\mathcal{F}_+ \setminus \mathcal{C}}$ such that

$$n \Pr(a_n^{-1} X \in \cdot) \xrightarrow{\hat{w}} \Lambda(\cdot), \quad n \rightarrow \infty; \quad (4.4)$$

then we write $X \in \text{RV}(\mathcal{F}_+ \setminus \mathcal{C}, a_n, \Lambda)$. Equation (4.4) can be seen as a generalization of equation (4.1), where we supposed that $a_n \rightarrow \infty$ and $b_n = 0$, corresponding to the Fréchet regime. The limiting measure Λ satisfies a homogeneity property: for any positive scalar $t > 0$, the measure of the set $t\mathcal{A}$, where \mathcal{A} is an element of $\mathcal{B}(\mathcal{F}_+ \setminus \mathcal{C})$, is equal to $t^{-1/\xi} \Lambda(\mathcal{A})$, for some $\xi > 0$, called the tail index. Homogeneity is a key component of functional extreme value theory, because it allows the extrapolation of the measure of any set containing observed extreme events to sets containing only unobserved events; see Section 1.3.3 for more details.

In a scalar context, it is straightforward to define the exceedance of a random variable X over a threshold u . For functions, an appropriate notion of exceedance can be defined through the concept of r -exceedances (Dombry and Ribatet, 2015). A risk functional $r : \mathcal{F}_+ \rightarrow [0, \infty)$ is defined to be a continuous functional satisfying a homogeneity property, i.e., there exists $\kappa > 0$ such that $r(ax) = a^\kappa r(x)$ for $x \in \mathcal{F}_+$ and $a > 0$. An r -exceedance is an event of the form $\{r(X) \geq u\}$ for some $u \geq 1$. Under these hypotheses, it is straightforward to verify that the set $\mathcal{C}_r = \{x \in \mathcal{F}_+ : r(x) = 0\}$ is a cone

Chapter 4. Functional peaks-over-threshold analysis

of \mathcal{F}_+ and thus following Lindskog et al. (2014, Theorem 2.3), regular variation on $\mathcal{F}_+ \setminus \{0\}$ implies regular variation on $\mathcal{F}_+ \setminus \mathcal{C}_r$.

Suppose that X denotes a regularly varying stochastic process on $\mathcal{F}_0 = \mathcal{F}_+ \setminus \{0\}$ with limiting measure Λ . Then, there exist $\xi > 0$ and a probability measure σ_r on $\mathcal{S}_r = \{x \in \mathcal{F}_+ : \|x\|_{\text{ang}} = 1\}$, such that for any $r' \geq 1$ and $\mathcal{W} \subset \mathcal{S}_r$ (de Fondeville and Davison, 2018),

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{X}{a_n} \in \left\{ x \in \mathcal{F}_0 : r(x) \geq r', \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W} \right\} \middle| r \left(\frac{X}{a_n} \right) \geq 1 \right] = r'^{-1/\xi} \sigma_r(\mathcal{W}).$$

The factorization in (4.5) is called a pseudo-polar decomposition and separates the intensity of the r -exceedances, measured by $r'^{-1/\xi}$, with their dependence driven by the angular component $X/\|X\|_{\text{ang}}$. The stochastic process P on $\{x \in \mathcal{F}_0 : r(x) \geq 1\}$ with probability measure given by equation (4.5) is called an r -Pareto process and can be used to approximate the distribution of r -exceedances over a sufficiently high threshold $u > 0$, see Dombry and Ribatet (2015) for more details. de Fondeville and Davison (2018) develop and compare high-dimensional inference procedures for r -Pareto processes, and apply their results to extreme rainfall over Florida for two types of exceedances.

The r -Pareto processes lack flexibility: for any location $s_0 \in S$ and a sufficiently high threshold $u_0 > 0$ satisfying $\{x \in \mathcal{F}_0 : x(s_0) > u_0\} \subset \{x \in \mathcal{F}_0 : r(x) \geq 1\}$, the marginal upper tail probability of P at location s_0 is

$$\Pr \{P(s_0) > r'\} = \left(\frac{r'}{u_0} \right)^{-1/\xi} \frac{\Lambda \{x \in \mathcal{F} \setminus \mathcal{C}_r : r(x) \geq 1, x(s_0) \geq u_0\}}{\Lambda \{x \in \mathcal{F} \setminus \mathcal{C}_r : r(x) \geq 1\}}, \quad r \geq u_0, \quad (4.5)$$

so above u_0 , the r -Pareto process has Pareto-type marginals with tail index ξ . From a practical point of view, this limits application of the r -Pareto process to datasets with Fréchet-type tails or it requires one to standardize the marginals to be unit Fréchet (Coles and Tawn, 1991, Section 5) or unit Pareto (Klüppelberg and Resnick, 2008). In the latter, r -exceedances are defined using the transformed data, potentially leading to rather unintuitive risk functionals to discriminate between the different processes generating the observations; see Chapter 2 for examples of functionals characterizing cyclonic against convective rainfall. Also, in this setting, the risk functionals must be continuous and homogeneous, which can be restrictive. In Section 4.3, we generalize r -exceedances to a wider class of functionals and derive the convergence theorem for the three possible tail decay regimes.

4.3 Limiting distributions of r -exceedances

Similarly to Section 4.2.2, $S \subset \mathbb{R}^D$ ($D \geq 1$) denotes a compact metric space but \mathcal{F} now denotes the Banach space of real-valued continuous functions on S , denoted $C(S, \mathbb{R})$, and we write $\mathcal{F}_0 = C\{S, [0, \infty)\} \setminus \{0\}$. Let $\xi \in \mathbb{R}$, $A \in C\{S, (0, \infty)\}$, $B \in C(S, \mathbb{R})$, and define the sets

$$\mathcal{F}^{\xi, A, B} = \begin{cases} \{x \in \mathcal{F} \setminus \{B - \xi A\} : x(s) \geq B(s) - \xi A(s), s \in S\}, & \xi > 0, \\ \mathcal{F}, & \xi = 0, \\ \{x \in \mathcal{F} : x(s) < B(s) - \xi A(s), s \in S\}, & \xi < 0, \end{cases}$$

$$\mathcal{S}_r^\xi = \begin{cases} \{x \in \mathcal{F} : r(x) \geq 0, \|x\|_{\text{ang}} = 1\}, & \xi \neq 0, \\ \{x \in \mathcal{F} : r(x) = 0\}, & \xi = 0, \end{cases}$$

and $\mathcal{U}_r^\xi = [0, \infty)$ if $\xi \geq 0$ and $\mathcal{U}_r^\xi = [0, r(B - \xi A))$ if $\xi < 0$. In this section, X denotes a stochastic process with sample path in the Banach space of real-valued continuous functions for which there exist $\xi \in \mathbb{R}$, sequences $\{a_n\}_{n=1}^\infty > 0$ and $\{b_n\}_{n=1}^\infty$ of continuous functions on S , and a measure $\Lambda \in M_{\mathcal{F}_0}$ such that

$$\left. \begin{array}{l} n \Pr \left[\left\{ 1 + \xi \left(\frac{X - b_n}{a_n} \right) \right\}_+^{1/\xi} \in \cdot \right], \quad \xi \neq 0 \\ n \Pr \left[\exp \left(\frac{X - b_n}{a_n} \right) \in \cdot \right], \quad \xi = 0 \end{array} \right\} \xrightarrow{\hat{w}} \Lambda(\cdot), \quad n \rightarrow \infty, \quad (4.6)$$

where $\{\cdot\}_+ = \max(\cdot, 0)$ is taken component-wise and a_n and b_n are chosen such that for any $s \in S$

$$n \Pr \left\{ \frac{X(s) - b_n(s)}{a_n(s)} > x \right\} \rightarrow \begin{cases} (1 + \xi x)^{-1/\xi}, & \xi \neq 0 \\ \exp(-x), & \xi = 0 \end{cases} \quad n \rightarrow \infty, \quad (4.7)$$

with $1 + \xi x > 0$ when $\xi \neq 0$, and $x > 0$ for $\xi = 0$. Equation (4.6) generalizes (4.1) and defines a general form of functional regular variation introduced by Ferreira and de Haan (2014); we write $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$. Similarly to classical regular variation, the limiting measure Λ is (-1) -homogeneous (Lindskog et al., 2014, Theorem 3.1).

We now extend the notion of risk functional by relaxing some assumptions of Section 4.2.2. Now a functional $r : \mathcal{F} \rightarrow \mathbb{R}$ is said to be a valid risk functional for the process $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$ if it is monotonic increasing, there exist continuous functions $A > 0$ and B such that

$$\limsup_{n \rightarrow \infty} \sup_{s \in S} \left| \frac{a_n(s)}{r(a_n)} - A(s) \right| = 0, \quad \limsup_{n \rightarrow \infty} \sup_{s \in S} \left| \frac{b_n(s) - r(b_n)}{r(a_n)} - B(s) \right| = 0, \quad s \in S, \quad (4.8)$$

Chapter 4. Functional peaks-over-threshold analysis

and

$$\begin{aligned} r \text{ is continuous at } \{B - \xi^{-1}A\}, \quad r(B - A\xi^{-1}) < 0, \quad \xi > 0, \\ r(x) \rightarrow -\infty \text{ as } x \rightarrow -\infty, \quad r(x+t) = r(x) + t, \quad t \in (0, \infty), \quad \xi \leq 0. \end{aligned} \quad (4.9)$$

These technical assumptions are the minimal requirements needed to prove the convergence of r -exceedances over a threshold $u \geq 0$ to a non-degenerate limit measure. Condition (4.8) implies that the functions a_n and b_n can be decomposed asymptotically into $r(a_n)A(s)$ and $r(b_n) + r(a_n)B(s)$ respectively, and also requires that the speed of convergence of $r(a_n)$ and $a_n(s)$ must be the same for all $s \in S$, restricting the class of valid risk functionals. For instance, when $\xi > 0$ and $B = 0$, the class of 1-homogeneous functionals properly shifted satisfies (4.8) and (4.9), while the class of linear functionals with $r(A) > 0$ and $r(B) = 0$ yields valid risk functionals for all $\xi \in \mathbb{R}$. Similar assumptions were used in Ferreira et al. (2012) and Engelke et al. (2019) and seem reasonable in many environmental applications. For instance, assuming that the distribution F_s at each $s \in S$ belongs to a location-scale family $F[\{x - B(s_0)\}/A(s_0)]$, describing the marginal behaviour of the underlying physical process characterized by the risk functional r , implies a common shape $\xi \in \mathbb{R}$ and that we can choose $a_n(s) = a(n)A(s)$ and $b_n(s) = A(s)b(n) + B(s)$ with $\{a(n)\}, \{b(n)\} \in \mathbb{R}^\infty$.

Theorem 4.1 *Let $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$, and consider a valid risk functional r . Then for any $\mathcal{W} \subset \mathcal{S}_r^\xi$, and $r' \in \mathcal{U}_r^\xi$,*

$$\left. \begin{aligned} n \Pr \left[\frac{X - r(b_n)}{r(a_n)} \in \left\{ x \in \mathcal{F}^{\xi, A, B} : r(x) \geq r', \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W} \right\} \right] \\ n \Pr \left[\frac{X - r(b_n)}{r(a_n)} \in \left\{ x \in \mathcal{F}^{0, A, B} : r(x) \geq r', x - r(x) \in \mathcal{W} \right\} \right] \end{aligned} \right\} \xrightarrow{\hat{w}} \Lambda \{(r', \mathcal{W})\}, \quad n \rightarrow \infty,$$

in $M(\mathcal{F}^{\xi, A, B})$, where

$$\Lambda \{(r', \mathcal{W})\} = \begin{cases} \Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq r', \frac{A(y^\xi - 1) + \xi B}{\|A(y^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}, & \xi \neq 0, \\ \Lambda \left\{ y \in \mathcal{F}_0 : r(A \log y + B) \geq r', A \log y + B - r(A \log y + B) \in \mathcal{W} \right\}, & \xi = 0. \end{cases}$$

Theorem 4.1 describes the limiting measure for r -exceedances of the rescaled process $\{X - r(b_n)\}/r(a_n)$ rather than that of the process X . This standardization is simpler than the marginal transform of the classical regular variation methodology described in Section 4.2.2, as it only requires two real-valued sequences, which can be unknown in practice, and does not modify the tail decay regime. Thus with Theorem 4.1, the risk functional r is defined on the properly rescaled process, which is close to the desired characterization of the risk through $r(X)$.

Also, let R_{eq} be the class of valid risk functionals for which there exists $u_n \geq 0$ for any

4.3. Limiting distributions of r -exceedances

$r_{\text{eq}} \in R_{\text{eq}}$ and $u \geq 0$ such that

$$\{x \in \mathcal{F}^{\xi, A, B} : r_{\text{eq}}[\{x - r_{\text{eq}}(b_n)\} / r_{\text{eq}}(a_n)] \geq u\} = \{x \in \mathcal{F}^{\xi, A, B} : r_{\text{eq}}(X) \geq u_n\}. \quad (4.10)$$

Then for this class of risk functionals, Theorem 4.1 describes the limiting measure of r_{eq} -exceedances over the threshold u_n of original process X as $n \rightarrow \infty$. The class R_{eq} includes linear functionals and the infimum and supremum. Similarly, for a threshold $u \geq 0$, if a valid risk functional r_{hom} is 1-homogeneous then Theorem 4.1 describes the r_{hom} -exceedances over the threshold $u_n = r_{\text{hom}}(a_n)u$ of the process $X - r_{\text{hom}}(b_n)$, which is just the original process X shifted by a constant. Moreover, in the Fréchet domain of attraction, the sequence b_n can be chosen to equal zero (Resnick, 1987, Proposition 0.2) and thus Theorem 4.1 retrieves the results of Dombry and Ribatet (2015) for homogeneous functionals. Finally, when considering the class R_{lin} of linear risk functionals r_{lin} , Corollary 4.1 gives a factorized representation of the limiting measure Λ .

Corollary 4.1 *Suppose that the conditions of Theorem 4.1 are satisfied, that r_{lin} is a valid linear risk functional with $r(A) = 1$ and $r(B) = 0$, and that for $\xi = 0$, r_{lin} satisfies $\exp\{r_{\text{lin}}(\log x)\} = r_{\text{lin}}(x)$. Then for any $\mathcal{W} \subset \mathcal{S}_r^\xi$, and $r' \in \mathcal{U}_r^\xi$,*

$$\left. \begin{aligned} n \Pr \left[\frac{r_{\text{lin}}(X) - r_{\text{lin}}(b_n)}{r_{\text{lin}}(a_n)} \geq r', \frac{X - r_{\text{lin}}(b_n)}{\|X - r_{\text{lin}}(b_n)\|_{\text{ang}}} \in \mathcal{W} \right], \quad \xi \neq 0 \\ n \Pr \left[\frac{r_{\text{lin}}(X) - r_{\text{lin}}(b_n)}{r_{\text{lin}}(a_n)} \geq r', \frac{X - r_{\text{lin}}(X)}{r(a_n)} \in \mathcal{W} \right], \quad \xi = 0 \end{aligned} \right\} \xrightarrow{\hat{w}} \Lambda^\xi \{\mathcal{U}^\xi(r')\} \times \sigma_{\text{ang}}(\mathcal{W}), \quad (4.11)$$

as $n \rightarrow \infty$ in $M(\mathcal{F}^{\xi, A, B})$, where

$$\Lambda^\xi \{\mathcal{U}^\xi(r')\} = \begin{cases} (1 + \xi r')^{-1/\xi} \Lambda \{ \{y \in \mathcal{F}_0 : r_{\text{lin}}(Ay^\xi) \geq 1\} \}, & \xi \neq 0, \\ \exp(-r') \Lambda \{ \{y \in \mathcal{F}_0 : r_{\text{lin}}(A \log y) \geq 1\} \}, & \xi = 0, \end{cases}$$

with $\mathcal{U}_r^\xi(r') = \{R \in \mathcal{U}_{r_{\text{lin}}}^\xi : R \geq r'\}$, and

$$\sigma_{\text{ang}}(\cdot) = \begin{cases} \frac{\Lambda \{y \in \mathcal{F}_0 : r_{\text{lin}}(Ay^\xi) \geq 1, \{A(y^\xi - 1) + \xi B\} / \|A(y^\xi - 1) + \xi B\|_{\text{ang}} \in (\cdot)\}}{\Lambda \{y \in \mathcal{F} \setminus \{0\} : r_{\text{lin}}(Ay^\xi) > 1\}}, & \xi \neq 0 \\ \frac{\Lambda \{y \in \mathcal{F}_0 : r_{\text{lin}}(y \exp A) \geq 1, A \log y + B - r(A \log y + B) \in (\cdot)\}}{\Lambda \{y \in \mathcal{F} \setminus \{0\} : r_{\text{lin}}(y \exp A) \geq 1\}}, & \xi = 0, \end{cases} \quad (4.12)$$

is a probability measure on $\mathcal{B}(\mathcal{S}_{r_{\text{lin}}}^\xi)$.

The conditions $r(A) = 1$ and $r(B) = 0$ in Corollary 4.1 ensure that the measure of the radial component has zero location and unit scale, but can be relaxed by properly

rescaling Λ^ξ . In general the pseudo-polar decomposition in (4.11) holds only for linear risk functionals, but it can also be derived when $\xi > 0$ to the class of 1-homogenous functionals R_{hom} by imposing $b_n = 0$; see Dombry and Ribatet (2015). Our result might seem less general as Corollary 4.1 holds only for R_{lin} , but this small restriction allows us to link Theorem 4.1 to the classical univariate results of Section 4.2.1, and thus allows greater flexibility for modelling by covering the three regimes of tail decay. Indeed, projections $r_{s_0}(X) = X(s_0)$ for $s_0 \in S$ are an important class of linear functionals that can be used to derive the limiting marginal tail behaviour of X : For any $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$ and a sufficiently high non-negative threshold u , Corollary 4.1 yields

$$\Pr\{r_{s_0}(X) - u > r' \mid r_{s_0}(X) > u\} \approx \begin{cases} \left(1 + \xi \frac{r'}{\sigma(s_0)}\right)^{-1/\xi}, & \xi \neq 0, \\ \exp\left(-\frac{r'}{\sigma(s_0)}\right), & \xi = 0, \end{cases} \quad (4.13)$$

where $\sigma(s_0) > 0$. Equation (4.13) links the functional framework to the classical univariate results described in equation (4.2).

4.4 Generalized r -Pareto processes

In this section, let $\xi \in \mathbb{R}$, $u \geq 0$, $A > 0$ and B denote a tail index, a threshold in \mathcal{U}_r^ξ , and two functions continuous on S .

4.4.1 Definition

For generalized regularly varying stochastic processes, Theorem 4.1 describes the limiting measure of r -exceedances, and can be used to express the limit distribution of conditional r -exceedances $\Pr\{X \in \cdot \mid r(X) \geq u\}$, for some $u \geq 0$, through the generalized r -Pareto process.

Definition 4.1 *Let $r : \mathcal{F} \rightarrow \mathbb{R}$ be a continuous functional satisfying condition (4.8), let $A > 0$ and B be continuous functions over S , and let Λ be a (-1) -homogeneous probability measure on $\mathcal{F} \setminus \{0\}$. The generalized r -Pareto process P with tail index $\xi \in \mathbb{R}$ is a stochastic process on $\{x \in \mathcal{F}^{\xi, A, B} : r(x) \geq u\}$, $u \in \mathcal{U}_r^\xi$, with probability distribution function*

- for $\xi \neq 0$,

$$\Pr\left\{r(P) \geq r', \frac{P}{\|P\|_{\text{ang}}} \in \mathcal{W}\right\} =$$

$$\frac{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + \xi B \right) \geq r', \frac{A(y^\xi - 1) + B}{\|A(y^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq u \right\}}, \quad (4.14)$$

where $r' \in \mathcal{U}^\xi(u)$ and $\mathcal{W} \subset \mathcal{S}_r^\xi$; and

- for $\xi = 0$,

$$\Pr \{ r(P) \geq r', P - r(P) \in \mathcal{W} \} = \frac{\Lambda \left\{ y \in \mathcal{F}_0 : r(A \log y + B) \geq r', A \log y + B - r(A \log y + B) \in \mathcal{W} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r(A \log y + B) \geq u \right\}}, \quad (4.15)$$

where $r' \in \mathcal{U}^\xi(u)$ and $\mathcal{W} \subset \mathcal{S}_r^0$.

Theorem 4.1 implies that the generalized r -Pareto process is the only possible limiting process for r -exceedances of properly rescaled regularly varying stochastic processes. Hence, for any $X \in \text{GRV}(\mathcal{F} \setminus \{0\}, \xi, a_n, b_n, \Lambda)$ and sufficiently large n , the distribution of r -exceedances of the process $\{X - r(b_n)\}/r(a_n)$ over a threshold $u \geq 0$ can be approximated by the corresponding generalized r -Pareto process P . If we further suppose that $r_{\text{eq}} \in R_{\text{eq}}$, for instance if r_{eq} satisfies $r(x + t) = r(x) + t$ for any scalar $t \in \mathbb{R}$, then

$$\Pr \{ X \in \cdot \mid r(X) \geq u_n \} \approx \Pr \{ P \in \cdot \}, \quad (4.16)$$

where u_n was defined in (4.10). Thus, with equation (4.16), the generalized r_{eq} -Pareto process is the only possible limit of large r -exceedances for any generalized regularly varying stochastic process X . For 1-homogeneous functionals r_{hom} , the generalized r_{hom} -Pareto process is in general limited to approximating the r -exceedances of $X - r_{\text{hom}}(b_n)$. When $r_{\text{lin}} \in R_{\text{lin}}$, the pseudo-polar decomposition introduced in Corollary 4.1 gives an equivalent definition of the generalized r -Pareto process.

Definition 4.2 Let r_{lin} be a valid linear risk functional, let $A > 0$ and B be continuous functions on S satisfying $r(A) = 1$ and $r(B) = 0$, and let $u \in \mathcal{U}_r^\xi$. The generalized r_{lin} -Pareto process P with tail index $\xi \in \mathbb{R}$ is a stochastic process on $\{x \in \mathcal{F}^{\xi, A, B} : r_{\text{lin}}(x) > u\}$ with distribution

- for $\xi \neq 0$,

$$\Pr \left\{ r_{\text{lin}}(P) \geq r', \frac{P}{\|P\|_{\text{ang}}} \in \mathcal{W} \right\} = \left(1 + \xi \frac{r' - u}{\sigma} \right)^{-1/\xi} \times \sigma_{A, B}(\cdot), \quad r \geq u, \mathcal{W} \subset \mathcal{S}_{r_{\text{lin}}}^\xi, \quad (4.17)$$

Chapter 4. Functional peaks-over-threshold analysis

where $r' \in \mathcal{U}^\xi(u)$, $\mathcal{W} \subset \mathcal{S}_{r_{\text{lin}}}^\xi$ and $\sigma_{A,B}$ is the angular probability measure for a stochastic process W on $\mathcal{S}_{r_{\text{lin}}}^\xi$ such as that defined in Corollary 4.1;

- for $\xi = 0$,

$$\Pr\{r_{\text{lin}}(P) \geq r', P - r_{\text{lin}}(P) \in \mathcal{W}\} = \exp\left(-\frac{r' - u}{\sigma}\right) \times \sigma_{A,B}(\cdot), \quad (4.18)$$

where $r' \in \mathcal{U}^\xi(u)$, $\mathcal{W} \subset \mathcal{S}_{r_{\text{lin}}}^0$ and $\sigma_{A,B}$ is the angular probability measure for a stochastic process W on $\mathcal{S}_{r_{\text{lin}}}^0$ such as that defined in Corollary 4.1.

The pseudo-polar decomposition in Definition 4.2 reveals the structure of the generalized r_{lin} -Pareto process, which is the product of a radial component, with univariate generalized Pareto distribution representing the intensity of the exceedance, and an angular component driving the dependence structure. With this characterization and when $\xi > 0$, the r -Pareto process of Dombry and Ribatet (2015) is retrieved by setting $A = \xi$, $B = 1$ and $u = \sigma/\xi$, in a similar fashion to the univariate equivalence between generalized and classical Pareto distributions.

4.4.2 Construction and marginal properties

As suggested by Theorem 4.1 and Definition 4.1, the generalized r -Pareto process is closely related to the stochastic process Y_u ($u \geq 0$) defined on

$$\mathcal{A}_r(u) = \begin{cases} \{y \in \mathcal{F}_0 : r\left(A\frac{y^\xi - 1}{\xi} + B\right) \geq u\}, & \xi \neq 0, \\ \{y \in \mathcal{F}_0 : r(A \log y + B) \geq u\}, & \xi = 0, \end{cases}$$

with probability measure $\Lambda(\cdot)/\Lambda\{\mathcal{A}^r(u)\}$ and where Λ is a (-1) -homogenous measure on \mathcal{F}_0 . A standard approach to modelling dependence in multivariate statistics relies on copulas, and requires that all the components of a random vector follow a uniform distribution. Similarly, in extremes the marginal behaviour and dependence structure are handled separately, but contrary to classical copula modelling, the data are often standardized to have a common heavy-tailed distribution such as the unit Pareto. For this reason, Y_u , whose marginals are in the Fréchet domain of attraction with unit tail index, is used as the process of reference. A natural construction of a generalized r -Pareto process derived from Definition 4.1 is

$$P = \begin{cases} A\xi^{-1}(Y_u^\xi - 1) + B, & \xi \neq 0, \\ A \log Y_u + B, & \xi = 0. \end{cases} \quad (4.19)$$

Similarly, following Definition 4.2, if r is linear, then the generalized r_{lin} -Pareto process can also be constructed as

$$P = \begin{cases} R \frac{W}{r(W)}, & \xi \neq 0, \\ R + W, & \xi = 0, \end{cases} \quad (4.20)$$

where R is a univariate generalized Pareto variable with tail parameter ξ , scale $\sigma > 0$ and location $u \geq 0$ and

$$W = \begin{cases} \frac{A(Y_{u'}^\xi - 1) + \xi B}{\|A(Y_{u'}^\xi - 1) + \xi B\|_{\text{ang}}}, & \xi \neq 0, \\ A \log Y_{u'} + B - r(A \log Y_{u'} + B), & \xi = 0, \end{cases} \quad (4.21)$$

for any $u' \geq 0$ that can be chosen for convenience, for instance $u' = 0$. The representations (4.19) and (4.20) are used in Section 4.4.3 to derive simulation algorithms.

What makes generalized r -Pareto processes useful are their marginal properties: consider a location $s_0 \in S$ and a sufficiently high threshold $\nu_0 > 0$ such that the set $\{y \in \mathcal{F}_0 : y(s_0) > \nu_0\}$ is included in $\{y \in \mathcal{F}_0 : r[A(y^\xi - 1)/\xi + B] \geq u\}$. Then

$$\Pr\{P(s_0) > x\} \propto \left\{1 + \xi \frac{x - u_0}{\sigma(s_0)}\right\}^{-1/\xi}, \quad x \geq u_0, \quad (4.22)$$

where \propto stands for proportionality, $\sigma(s_0) = A(s_0) + \xi\{u_0 - B(s_0)\}$ and the threshold u_0 equald $\{1 + \xi A(s)^{-1} \nu_0\}^{1/\xi}$. This means that using equation (4.22), the conditional distribution of exceedances above the threshold u_0 is generalized Pareto. The univariate distribution of the aggregated process $\Pr\{r(P) > r'\}$, $r' \in \mathcal{U}^\xi(u)$ is not in general available in closed form but is obtainable by numerical evaluation of $\Lambda\{\mathcal{A}_r(r')\} / \Lambda\{\mathcal{A}_r(u)\}$. When $r_{\text{lin}} \in R_{\text{lin}}$, the marginal distribution of $r_{\text{lin}}(P)$ simplifies to a generalized Pareto distribution $H_{\xi-1, \sigma}(x - u)$ and the generalized r_{lin} -Pareto process P is threshold-stable, i.e.,

$$\Pr\{P \in [r' + u', \infty) \times \mathcal{W} \mid r_{\text{lin}}(P) \geq u'\} = \Pr\{\kappa(u') \times P \in [r', \infty) \times \mathcal{W}\}, \quad r' \geq 0, \mathcal{W} \subset \mathcal{S}_r^\xi. \quad (4.23)$$

where $\kappa(u') = \{\sigma + \xi(u' - u)\}/\sigma$, assuming $u' \geq u$ and $\xi \neq 0$. Thus the distribution of r_{lin} -exceedances, similarly to the univariate case, is stable in distribution up to a scaling parameter κ that is a function of the new threshold ν .

4.4.3 Simulation

The process Y_u defined in Section 4.4.2, is key to the construction of generalized r -Pareto processes and is also central to their simulation. Thus, it is necessary to be

Chapter 4. Functional peaks-over-threshold analysis

able to draw samples from Y_u , which is usually possible for a restricted class of risk functionals. The case $r(x) = \|x\|_1$ with $\xi = 1$ is usually convenient because $E(\|Y_1\|_1)$ is constant and independent of the dependence structure. From this observation, simple simulation algorithms that rely on the pseudo-polar decomposition in equation (4.5) have been developed; see for example Asadi et al. (2015) for the Brown–Resnick model. If a simulation algorithm for Y_u with risk functional $r(X) = \|X\|_1$ is available, then we can sample from the angular component W of Y_u whose probability measure $\sigma_{\|\cdot\|_1}$ is defined in equation (4.5). We generalize the principle described in de Fondeville and Davison (2018, Section 2.3) to develop an accept-reject algorithm for the generalized r -Pareto process. We detail the case $\xi \neq 0$, but generalization for $\xi = 0$ is straightforward.

Let r be a risk functional and let P be the corresponding generalized r -Pareto process with limiting measure Λ , tail index $\xi \in \mathbb{R}$, scale function $A > 0$ and location function B . We suppose that we have a threshold $u' > 0$ such that

$$\left\{ y \in \mathcal{F}_0 : r\left(A \frac{y^\xi - 1}{\xi} + B\right) > u \right\} \subset \{y \in \mathcal{F}_0 : \|y\|_1 > u'\}. \quad (4.24)$$

We stress that u' is a deterministic quantity that can be found analytically. Algorithm 1 gives an accept-reject algorithm for simulation of P when an algorithm for Y_u with the L_1 -norm is available.

Algorithm 1: Simulation of generalized r -Pareto process, P

Set $Y_u = 0$;

while $r[A\xi^{-1}\{(Y_{u'})^\xi - 1\} + B] < u$ **do**

generate a unit Pareto variable R on $[u', \infty)$;

generate W on $\mathcal{S}_{\|\cdot\|_1} = \{y \in \mathcal{F}_0 : \|x\|_1 = 1\}$ with probability measure $\sigma_{\|\cdot\|_1}$ defined in equation (4.5);

set $Y_{u'} = RW$;

end

Set $P = A\xi^{-1}\{(Y_{u'})^\xi - 1\} + B$;

The efficiency of Algorithm 1 is determined by the capacity to find the largest possible threshold u'_{sup} such that (4.24) is satisfied, and its rejection rate is given by the ratio of the measures of the two corresponding sets. If r is linear, then we suppose that we have a threshold $u' > 0$ such that

$$\left\{ y \in \mathcal{F}_0 : r_{\text{lin}}(Ay^\xi) \geq 1 \right\} \subset \{y \in \mathcal{F}_0 : \|y\|_1 > u'\}.$$

Algorithm 2: Simulation of generalized r_{lin} -Pareto process, P

 Set $Y_1 = 0$;

while $r_{\text{lin}} \{A(Y_1)^\xi\} < 1$ **do**

 generate a unit Pareto variable R on $[u', \infty)$;

 generate W on $\mathcal{S}_{\|\cdot\|_1} = \{y \in \mathcal{F}_0 : \|x\|_1 = 1\}$ with probability measure $\sigma_{\|\cdot\|_1}$ defined in equation (4.5);

 set $Y_1 = RW$;

end

 set $W_2 = \frac{A\{Y_1\}^\xi - 1 + \xi B}{\|A\{Y_1\}^\xi - 1 + \xi B\|_{\text{ang}}}$;

 generate a generalized Pareto random variable $R_2 \sim H_{(\xi, \sigma)}(x)$;

 set $P = (R_2 + u)W_2 / r_{\text{lin}}(W_2)$;

While Algorithm 2 is more complex than Algorithm 1, it also allows the simulation of events for a given intensity of the risk functional, which is not possible in general. Indeed, the first steps of Algorithm 2 describe how to sample from W_2 , i.e., how to simulate from the spectral measure σ_{ang} defined in Corollary 4.1.

4.4.4 Link to max-stable processes

In the univariate theory, marginal assumptions given by equation (4.7) are equivalent to convergence of rescaled block maxima toward the generalized extreme value distribution, i.e.,

$$\lim_{n \rightarrow \infty} \Pr \left\{ \frac{\max_{i=1, \dots, n} X_i(s) - b_n(s)}{a_n(s)} \leq z \right\} = \begin{cases} \exp \left[- \left\{ 1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right\}^{-1/\xi} \right], & \xi \neq 0, \\ \exp \left[- \left\{ \exp \left(\frac{z - \mu}{\sigma} \right) \right\} \right], & \xi = 0, \end{cases}$$

for $z \in \mathbb{R}$ if $\xi = 0$ or on $\{z \in \mathbb{R} : 1 + \xi(z - \mu)/\sigma > 0\}$ if $\xi \neq 0$. Similarly, it is possible to link generalized r -Pareto processes to the functional extensions of the generalized extreme value distributions known as max-stable processes. The representation of these processes is not unique; we use that introduced by de Haan (1984) which relies on the Poisson point processes $(R_n, W_n)_{n=1, \dots}$ on $(0, \infty) \times \mathcal{S}_{\|\cdot\|_1}$ with intensity measure $r^{-2} dr \times \sigma_{\|\cdot\|_1}(dw)$, where $\sigma_{\|\cdot\|_1}$ is defined in equation (4.5). Then the process

$$M(s) = \begin{cases} \max_{n \geq 1} A(s) \frac{\{R_n W_n(s)\}^\xi - 1}{\xi} + B(s), & \xi \neq 0, \\ \max_{n \geq 1} A(s) \log \{R_n W_n(s)\} + B(s), & \xi = 0, \end{cases} \quad s \in S, \quad (4.25)$$

Chapter 4. Functional peaks-over-threshold analysis

is max-stable with exponent measure $\Lambda \circ T_{\xi,A,B}(\cdot)$ (Resnick, 1987, Proposition 3.7), where $T_{\xi,A,B}(z)$ is the non-atomic map

$$T_{\xi,A,B}(z) = \begin{cases} z \rightarrow \{1 + \xi(z - B)/A\}^{1/\xi}, & \xi \neq 0, \\ z \rightarrow \exp\{(z - B)/A\} & \xi = 0. \end{cases}$$

With this notation, the finite-dimensional distribution function of M at locations $s_1, \dots, s_L \in S$ is

$$\Pr\{M(s_l) < z_l, l = 1, \dots, L\} = \exp \left[-\Lambda \circ T_{\xi,A(s_l),B(s_l)} \left\{ (0, z_l]_{l=1,\dots,L}^c \right\} \right],$$

where in the exponential term we recognize the measure of a generalized r -Pareto process with risk functional $r(x) = \max_{l=1,\dots,L} x(s_l)$. With this representation, the process M is constituted by infinitely many single events whose intensity is Poisson-distributed, and the r -exceedances distribution of these events above a threshold $u \geq 0$ corresponds to a generalized r -Pareto process. When r is linear, this link is even clearer, as we can replace the Poisson process in equation (4.25) by the pseudopolar representation $R_n W_n / r(W_n)$ introduced in equation (4.20), whose measure is $\sigma^{-1} \{1 + \xi(r' - u)/\sigma\}^{-1/\xi-1} dr' \times \sigma_{\text{ang}}(dw)$. In short, the generalized r -Pareto process arises as the distribution of r -exceedances of the single events constituting a max-stable process. The Poisson intensity, which is necessary to model the occurrence of single events in the max-stable process, disappears through the r -exceedance conditioning, and thus the distribution of the number of exceedances can be chosen independently of the generalized r -Pareto process model. Further details are given in Section 4.5.

4.5 Statistical inference

Let $X_n \in \mathbb{R}_+^L$ ($n = 1, \dots, N$) be realizations of a generalized regularly varying stochastic process X sampled at locations $s_1, \dots, s_L \in S$ and let r be a valid risk functional, as in Section 4.3. Here we explain how to fit generalized r -Pareto process as an approximation for the distribution of r -exceedances of X over a threshold $u > 0$. Theorem 4.1 suggests that, from a theoretical point of view, the choice of risk functional should not impact the model parameters, but it affects what data are considered extreme especially in the presence of a mixture in the tail behaviour, as illustrated by Figure 4.1. Designing a risk functional r enables us to focus on one component of this mixture by incorporating field-specific expertise, while improving sub-asymptotic behaviour by fitting the model using only those observations closest to the chosen type of extreme event. Thus, suppose that we have a risk functional r , for which exceedances

correspond to only one physical process, such as cyclonic rainfall, and that under this hypothesis it is reasonable to consider a tail index $\xi \in \mathbb{R}$ constant over space.

More specifically, we assume that the marginal distributions of X belong to a location-scale family, that is, for some distribution function F and continuous $A : S \rightarrow (0, \infty)$, $B : S \rightarrow \mathbb{R}$ we have

$$\mathbb{P}\{X(s) \leq x\} = F\left\{\frac{x - B(s)}{A(s)}\right\},$$

for any fixed $s \in S$ and where F satisfies equation (4.1). This implies that the normalizing functions satisfy

$$a_n(s) = A(s)a(n), \quad b_n(s) = B(s) + A(s)b(n), \quad s \in S,$$

where $a(n) > 0$ and $b(n) \in \mathbb{R}$, ensuring the asymptotic decomposition required in condition (4.8).

We impose a parametric structure on the marginal scale and location parameters, i.e., on the unknown functions A and B , and on the extremal dependence of X . For the marginal distributions, we assume that A and B belong to parametric families of functions $\{A_{\theta_A}; \theta_A \in \Theta_A\}$ and $\{B_{\theta_B}; \theta_B \in \Theta_B\}$ where Θ_A and Θ_B are appropriate subsets of \mathbb{R}^{d_A} and \mathbb{R}^{d_B} . Similarly, we suppose that the limiting measure Λ belongs to a parametric class $\{\Lambda_{\theta_\Lambda}; \theta_\Lambda \in \Theta_\Lambda\}$ with $\theta_\Lambda \in \mathbb{R}^{d_\Lambda}$. In the following, we describe a method to jointly infer the complete parameter vector

$$\theta = \{a(n), b(n), \theta_A, \theta_B, \theta_W\} \in (0, \infty) \times \mathbb{R} \times \Theta_A \times \Theta_B \times \Theta_W.$$

Identifiability issues may arise with the parametric models for A and B , which thus need to be carefully designed, for instance by assuming $r(A) = 1$ and $r(B) = 0$; see Chapter 3.

Statistical inference for r -exceedances of stochastic processes $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$ is based on the approximation

$$\begin{aligned} & \Pr\left\{\frac{X - r(b_n)}{r(a_n)} \in \mathcal{A}\right\} \\ &= \Pr\left[r\left\{\frac{X - r(b_n)}{r(a_n)}\right\} \geq u\right] \times \Pr\left[\frac{X - r(b_n)}{r(a_n)} \in \mathcal{A} \mid r\left\{\frac{X - r(b_n)}{r(a_n)}\right\} \geq u\right], \\ &\approx \Pr\left[r\left\{\frac{X - r(b_n)}{r(a_n)}\right\} \geq u\right] \times \Pr(P \in \mathcal{A}), \end{aligned} \tag{4.26}$$

with $\mathcal{A} \subset \{x \in \mathcal{F}^{\xi, A, B} : r(x) \geq u\}$, and where $1\{\cdot\}$ is the indicator function and the probability of observing the event $\{r(X) > u\}$ is replaced by the distribution of the

Chapter 4. Functional peaks-over-threshold analysis

binary variable $\{r(X) \geq u\}$. Thus, following (4.26), a natural choice for fitting a generalized r -Pareto process to r -exceedances over the threshold $u \geq 0$ is to maximize the log-likelihood function

$$\mathcal{L}_{\text{Thres}}(\theta) = \log \Pr(N_u = n_u) + \sum_{n \in K_u} \log f_u^r(x_n),$$

where N_u is the random number of exceedances, K_u is the index set of r -exceedances over u , i.e., $\{n \in 1, \dots, N : r(X_n) \geq u\}$, with n_u being its cardinality, and f^r is the finite-dimensional density function of the generalized r -Pareto process with threshold $u \in \mathcal{U}_r^\xi$ sampled at locations s_1, \dots, s_L , i.e.,

$$f_u^r(x) = \frac{\lambda \left\{ \left(1 + \xi \frac{x-B}{A} \right)^{1/\xi} \right\}}{\Lambda \{ \mathcal{A}_r(u) \}} \prod_{l=1}^L A(s_l)^{-1} \left\{ 1 + \xi \frac{x(s_l) - B(s_l)}{A(s_l)} \right\}^{1/\xi-1}, \quad x \in \mathcal{A}_r(u), \quad (4.27)$$

where $\Lambda \{ \mathcal{A}_r(u) \} = \int_{\mathcal{A}_r(u)} \lambda(x) dx$. The derivation of equation (4.27) and the expression for linear risk functionals can be found in Appendix D.6. A model needs to be specified for the random number of exceedances N_u appearing in the log-likelihood function: Wadsworth and Tawn (2014) and Engelke et al. (2015) use a Poisson distribution, inspired by Poisson point processes, yielding log likelihood

$$\begin{aligned} \mathcal{L}_{\text{Poiss}}(\theta) = & -\log n_u! + n_u \log \Lambda \left[\mathcal{A}_r^n \left\{ \frac{u - r(b_n)}{r(a_n)} \right\} \right] - \\ & \Lambda \left[\mathcal{A}_r^n \left\{ \frac{u - r(b_n)}{r(a_n)} \right\} \right] + \sum_{m \in K_u} \log f_u^r \left\{ \frac{x_m - r(b_n)}{r(a_n)} \right\}, \end{aligned} \quad (4.28)$$

with

$$\mathcal{A}_r^n(u) = \left\{ \left(1 + \xi \frac{x - r(b_n) - r(a_n)B}{r(a_n)A} \right)^{1/\xi} \in \mathcal{F}_0 : r \left(\frac{x - r(b_n)}{r(a_n)} \right) \geq u \right\}.$$

The Pareto methodology is more flexible than the Poisson point process approach because various models for the distribution of the number of exceedances N_u can be considered. For instance, when supposing that the number of exceedances is fixed, choosing a binomial distribution yields

$$\begin{aligned} \mathcal{L}_{\text{Bin}}(\theta) = & (n - n_u) \log \left(1 - \frac{1}{n} \Lambda \left[\mathcal{A}_r^n \left\{ \frac{u - r(b_n)}{r(a_n)} \right\} \right] \right) + \\ & n_u \log \left(\frac{1}{n} \Lambda \left[\mathcal{A}_r^n \left\{ \frac{u - r(b_n)}{r(a_n)} \right\} \right] \right) + \sum_{m \in K_u} \lambda^r \left\{ \frac{x_m - r(b_n)}{r(a_n)} \right\}. \end{aligned} \quad (4.29)$$

Equation (4.29) slightly differs from the formulation in Thibaud and Opitz (2015) where the probability of exceedance is $\Pr[1\{r(X) > u\} = 1] = \Lambda [\mathcal{A}_r^n(u)]$ instead of $n^{-1} \Lambda \{ \mathcal{A}_r^n(u) \}$. By contrast, our expression yields a valid probability distribution for

any $u > n^{-1}$ and can be linked to the Poisson point process model; see Appendix D.7.

For likelihood-based inference, maximizing equations (4.28) and (4.29) will be numerically unstable if K_u depends on $r(a_n)$ and $r(b_n)$, which are unknown in general. One way to ensure numerical stability is to focus on risk functionals for which we can find $u' \geq 0$ such that

$$r(x) \geq u'_n \iff r\left\{\frac{x - r(b_n)}{r(a_n)}\right\} \geq u, \quad (4.30)$$

where \iff stands for equivalence. This property is satisfied by any functional such that $r(x + t) = r(x) + t$ for any scalar $t \in \mathbb{R}$, and then $u'_n = r(a_n)u + r(b_n)$. In this case, as explained in Section 4.4, r -exceedances of X above a large enough threshold $u' \geq 0$ can be approximated by a generalized r -Pareto process, so log-likelihood functions (4.28) and (4.29) can be directly exploited by replacing u by $\{u'_n - r(b_n)\}/r(a_n)$.

When equation (4.30) is not satisfied, for instance with the functional $\{\int_S X(s)^2 ds\}^{1/2}$, a two-step procedure might be used to first estimate $a(n)$, A , $b(n)$ and B , and then to find the remaining components of the vector θ . Other procedures might be considered depending on the choice of functional and model hypotheses.

Lastly, applying the risk functional directly to X might not always be desirable. For instance, with $\max_{s \in S} X(s)$, extreme events at locations with the largest scales and locations might dominate the set of exceedances. This effect can be removed by using,

$$r_q(x_n) = \max_{s \in S} \left\{ \frac{x_n(s)}{u^q(s)} \right\} \geq 1,$$

where $u^q(s)$ denotes the empirical q -quantile at location $s \in S$. The marginal parameters a_n and b_n do not influence r_q -exceedances, and thus the corresponding set $\mathcal{A}_{r_q}^n(u)$ in (4.28) is stable and equals

$$\mathcal{A}_{r_q}^n(u) = \left(0, \left\{ 1 + \xi \frac{u^q(s_{1:L}) - b(n) - a(n)B(s_{1:L})}{a(n)A(s_{1:L})} \right\}^{1/\xi} \right)^c,$$

where $s_{1:L} = s_1, \dots, s_L$ and \mathcal{A}^c denotes the complement of \mathcal{A} in $r^L \setminus \{0\}$. Maximum likelihood estimators based on equations (4.28) and (4.29) have been studied for specific choices of risk functional and were found to perform poorly in practice because the limiting process is used as an approximation for finite n , yielding a naturally misspecified model (Engelke and Malinowski, 2014; Huser et al., 2016). Classical maximum likelihood estimators satisfy the Cramer–Rao lower bound and are thus efficient but can be sensitive to model misspecification. For this reason more robust alternatives have been considered, using censoring of low components (e.g. Wadsworth and Tawn,

2014), composite methods (Padoan et al., 2010; Huser and Davison, 2013; Castruccio et al., 2016) and M-estimation using pairwise tail indexes (Einmahl et al., 2016a,b). All are more robust with regard to sub-asymptotic observations, but they work for specific risk functionals and are dimensionally limited, either by the computational burden due to the numerical evaluation of the scaling constant $\Lambda\{\mathcal{A}_r^n(u)\}$ and the censoring, or, for pairwise procedures, by combinatorial considerations. For the Brown–Resnick and extremal t models, efficient algorithms for censored likelihood are available (de Fondeville, 2016) and tractable for hundreds of dimensions. Gradient scoring (de Fondeville and Davison, 2018) can be applied to a large class of risk functionals and avoids the computation of $\Lambda\{\mathcal{A}_r^n(u)\}$, making inference tractable for thousands of dimensions: for the Brown–Resnick case, its numerical complexity is driven only by that of matrix inversion.

4.6 Model validation

Suppose that we have an estimate $\widehat{\theta}$ of the complete vector of parameters introduced in Section 4.5, and a measure of its uncertainty, obtained for instance using a block jackknife or bootstrap, and that we wish to check the quality of the model. For each sampled location s_1, \dots, s_L , we can compare the observations with the theoretical quantiles of the marginal model. Let $u_q(s_l)$ denote the q^{th} empirical quantile of r -exceedances at location s_l , i.e., estimated using only observations satisfying $r(X_n) \geq u$, and let n_q denote the number of observations exceeding $u_q(s_l)$. Following equation (4.13), we have

$$\Pr\{X(s_l) - u_q(s_l) \geq x \mid X(s_l) \geq u_q(s_l)\} \approx H_{\widehat{\xi}, \widehat{\sigma}(s_l)}(x), \quad x \geq 0,$$

with $\widehat{\sigma}(s_l) = \widehat{a(n)}\widehat{A}(s_l) + \widehat{\xi}\{u_q(s_l) - \widehat{B}(s_l) - \widehat{b(n)}\}$. Then we use quantile-quantile plots to check the quality of the marginal fit. Confidence intervals can be obtained by resampling: we draw N_s samples of size N_q $(Z_1^1, \dots, Z_{N_q}^1), \dots, (Z_1^{N_s}, \dots, Z_{N_q}^{N_s})$ from the fitted distribution and let $Z_{(n)}^1, \dots, Z_{(n)}^{N_s}$ denote the n^{th} order statistic of each sample. A 95% confidence interval for the generalized Pareto fit is then defined as the 2.5 and 97.5 empirical percentiles of the sets $\{Z_{(n)}^m : m = 1, \dots, N_s\}$. When the estimator used to obtain $\widehat{\theta}$ is asymptotically normal, the uncertainty of the model parameters can be taken into account by drawing the N_s samples from different generalized Pareto distributions whose parameters are normally distributed with means $\widehat{\xi}$ and $\widehat{\sigma}(s_l)$ respectively and standard deviations corresponding to the uncertainty of the vector $\widehat{\theta}$; strictly positive parameters being handled on the log-scale.

A check of the dependence model can be based on a generalization of the extremogram

(Davis and Mikosch, 2009) introduced in de Fondeville and Davison (2018), i.e.,

$$\pi(s_l, s_k) = \Pr[X(s_l) > u_q(s_l) \mid \{X(s_k) > u_q(s_k)\} \cap \{r(X) > u\}], \quad l, k = 1, \dots, L.$$

If the model is stationary and isotropic, π depends only the distance $h = |s_l - s_k|$ between the locations s_l and s_k . The theoretical values of the model are then compared to empirical estimates of π and summarized using what we call an extremogram cloud: a graphical diagnostic, which displays π as a function of the distance, and if relevant, the orientation of the pair (s_l, s_k) .

Model comparison can be performed using classical likelihood criteria, such as the Akaike Information Criterion (AIC), the Composite Likelihood Information Criterion (CLIC) (Davison and Gholamrezaee, 2012) or the Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007). Formal testing is possible for nested models (de Fondeville and Davison, 2018).

4.7 Discussion

Peaks-over-threshold analysis is widely used for modelling tails of univariate distributions through the generalized Pareto distribution, but natural hazards cannot be studied using only univariate results. In this chapter, we have extended peaks-over-threshold analysis to extremes of functional data. Exceedances are defined using a real-valued functional r , and modelled with the generalized r -Pareto process, a functional generalization of the generalized Pareto distribution, covering the three possible tail decay regimes. This family appears as the limit for r -exceedances of a properly rescaled process. We derive construction rules for generalized r -Pareto processes, give simulation algorithms and highlight their link to max-stable processes. Finally we discuss inference procedures and model validation.

The strength of the theoretical results developed in this chapter depends on the relevance of the properties satisfied by r . The class of linear functionals is particularly attractive because in this case the risk is directly defined on the original process and Corollary 4.1 gives the limit distribution of large r_{lin} -exceedances of X . Also, the generalized r_{lin} -Pareto process can be factorized into two components: a radial part measuring the intensity of the excess and an angular component modelling the dependence. This decomposition enables simulations for fixed intensities, i.e. for determined values of $r(X)$, and allows the generation of catalogues of extreme events for fixed return periods; such events can later be used as input for stress tests either on human infrastructure or insurance portfolios. In Chapter 5, we illustrate this methodology by developing a stochastic weather generator for extreme windstorms

Chapter 4. Functional peaks-over-threshold analysis

over Europe.

While the class R_{lin} might seem restrictive, for spatial applications it can be combined with tools from image processing such as Fourier or wavelet transforms, that have been successfully used to classify large and complex datasets of images. This chapter opens the development of flexible, and if possible linear, risk functionals able to discriminate between different meteorological phenomena.

5 Statistical modelling of extreme wind-storms over Europe

5.1 Introduction

On 25 January 1990, the wind storm Daria, also known as the ‘Burns Day Storm’ as it started on the birthday of Scottish poet Robert Burns, struck the United Kingdom. Daria is famous for being one of the most severe extra-tropical cyclones in this region. During the two days where the storm was active, 97 deaths were reported and damage was valued at around 8.2 billion dollars. The strongest gusts were measured to be 170 km.h^{-1} , a speed equivalent to a category 1 hurricane. Figure 5.1 shows the maximum speed over the past $3h$ hours of the wind gusts sustained for at least 3s. The selected time steps correspond to the 24 hours during which the storm was at its peak.

About 10 years later, on 26 December 1999, the storm Lothar swept across western and central Europe during a period of 36 hours. A wind speed of 169 km.h^{-1} was recorded in Paris, and at the summit of the ‘Dole’ in Switzerland, the weather station reported a maximum wind gust of 201.2 km.h^{-1} . Lothar was classified as a category 2 cyclone, and caused 8 billion dollars loss and more than 100 deaths.

Estimating the risk linked to such extreme natural hazards has become a major question in recent decades, because of the possible influence of global warming. Even if the influence of human activity on the climate has been established, according to the IPCC (Pachauri et al., 2014), its impact on specific types of events is much less certain. To issue long-range projections or to minimize risks linked to wind storms, both climatologists and insurers want to better understand the extremal behaviour of weather events.

In this chapter, we use the theory presented in Chapter 4 to develop a stochastic weather generator of extreme wind storms over Europe. The model can create catalogues of wind storms with unobserved shapes and tracks, and potentially unobserved

Chapter 5. Statistical modelling of extreme windstorms over Europe

intensities. These catalogues can then be used as ‘stress tests’ for physical infrastructure or insurance portfolios.

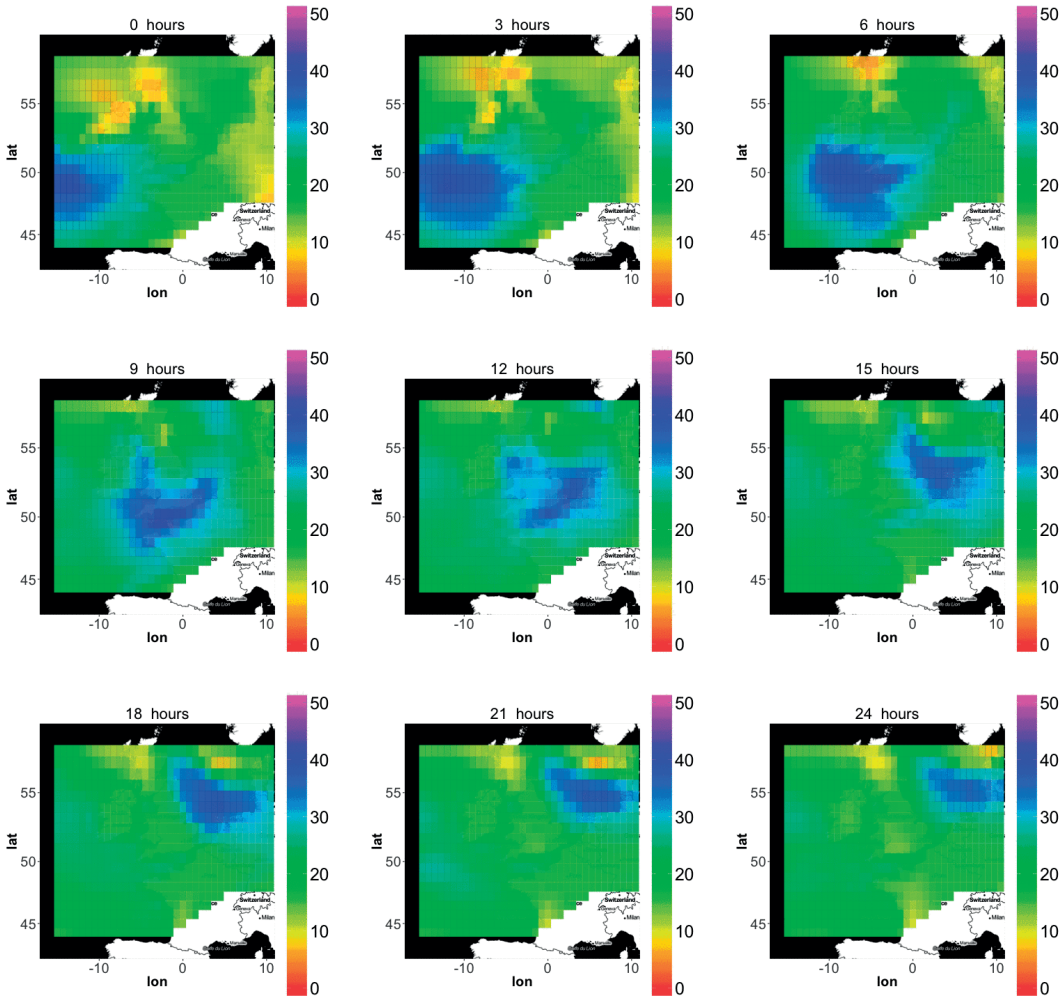


Figure 5.1 – Maximum speed (m.s^{-1}) over the past 3h hours of the wind gusts sustained for at least 3s from ERA-Interim reanalysis during the peak of wind storm Daria, which swept over Europe during January 1990.

5.2 Risk estimation for extreme windstorms

Up until now, risk estimation for extreme wind storms has been limited to straightforward exploitation of historical catalogues (Haylock, 2011; Pinto et al., 2012). These sources of data are limited because records usually do not exceed 40 to 50 years in length. In order to increase the size of those catalogues, one approach has been to statistically perturb the wind field intensity, shape and location (Hall and Jewson, 2008). A second method detects extreme storms in reanalysis data and climate model projections (Della-Marta et al., 2010). With this methodology, the same events may be used twice or more but the climatological indexes always differ slightly because of different hypotheses and approximations between models. Uncertainties and bias linked to this approach are likely to be large and difficult to estimate. Moreover recent studies on climatological projections stressed the inability of these models to accurately reproduce the behaviour of extreme events (e.g., Weller et al., 2013). More recently, an approach has been developed to create new events from historical catalogues by reordering time steps based on spatial analogues (Yiou, 2014). All these methods generate extreme wind storms with a tail behaviour that is not mathematically justified.

Extreme value theory provides a theoretical basis to study and develop models for tail distributions. Della Marta and Mathis (2008) performed a peaks-over-threshold analysis on univariate summaries characterizing extreme wind storms, but they ignore spatial dependence. Ferreira and de Haan (2014) developed a method to upscale historical wind storm records to higher intensities using Pareto processes. Economou and David (2014) adapted Bayesian hierarchical models to extra-tropical cyclones, but in this case dependence is included using covariates such as mean sea level pressure, which limits the capability of the model to generate new patterns and intensities.

We propose a different approach based on generalized r -Pareto processes, which generalizes the Della Marta and Mathis (2008) approach to a functional setting and allows not only local risk estimation but also the simulation of completely new and spatially and temporally consistent extreme events.

5.3 Data set and region of study

To build our stochastic weather generator, we follow the methodology of the extreme wind storms (XWS) catalogue (Roberts et al., 2014). This database, publicly released in 2013, is the first, and still as of today the only, one of its kind. It provides historical records on the 50 most extreme storms over Europe between 1979 and 2012; more

precisely it contains maps of 72-hour maximum wind gusts over northern Europe. In this catalogue, the notion of extreme storm is chosen to focus on events with high impact on infrastructure. Indeed storms with the overall highest maximum wind speed are not necessarily the ones that induce the most damage, as their trajectories may not include inhabited areas. The difficulty is thus to define a meaningful univariate summary to characterize such impactful events. Roberts et al. (2014) compare methods for storm tracking and build an index to characterize their catalogue, which motivates our choice of risk functional in Section 5.4.

The XWS catalogue tracks storms in the ERA-Interim reanalysis (Dee et al., 2011), a real-time climate model whose records start in 1979. This catalogue provides time series for many climatological indexes, and in particular for the maximum speed over the past 3 hours of the wind gusts sustained for at least 3s. The model is run every six hours on a worldwide grid, whose cells are squares with a side that can be chosen between 3° and 0.125° ; 0.75° is the native grid size, the other resolutions being obtained by interpolation. In addition to the 6-hourly fields obtained by data assimilation, i.e., by constraining the grid values to station measurements, 256-hour forecasts are generated each day at 00UTC and 12UTC, and can be used to obtain a 3-hourly database. Most European winter storms are limited in time and evolve quickly, so such a fine time resolution is necessary to accurately detect them.

Our study focuses on western Europe, i.e., an area with boundaries N57.75, S44.25, E25 and W10.5, from which the mountainous regions such as the Pyrenees and the Alps are removed; see Figure 5.2. The reanalysis model is known to have a systematic bias over regions with high variations of altitude (Donat et al., 2011). Similarly to the XWS catalogue methodology, we use the maximum of wind gusts sustained for at least 3s since previous post-processing from the ERA-Interim reanalysis and use the forecasts to obtain a 3-hourly database. We retrieve the data with the native resolution of 0.75° , yielding a grid with 605 cells. The study is restricted to October–March in order to avoid any seasonality effects, as extra-tropical wind storms over Europe occur only during the winter. For illustration, Figure 5.1 shows the 3s maximum wind gust during the storm Daria. To give an idea of the severity of this extreme event, damaging windspeeds are considered to start at 25 m.s^{-1} (Roberts et al., 2014).

5.4 Storm definition and frequency modelling

Following the comparison of Roberts et al. (2014), we define a storm as the exceedance of the spatial mean over a region with very dense human infrastructure during a 24-hour temporal frame. The spatio temporal process $X(s, t)$, $s \in S$ and $t \in T$, represents

5.4. Storm definition and frequency modelling

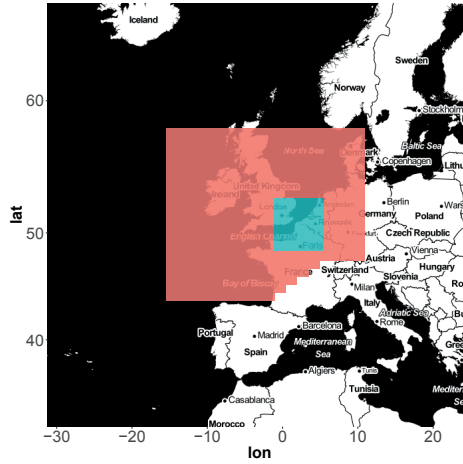


Figure 5.2 – Area of study (coloured cells) for modelling extreme wind storms over Europe. Mountainous regions were removed to avoid the systematic bias of the reanalysis model. Green cells represent the selected region whose density of human infrastructure is high.

the wind field over the region S , here a subset of Europe, and over the period $T = [1979; 2014]$. Our mathematical definition of risk functional r at a time $t \in T$ is

$$r(X)(t) = \max_{t'=-12,9,\dots,12} |S_{\text{PLBA}}|^{-1} \int_{S_{\text{PLBA}}} X(s, t+t') ds, \quad x \in C(S, \mathbb{R}),$$

where S_{PLBA} refers to the green region in Figure 5.2, which includes Paris, London, Brussels and Amsterdam. To suppress the effect of temporal clustering, we center the time frame on the spatial mean maxima and keep only events that are at least 48 hours apart, yielding $n = 1561$ observations. For illustration, storm Daria corresponds to an intensity $r(x) = 32.1 \text{m.s}^{-1}$.

The stochastic weather generator that we develop is based on the approximation (4.26), which, with the properties of the risk functional and the same notations, simplifies to

$$\Pr\{X \in \cdot\} \approx \Pr[1\{r(X) \geq u_n\} = 1] \times \Pr(P \in \cdot), \quad (5.1)$$

where $u_n = ur(a_n) + r(b_n)$. In (5.1) three components must be modelled: the margins, which include a tail index ξ and the functions a_n and b_n ; the distribution of $1\{r(X) \geq u_n\}$; and the dependence of the generalized r -Pareto process P .

To model the probability of observing an r -exceedance, for simplicity we choose $u = 0$ such that r -exceedances are defined as events for which $r(X) \geq r(b_n)$. In this case, a high quantile of the random variable $r(X)$ is a natural choice for $u_n = r(b_n)$, as it

ensures if the quantile level is sufficiently high that n is large. In order to include most of the XWS storms in our set of exceedances, we take $u_n = q_{0.96}\{r(X)\} = 24\text{m}\cdot\text{s}^{-1}$, yielding 63 events for the period 1979 to 2016. This quantile level also corresponds to a stability region in the estimated tail index of the univariate variable $r(X)$. The risk functional, the r -exceedances and the XWS storms are shown in Figure 5.3. The 63 events, depicted by the red dots, coincide with most of the wind storms from the XWS catalogue represented by the vertical lines: the exceedances of the risk functional r successfully characterise extreme wind storms hitting Europe in the region S_{PLBA} . The events from the catalogue that do not match an exceedance mostly pass over southern regions of Europe and thus are logically not captured by r .

Figure 5.3 also shows that the temporal distribution of $1\{r(x) \geq u\}$ is not stationary, probably due to the influence of external factors. Several studies point out the importance of climatic circulation patterns such as the North Atlantic Oscillation index (NAO) (Donat et al., 2010; Pfahl, 2014) in the frequency of extreme wind storms. Thus to include the influence of potential covariates, we choose to model the distribution of $1\{r(x) \geq u\}$ with logistic regression.

To compute the NAO index, the 3-hourly mean sea level pressure is extracted from the ERA-interim reanalysis and the North Atlantic Oscillation index is computed using its definition based on Empirical Orthogonal Functions (EOF) (Blessing et al., 2005), i.e., the first eigenvalue of the mean sea level pressure anomaly at time t . We proceed similarly to compute the Antarctic Oscillation (AAO) index and create indexes for the temperature anomaly based on EOF analysis. Time is also considered as a potential covariate.

An analysis of deviance reveals that the NAO index and the first and third eigenvalues of the temperature anomaly have a significant influence on the occurrence of winter storms at the 0.1% confidence level. Figure 5.4 shows a yearly summary of the model, and more detailed plots can be found in Appendix E.

5.5 Marginal model

Following equation (5.1), fitting the marginal model is equivalent to the estimation of a tail index $\xi \in \mathbb{R}$ and the functions a_n and b_n . To ensure condition (4.8), we define the functions $A > 0$ and $B \in C(S, \mathbb{R})$, such that

$$a_n(s) = a(n)A(s), \quad b_n(s) = b(n) + a(n)B(s), \quad s \in S,$$

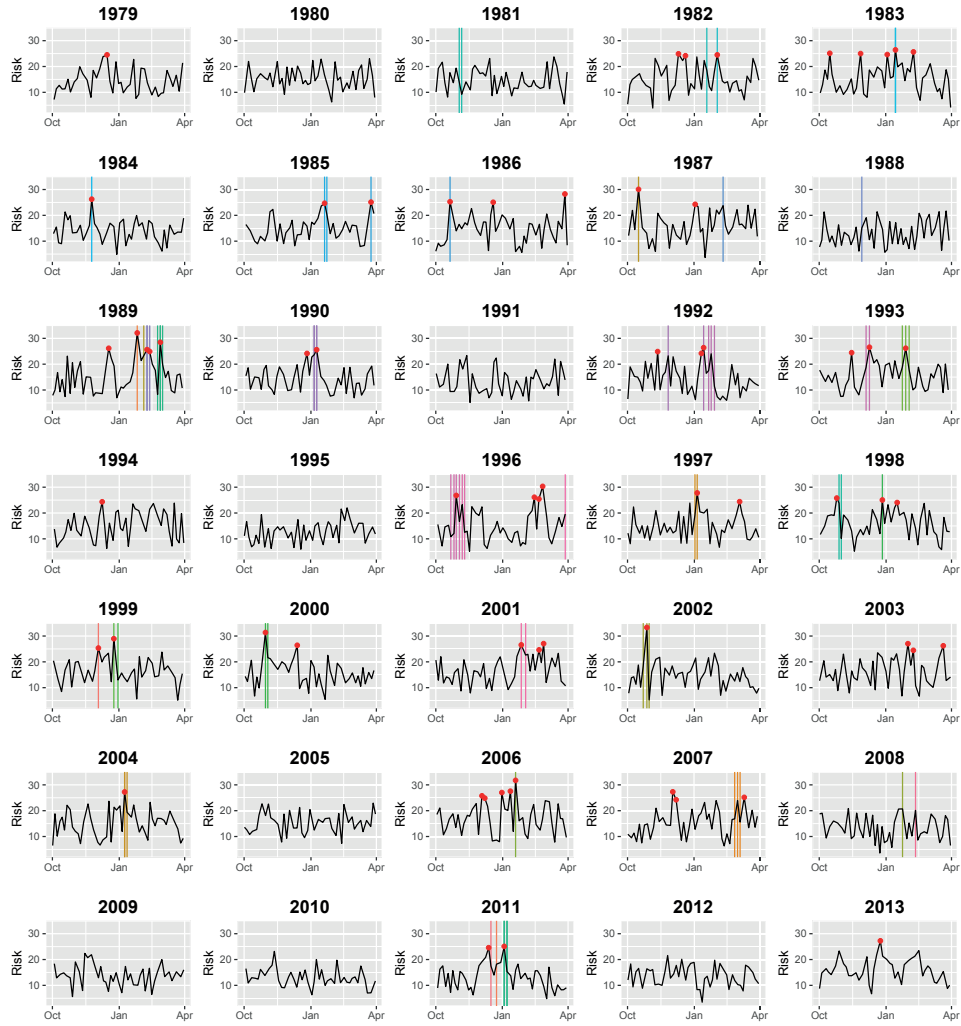


Figure 5.3 – Risk functional $r(X)(t) = \max_{t'=-12,9,\dots,12} |S_{\text{PLBA}}|^{-1} \int_{S_{\text{PLBA}}} X(s, t + t') ds$ ($\text{m}\cdot\text{s}^{-1}$) computed on the ERA–Interim data set for each winter. r -exceedances above the 0.96 empirical quantile are represented by red dots and wind storms from the XWS catalogue are represented by vertical lines coloured by dates.

with $a(n) \in [0, \infty)$, $b(n) \in \mathbb{R}$. To avoid any identifiability issues, we further suppose that $r(A) = 1$ and $r(B) = 0$, which implies that $r(a_n) = a(n)$ and $r(b_n) = b(n)$. In general, a parametric model for the functions A and B is necessary, see Section 3.5 for an example. But, for simplicity, we choose to have two parameters $A(s_l) = a_l > 0$ and $B(s_l) = b_l \in \mathbb{R}$ for every location s_l ($l = 1, \dots, 605$).

With the model for the probability of storm occurrence in Section 5.4, the parameter $b(n) = r(b_n)$ is fixed to the 0.96 empirical quantile of the series $r(X)$. The threshold-stability property of the generalized r -Pareto process makes the function B non-

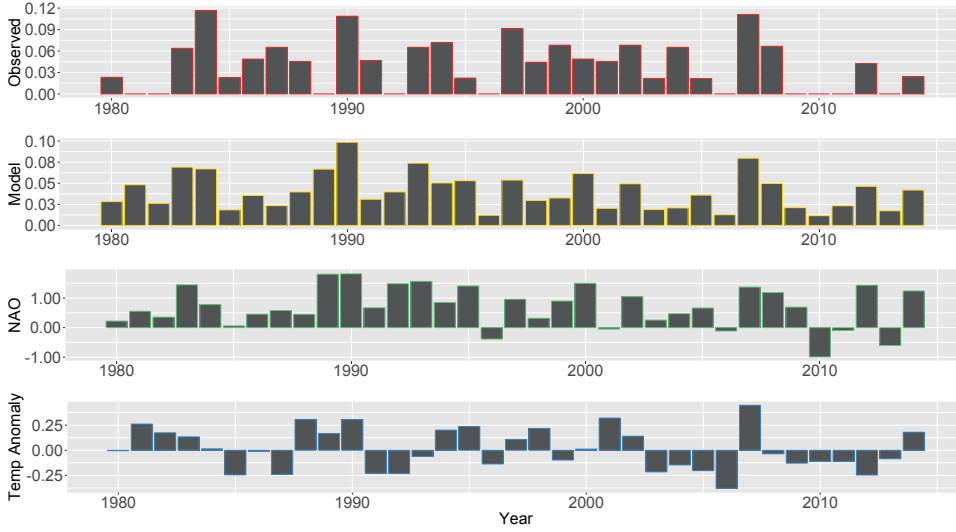


Figure 5.4 – Yearly summary of the model for the probability of storm occurrence: Observed frequency (Top), modelled frequency (Second row), North Atlantic Oscillation index (Third row) and aggregated temperature anomaly indexes (Bottom).

identifiable without further hypotheses on the distribution of $1\{r(X) \geq u_n\}$. With a logistic regression, the greater flexibility enables us to account for non-stationarity, but the model defined in Section 5.4 does allow us to identify B . Thus, we set

$$b_l = u_{q'}\{X(s_l)\} - b(n), \quad l = 1, \dots, 605,$$

where $u_{q'}\{X(s_l)\}$ is the q'^{th} local empirical quantile at location s_l of the set of r -exceedances above threshold u_n and q' is chosen such that $r(B) - b(n) = 0$. For our data set, which includes 63 storms measured every 3 hours over a 24 hour frame, we find $q' = 0.675$, yielding $9 \times 63 \times 0.675 = 382$ excesses; the estimated function B is shown in Figure 5.5.

For an accurate quantification of the model's uncertainties, the marginal and dependence models should be estimated jointly. With the theory described in Chapter 4, full uncertainty quantification is possible, but for simplicity we prefer a step-wise procedure where the marginal model is estimated first. This simplified approach should be seen as a preliminary analysis that can later be used as a starting point for more complex procedures. Thus the tail index $\xi \in \mathbb{R}$ and the scale parameters $a_l > 0$ ($l = 1, \dots, 605$) are estimated by maximizing the independence log-likelihood

$$\ell_{\text{indep}}(\xi, a_1, \dots, a_L) = \sum_{i=1}^{567} \sum_{l=1}^L 1\{X_i(s_l) \geq b_l\} \log a_l^{-1} \left\{ 1 + \xi \frac{X_i(s_l) - b_l}{a_l} \right\}^{1/\xi - 1}.$$

For the maximization, we proceed with a grid search: for a given tail index x_i , the likelihood function of the exceedances at location s_l above the threshold b_l is optimized independently for all s_l ($l = 1, \dots, L$). We find the maximum for $\hat{\xi} = -0.22$; the corresponding scale function A is shown in Figure 5.5.

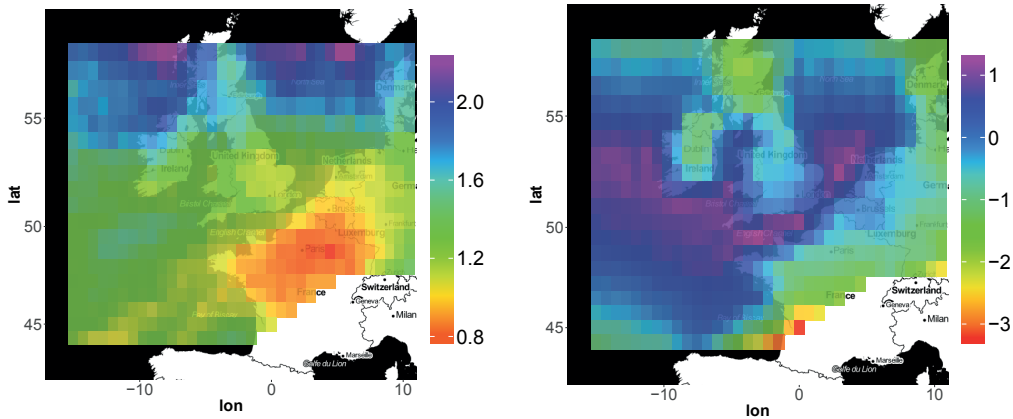


Figure 5.5 – Estimated functions A (left) and B (right) of the generalized r -Pareto process for modelling extreme windstorms over Europe. Estimates are obtained by shifting the local empirical quantiles $u_{0.675}\{X(s_l)\}$ by $b(n) = 24m.s^{-1}$.

To check the quality of the marginal model, Figure 5.6 displays QQ-plots of the local tail distribution at eight selected locations. The overall fit is convincing, as the observations mostly remain within the 99% confidence intervals; see Section 4.6 for details about the methodology used to obtain these diagnostics. Finally, the quality of the model fit for the distribution of $r(X)$ above the threshold u_n is shown in Figure 5.7. The model is reasonable but shows a systematic bias for low values with a few observations lying outside the point-wise 99% confidence intervals. However, these intervals do not account for the uncertainty of the parameter estimates and are thus likely to be too narrow.

5.6 Dependence model

At this point, only the dependence of the generalized r -Pareto process in equation (5.1) remains to be specified. For the angular component W , we choose a process with

Chapter 5. Statistical modelling of extreme windstorms over Europe

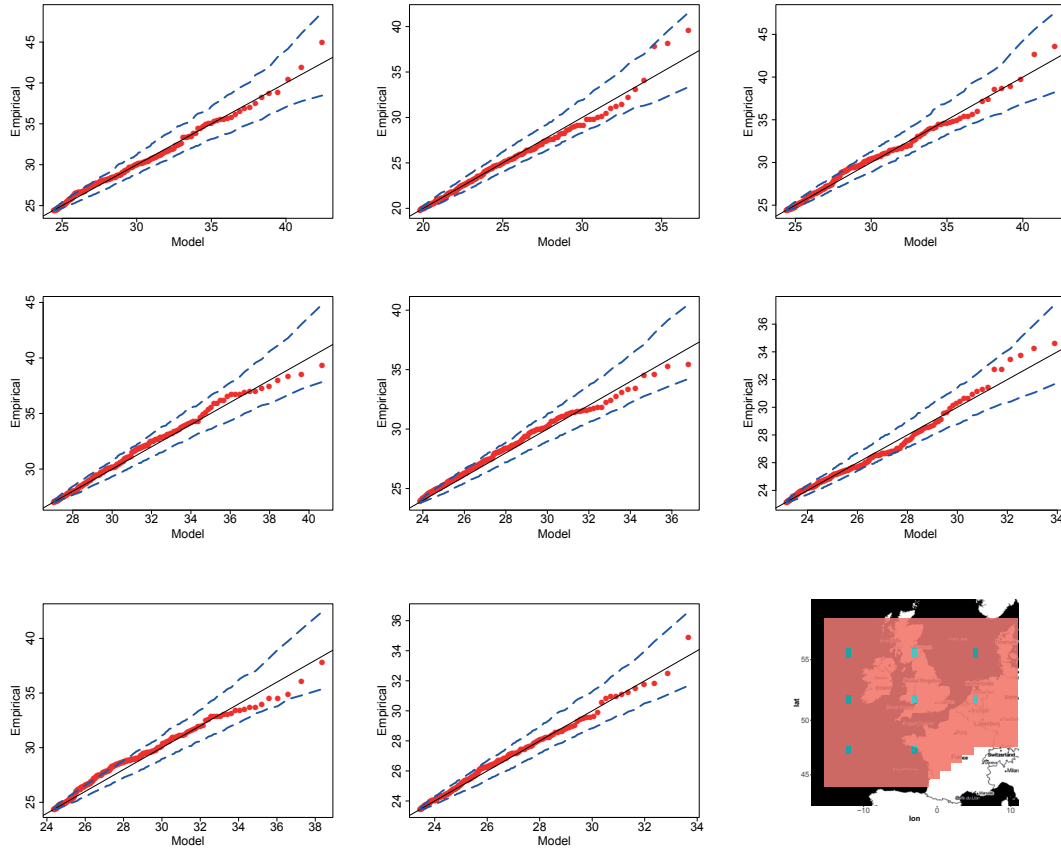


Figure 5.6 – QQ-plots of the local tail distributions for the six locations represented by the green cells on the bottom-left corner map. The positions in the Table match the relative position in space. The blue dashed lines corresponds to point-wise 99% confidence intervals. The thresholds correspond to the local 0.675 quantiles of the r -exceedances, yielding 382 excesses for each cell.

log-Gaussian random functions, see Section 1.6.2 for more details. In this case, the extremal dependence is characterized by the semi-variogram function γ , for which we choose a Bernstein model (Schlather and Moreva, 2017)

$$\gamma(s, s', t, t') = \frac{(1 + \|h\|^\alpha)^{\beta/\alpha} - 1}{2^{\beta/\alpha} - 1}, \quad 0 < \alpha \leq 2, \quad \beta \leq 2, \quad (5.2)$$

where (Gelfand et al., 2010, p. 428, p. 432)

$$\|h\| = \left\{ \left\| \frac{\Omega\{s' - s\} - V(t' - t)}{\tau_s} \right\|_2^2 + \left| \frac{t' - t}{\tau_t} \right|^2 \right\}^{1/2}, \quad s, s' \in S, \quad t, t' \in \{0, 3, \dots, 24\}, \quad (5.3)$$

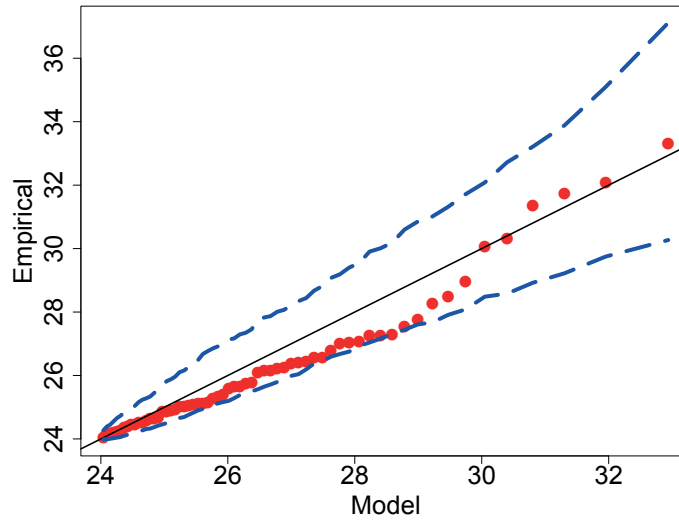


Figure 5.7 – QQ-plot for the distribution of exceedances of $r(X)$ above the threshold $u_n = 24$ modelled by a generalized Pareto distribution with scale $a(n)$ and tail index $\xi = -0.22$. The blue dashed lines corresponds to the point-wise 99% confidence intervals.

with $\tau_s, \tau_t > 0$, a wind vector $V \in \mathbb{R}^2$, and an anisotropy matrix

$$\Omega = \begin{bmatrix} \cos \eta & -\sin \eta \\ a \sin \eta & a \cos \eta \end{bmatrix}, \quad \eta \in \left(-\frac{\pi}{4}; \frac{\pi}{4}\right], \quad a > 0.$$

As explained in Section 1.6.1, the Bernstein semi-variogram is attractive for its flexibility: the parameter α determines the smoothness of the process while β drives the long-range dependence. Also, in (5.3), the spatial dependence is allowed to decrease faster in a particular direction by the introduction of the anisotropy matrix Ω with scaling factor a and angle η . To model the displacement of the storm, a wind vector V is included while the temporal component with scale τ_t accounts for the weakening of extremal dependence with time.

The semi-variogram function (5.2) is motivated by an exploratory analysis in which the extremogram

$$\pi(h_s, h_t) = \Pr\{X(s', t') \geq u' \mid X(s, t) \geq u\}, \quad h_s = s' - s \in S, \quad h_t = t' - t \in \{0, 3, \dots, 24\},$$

with thresholds equal to local 0.675 empirical quantiles of the set of r -exceedances, is estimated using the empirical estimator (1.36) of Chapter 1; see Figure 5.8.

Chapter 5. Statistical modelling of extreme windstorms over Europe

Table 5.1 – Estimated parameters for the semi-variogram function γ . Estimates are obtained by minimizing the squared difference of the empirical extremogram with its theoretical value from the model.

α	β	$\tau_s(\text{km})$	$\tau_t(\text{h})$	\hat{a}	$\hat{\eta}(\circ)$	$V_1(\text{km.h}^{-1})$	$V_2(\text{km.h}^{-1})$
1.7	-0.1	458.3	17.0	1.39	-4.6	51.7	14.7

To estimate the parameters θ of the semi-variogram function, we minimize the least squares criterion

$$\ell_{\text{extr}}(\theta) = \sum_{t, t' \in \{0, 3, \dots, 24\}} \sum_{l, l' = 1, \dots, L} \{\hat{\pi}(s_{l'} - s_l, t - t) - \pi_{\theta}(s_{l'} - s_l, t - t)\}^2.$$

Estimates are shown in Table 5.1 and the corresponding model for extremal dependence is illustrated in Figure 5.8. Due to the negativity of β , γ is bounded, but the closeness of $\hat{\beta}$ to zero yields a model near asymptotic independence for increasing distance between sites. The fitted model is quite smooth because \hat{a} is relatively high. The anisotropy matrix has a major axis with angle $\hat{\eta} = -4.6^\circ$, i.e., a South–North direction, and the scale $\hat{a} = 1.39$ implies that dependence decreases about 50% faster in this direction. Storms are born over the Atlantic and usually move towards the North Sea, which is consistent with the estimated wind vector \hat{V} , whose direction is East/North-East. In Figure 5.8, the overall fit looks reasonable: the modelled spatial dependence is similar to empirical estimates but the model does not reproduce the diamond pattern observed at long distance. The estimated semi-variogram function successfully captures the displacement of the storms in time, but looks too regular compared to local variations of the empirical probabilities.

5.7 Simulations

The relevance of our stochastic storm generator is checked by simulating from the fitted model. To do so, Algorithm 1 of Chapter 4 is modified to ensure that the maxima of the spatial mean is reached at $t = 12$ hours. First the angular component of the spatial process at time $t = 12$ is simulated. Then we iteratively simulate the remaining time steps by consecutively generating the spatial process at times $t = 9, 6, 3, 0, 15, 18, 24$ conditionally on the variables already simulated. If the new time step yields a spatial mean greater than its value at time $t = 12$, the sample is rejected and we repeat the procedure until a suitable candidate is found.

For an angular process with log-Gaussian random functions, such a simulation al-

gorithm is easy to implement because it is equivalent to conditional simulation of multivariate Gaussian random vectors. A simulated storm from our generator with intensity $r(X) = 29.1\text{m}\cdot\text{s}^{-1}$ is shown in Figure 5.9. Visually the simulation is convincing as it produces an overall spatio-temporal pattern similar to that of storm Daria in Figure 5.1, but the simulation is rougher than the original process.

5.8 Discussion

In this chapter we presented a stochastic weather generator for extreme wind storms over Europe. The model is based on the generalized r -Pareto process introduced in Chapter 4 and is capable of generating storms with unobserved patterns and intensities. Goodness-of-fit diagnostics in Figure 5.6 for the local tail margins and in Figure 5.8 for the dependence structure are fairly convincing.

The simulated storms look consistent with historical records, though the model might still be too simplistic from a climatological point of view: The current dependence function might not be flexible enough with regard to the complexity of the spatio-temporal structure of extreme wind storms. Indeed, further exploratory analysis reveals that the dependence changes over space. Thanks to Oesting et al. (2017b), we know that the potential types of non-stationarity models are limited, but models with varying local anisotropy, such as in Fuglstad et al. (2013) or Fouedjio et al. (2016), should be considered.

Fixing the wind vector V in the semi-variogram function is an oversimplification of the movement of a wind storm. Supposing that V is drawn from an underlying random distribution would be a first step toward more realistic storm tracks. In this case, V would be a latent variable whose model must be specified and for which inference might be delicate. Alternatively, the methodology of Lindgren et al. (2011) based on approximations of stochastic partial differential equations could be generalized to build physically inspired spatio-temporal dependence structures, for instance using the diffusion equation. This approach has the advantage of being computationally efficient and of bringing physics into the model.

In this chapter, we use a two-step fitting procedure: the marginal and the dependence models are estimated independently. From a statistical standpoint, for a better quantification of uncertainty, and obtain the standard errors of the parameters estimates, joint estimation of the model should be performed using, for instance, the methodology developed in Chapter 2.

Finally, designing a risk functional tailored to the risk of interest is a topic that remains

Chapter 5. Statistical modelling of extreme windstorms over Europe

generally unexplored, and potential risk functionals taking more meteorological and climatological expertise into account should be investigated.

The model developed in this chapter is, so far as we are aware, the first of its kind and opens up the field of stochastic generators for extreme events. We focused on extreme wind storms, but similar models could be developed for other types of extremes, such as heat waves, rainfall, and more generally, financial markets or network loads.

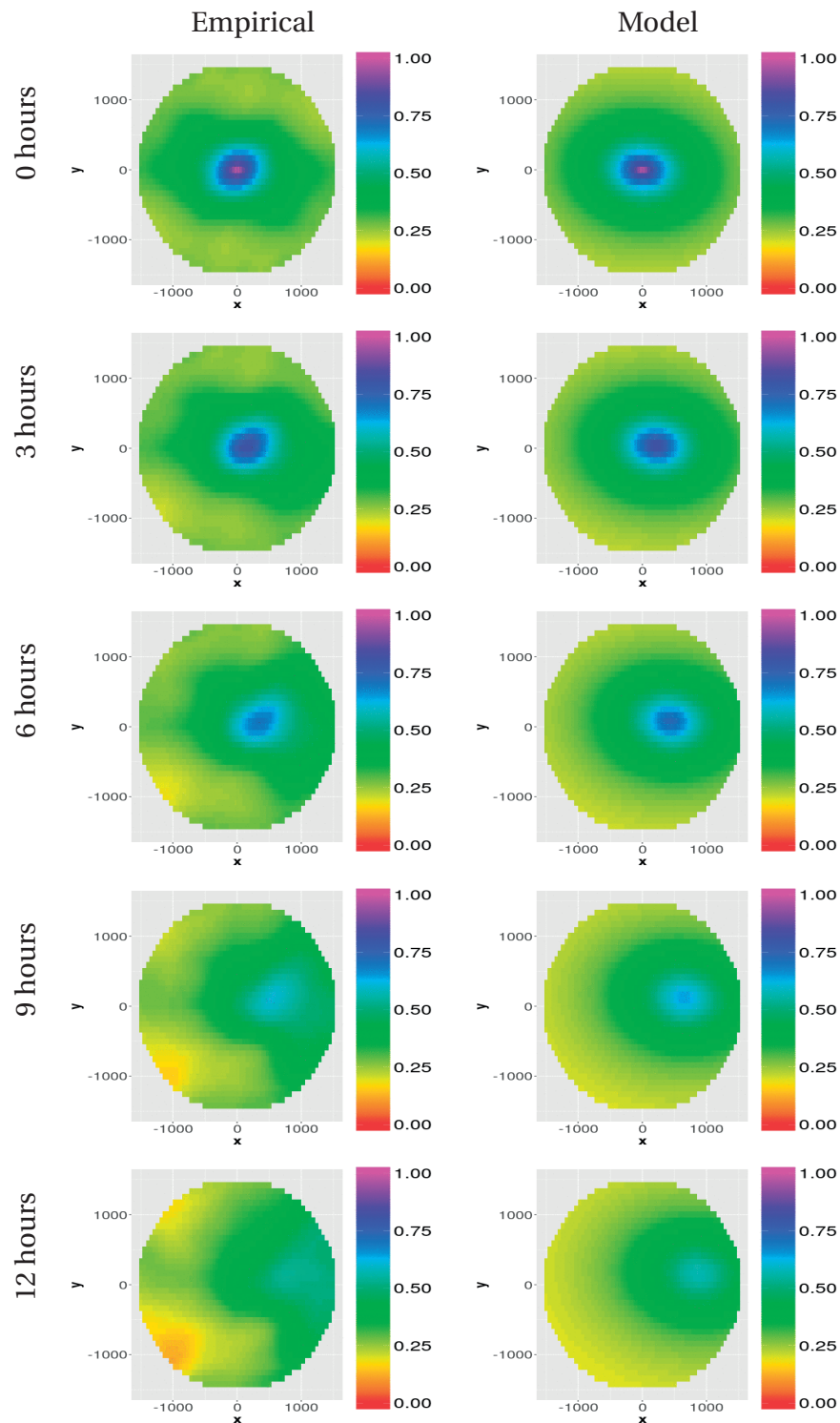


Figure 5.8 – Extremogram $\pi(s, t) = \Pr\{X(s, t) > u \mid X(0, 0) > u\}$ estimated empirically (left) and modelled (right). Each row represents a 3-hour time step. The model parameters are obtained by minimum least squares.

Chapter 5. Statistical modelling of extreme windstorms over Europe

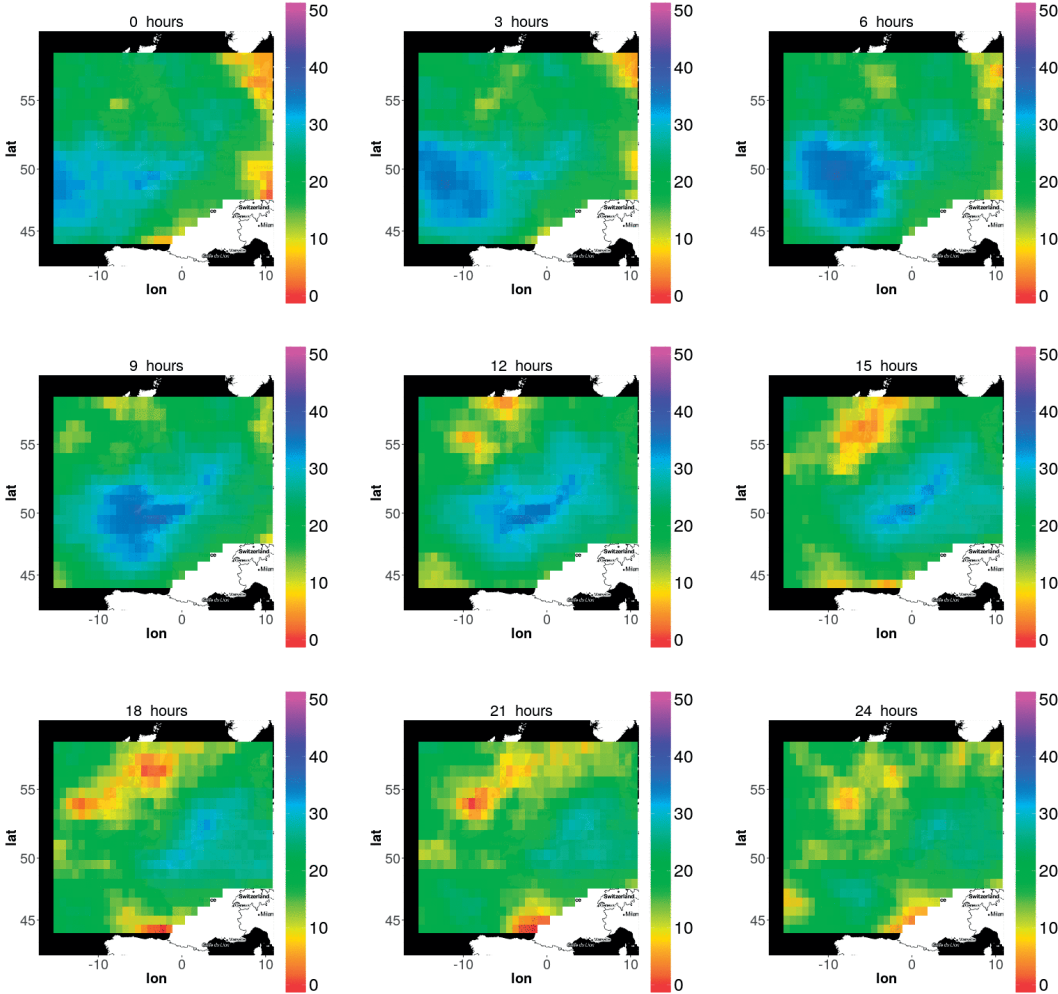


Figure 5.9 – Simulated maximum speed (m.s^{-1}) over the past 3h hours of wind gusts sustained for at least 3s. The storm has an intensity $r(X) = 29.1 \text{ m.s}^{-1}$.

6 Conclusion and future work

The research presented in this thesis describes methodological contributions to the development of stochastic generators of extreme events. In this approach, extremes are defined through an exceedance over a threshold of a well-chosen univariate summary. Simulations from these models can then be used as stress test scenarios to quantify the risk characterized by the chosen definition of exceedance.

In Chapter 2, we develop statistical inference procedures tractable in medium to high dimensions. An efficient implementation of censored likelihood allows us to work with a few hundred dimensions, while the gradient scoring rule is tractable for several thousand. With the latter, similarly to Gaussian models, the size of the problem that can be handled is limited only by the time required to invert a dense matrix. Data sets with hundreds of thousands of dimensions could possibly be handled by using sparse matrices or directly modelling the inverse covariance matrix. In a simulation study we show that the the gradient score is more robust to model misspecification than the classical log-likelihood, but is on average outperformed by the censored procedure. Empirical work shows that the robustness of the score and its efficiency are determined by the choice of the weighting function. Thus a thorough comparison through an extensive simulation study would be necessary to evaluate the impact of the different weighting functions on the performance of the method.

Chapter 3 describes the limit tail distribution of aggregated data and introduces the r -extremal coefficient, a measure of dependence that quantifies the impact of aggregation on the marginal tail distribution of r -exceedances. These theoretical contributions are used to build a statistical model for downscaling extreme temperatures in the South of France. The model relies on a specific tail decay regime and choice of risk functional for which the likelihood function can be written in closed form. For other tail decay regimes and risk functionals, model fitting is possible with the independence likelihood, but generalization of the full likelihood procedure would enlarge

Chapter 6. Conclusion and future work

the horizon of applications and the impact of this contribution. Downscaling is key to understanding the relationship between climate models and station measurements in order to produce more accurate projections of the future state of the climate.

The research focus of this thesis was directed toward building a generator of extreme wind storms to create catalogues of events usable by insurers and regulators to quantify their risk. To this end, Chapter 4 derives the theoretical convergence results necessary for such generators: we introduce the generalized r -Pareto process, the functional equivalent of the generalized Pareto distribution. This new process describes the asymptotic behaviour of r -exceedances of a properly rescaled regularly-varying stochastic process. We discuss statistical inference and fitting diagnostics. Chapter 5 illustrates how this methodology applies to the analysis the severe wind storms over Europe. Currently the model is fairly convincing, but quantification of uncertainties using the inference procedures described in Chapter 2 should be performed.

The size of the data sets used for this research makes the limitations of existing dependence models clear. For this reason, future research should focus on the development of new, more flexible, dependence models with spatially-varying parameters. Only a few such models exist, but some derived from stochastic partial differential equations (Lindgren et al., 2011) are attractive for their flexibility and their low computational complexity. This approach would allow us to incorporate physics into the model through the choice of a partial differential equation: generalization to a spatio-temporal setting with the diffusion equation would strongly increase the appeal of our approach to a non-statistical audience. Also to analyse the risk of extreme cyclones in the U.S. or in Asia, which have a more specific structure than European storms, models based on partial differential equations are handy as they can reproduce their spinning pattern. Finally, the model described in Chapter 5 relies on a stationarity assumption which, in a changing climate, is likely to be too restrictive. The methodology presented is flexible enough to incorporate the influence of global warming using covariates such as time, temperature or anthropogenic forcing; a possibility that should be investigated in future work.

This research opens up the field of stochastic generators for complex extreme events, where field-specific expertise can be exploited. These results apply not only to the environment but also motivate applications in finance, insurance, computer science, civil engineering, and other fields, which will give strong insights in the mechanisms driving extreme events, and will provide more accurate risk estimation and solutions for better mitigation.

A Proofs for Chapter 1

A.1 Theorem 1.6

Let $r' > 0$ and $\mathcal{W} \in \mathcal{S}_r$. Recall that for any set $\mathcal{A} \in \mathcal{F}$ and cone \mathcal{C} , $d(\mathcal{A}, \mathcal{C}) = \inf_{x \in \mathcal{A}, y \in \mathcal{C}} \|x - y\|_\infty$. Then

$$\begin{aligned} d\left[(h_r^{\text{pp}})^{-1}\{[r', \infty) \times \mathcal{W}\}, \{0\}\right] &= d\left[\{\rho w/r(w) \in \mathcal{F}_+ \setminus \{0\} : \rho \geq r', w \in \mathcal{W}\}, \{0\}\right], \\ &= r' \times d\left[\{\rho w/r(w) \in \mathcal{F}_+ \setminus \{0\} : \rho \geq 1, w \in \mathcal{W}\}, \{0\}\right], \\ &\geq r' \times d\left[\{x \in \mathcal{F}_+ \setminus \{0\} : r(x) \geq 1\}, \{0\}\right]. \end{aligned}$$

Also the continuity of the risk functional r ensures that $d_\infty[\{X \in \mathcal{F}_+ \setminus \{0\} \mid r(X) \geq 1\}, \{0\}] > 0$, we obtain that

$$d\left[(h_r^{\text{pp}})^{-1}\{[r', \infty) \times \mathcal{W}\}, \{0\}\right] > 0,$$

and the hypotheses for the mapping theorem are satisfied.

For the polar decomposition of the limit measure, we use the homogeneity of Λ . If $r' > 0$ and $\mathcal{W} \in \mathcal{S}_r$, then

$$\begin{aligned} \Lambda \circ (h_r^{\text{pp}})^{-1}(\mathcal{U}(r'), \mathcal{W}) &= \Lambda\{\rho w/r(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_r) : \rho \geq r', w \in \mathcal{W}\}, \\ &= \Lambda[r' \times \{\rho w/r(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_r) : \rho \geq 1, w \in \mathcal{W}\}], \\ &= (r')^{-1/\xi} \Lambda\{\rho w/r(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_r) : \rho \geq 1, w \in \mathcal{W}\}, \\ &= (r')^{-1/\xi} \Lambda(A_r) \times \frac{\Lambda\{\rho w/r(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_r) : \rho \geq 1, w \in \mathcal{W}\}}{\Lambda(A_r)}, \\ &= (r')^{-1/\xi} \Lambda(A_r) \times \sigma_r(\mathcal{W}). \end{aligned}$$

For the converse, let $\mathcal{A} \subset \mathcal{F}_+ \setminus \{0\}$. If $\bigcap_{l=1}^L \mathcal{C}_{r_l} = \{0\}$ then there exist $\mathcal{A}_1, \dots, \mathcal{A}_L$ such that

Appendix A. Proofs for Chapter 1

$\cup_{l=1}^L h_{r_l}^{-1}(\mathcal{A}_l) = \mathcal{A}$ and for all l , $\mathcal{A}_l \subset \mathcal{F}_+ \setminus \mathcal{C}_{r_l}$. Thus

$$\lim_{n \rightarrow \infty} \Lambda_n(\mathcal{A}) = \sum_{l=1}^L \lim_{n \rightarrow \infty} \Lambda_n \circ h_{r_l}^{-1}(\mathcal{A}_l) = \Lambda \left\{ \bigcup_{l=1}^L h_{r_l}^{-1}(\mathcal{A}_l) \right\} = \Lambda(\mathcal{A}).$$

□

A.2 Corollary 1.2

Let $dw \subset \mathcal{S}_r$. Using the definition of σ_r from Theorem 1.6,

$$\begin{aligned} & \frac{\sigma_{r_1}(dw)}{\sigma_{r_2}(dw)} \\ &= \frac{\Lambda \{ \rho w / r_1(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_{r_1}) \mid \rho > 1, w \in dw \}}{\Lambda(A_{r_1})} \times \\ & \quad \frac{\Lambda(A_{r_2})}{\Lambda \{ \rho w / r_2(w) \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_{r_2}) \mid \rho > 1, w \in dw \}} \\ &= \frac{\Lambda(A_{r_2})}{\Lambda(A_{r_1})} \frac{r_1(w)^{1/\xi} \Lambda \{ \rho w \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_{r_1}) \mid \rho > 1, w \in dw \}}{r_2(w)^{1/\xi} \Lambda \{ \rho w \in \mathcal{F}_+ \setminus (\{0\} \cup \mathcal{C}_{r_2}) \mid \rho > 1, w \in dw \}} \\ &= \frac{\Lambda(A_{r_2})}{\Lambda(A_{r_1})} \left\{ \frac{r_1(dw)}{r_2(dw)} \right\}^{1/\xi}, \end{aligned}$$

where we used the homogeneity of the measure to get from the first to the second line.

□

A.3 Corollary 1.3

We use Corollary 2.1 in Lindskog et al. (2014). It is straightforward to see that $h_{\text{proj}}(\{0\})$ is closed in \mathbb{R}_+^L , but we need to verify that h_{proj} is uniformly continuous. Let $\epsilon > 0$ and take $x, y \in \mathcal{F}_+ \setminus \mathcal{C}$ such that $\|x - y\|_\infty < \delta = \epsilon$. Then

$$\|h_{\text{proj}}(x) - h_{\text{proj}}(y)\|_\infty \leq \|x - y\|_\infty < \delta = \epsilon.$$

Conditions to apply Corollary 2.1 in (Lindskog et al., 2014) are then satisfied and we get the desired convergence result. □

A.4 Theorem 1.8

Let $r' > 0$ and $\mathcal{W} \in \mathcal{S}_r$. We use Theorem 1.6:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[r'; \infty), \mathcal{W}\} : r \left(\frac{X}{a_n} \right) \geq 1 \right] &= \lim_{u \rightarrow \infty} \frac{\Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[r'; \infty) \times \mathcal{W}\} \right]}{\Pr \left[r \left(\frac{X}{a_n} \right) \geq 1 \right]} \\
 &= \lim_{u \rightarrow \infty} \frac{n^{-1} \Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[r'; \infty) \times \mathcal{W}\} \right]}{n^{-1} \Pr \left[\frac{X}{a_n} \in (h_r^{\text{pp}})^{-1} \{[1; \infty) \times \mathcal{S}_r\} \right]} \\
 &= \frac{\Lambda_\beta(r') \times \sigma_r(\mathcal{W})}{\Lambda_\beta(1) \times 1} \\
 &= (r')^{-1/\xi} \sigma_r(\mathcal{W}).
 \end{aligned}$$

□

A.5 Equation (1.30)

Let $r' > 0$. We use the homogeneity of the Cartesian representation of the Pareto process measure:

$$\begin{aligned}
 \Pr \{P(s_0) > r'\} &= \frac{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1, x(s_0) \geq r'\}}{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1\}} \\
 &= \frac{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1, x(s_0) \geq \frac{r'}{u_0} u_0\}}{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1\}} \\
 &= \left(\frac{r'}{u_0} \right)^{-1/\xi} \frac{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1, x(s_0) \geq u_0\}}{\Lambda \{x \in \mathcal{F}_+ \setminus \mathcal{C}_r : r(x) \geq 1\}},
 \end{aligned}$$

where the hypothesis $\{x \in \mathcal{F}_+ : x(s_0) > u_0\} \subset \{x \in \mathcal{F}_+ : r(x) \geq 1\}$ ensures that

$$\{x \in \mathcal{F}_+ : x(s_0) > r' u_0, r(x) \geq 1\} = r' \times \{x \in \mathcal{F}_+ : x(s_0) > u_0, r(x) \geq 1\}.$$

□

B Supplementary material for Chapter 2

B.1 High-dimensional censored likelihood

B.1.1 Computational considerations

The algorithm due to Genz and Bretz (2009) and implemented in the R package `mvtnorm` (Genz et al., 2014) provides an unbiased estimate of a multivariate normal probabilities, with an indication of its largest probable error. An improved Matlab implementation (Genz, 2013) makes better use of quasi-Monte Carlo methods. We translated this code into C++ to speed it up; see Appendix B.1.2.

Function evaluation is independent for each sample, so we also adapted the algorithm for GPU computing and compared different implementations. Our C++ implementation is about four times faster than the `mvtnorm` implementation for a probable worst-case error of order 10^{-3} . GPU computing provides a slight improvement in speed compared to C++ for reasonably low error, but shows a significant speed-up for higher accuracies ($\lesssim 10^{-4}$). A computation time of 1 s for estimation of one integral seems reasonable for censored likelihood, and is achievable for $L \approx 500$ for probable worst-case errors of order 10^{-3} without GPU computing. GPU computing improves the computation time of a single estimation of multivariate normal distribution functions, but when estimating hundreds, a higher level of parallelization using the CPU is usually more efficient. For this reason, we decided to keep the CPU implementation for the R package but GPU code is available upon request.

Although Jensen's inequality implies that estimation of the log-likelihood function is biased for finite p , quasi-Monte Carlo estimation of an integral is unbiased, so for a

Appendix B. Supplementary material for Chapter 2

sufficiently high p ,

$$\log \Phi^p = \log(\Phi + \epsilon^p) = \log(\Phi) + \frac{\epsilon^p}{\Phi} + o_p\left(\frac{\epsilon^p}{\Phi}\right), \quad (\text{B.1})$$

where ϵ^p is a random error with zero mean and bounded variance. Using equation (B.1) with a small ϵ^p , we have $\hat{\theta}_{\text{cens}} \approx \text{E}(\hat{\theta}_{\text{cens}}^p)$. On a multi-node cluster, for scalability purposes, it is more efficient to combine independent estimates $\hat{\theta}_{\text{cens},q}^p$ ($q = 1, \dots, \bar{p}$) into $\tilde{\theta}^{\bar{p}} = \bar{p}^{-1} \sum_{i=1}^{\bar{p}} \hat{\theta}_{\text{cens},q}^p$ than to compute a single estimate $\hat{\theta}_{\text{cens}}^{p \times \bar{p}}$ with $p \times \bar{p}$ samples in the quasi-Monte Carlo procedures. Indeed, maximization of $\ell_{\text{cens}}^p(\theta)$ requires a reduction step, in which the computations performed on each node are assembled, for every evaluation of the objective function. Hence for a cluster with several nodes, where communication is usually slow and reduction steps expensive, $\tilde{\theta}^{\bar{p}}$ is more efficient because the computation of several $\hat{\theta}_{\text{cens},q}^p$ can be done independently on different nodes. Moreover, use of $\tilde{\theta}^{\bar{p}}$ allows $\text{var}(\hat{\theta}_{\text{cens}}^p)$ to be estimated.

We parallelized the ideas above on a cluster with 12 nodes each of 16 cores. First computation of $\ell_{\text{cens}}^p(\theta)$ was parallelized within each node using the R package `parallel`. The time needed to compute the censored likelihood for a 300-dimensional vector for a generalized Pareto process associated to a log-Gaussian random function with $p = 499$ and different dependence strengths dropped from minutes to a dozen seconds. Each node performs an independent maximization using the R routine `optim`. Even if slightly biased, this approach is computationally efficient for our cluster infrastructure. If the empirical variance of $\hat{\theta}_{\text{cens}}^p$ is too high then the number of samples p should be increased. For high accuracy and/or complex models, GPU computing may still be useful. Lastly, the tolerance of the optimization algorithm must be reduced for low p to ensure its convergence if the quasi-Monte Carlo estimates vary substantially.

B.1.2 Algorithm for multivariate normal distribution function estimation

This algorithm is a simplified version of that of Genz and Bretz (2009). To estimate the L -dimensional zero-mean multivariate normal distribution function $\Phi_L(x, \Sigma)$:

1. input covariance matrix Σ , upper bound x , number of deterministic samples p , number of random shifts p' and generating vector v' ;
2. compute lower triangular Cholesky factor L for Σ , permuting x , and rows and columns of Σ for variable prioritisation;
3. initialize $\Phi = 0$, $\delta = 0$ and $V = 0$;

4. for q' in $1, \dots, p'$:
 - (a) set $I_{q'} = 0$ and generate uniform random shift $\Delta \in [0, 1]^L$;
 - (b) for q in $1, \dots, p$:
 - i. set $z_q = (z_{q,1}, \dots, z_{q,L-1}) = \lfloor 2 \times \overline{(qv' + \Delta)} - 1 \rfloor$
 $e_1 = \Phi(b_1/l_{1,1})$
 $f_1 = e_1$;
 - ii. for i in $2, \dots, L$
 set $y_{i-1} = \Phi^{-1}(z_{q,i-1}e_{i-1})$
 $e_i = \Phi\left(\frac{b_i - \sum_{j=1}^{i-1} l_{i,j}y_j}{l_{i,i}}\right)$
 $f_i = e_i f_{i-1}$
 End i loop;
 - iii. set $I_{q'} = I_{q'} + (f_i - I_{q'})/q$;
 End q loop;
 - (c) Set $\delta = (I_{q'} - \Phi)/i$, $\Phi = \Phi + \delta$, $V = (q' - 2)V/i + \delta^2$ and $\text{ERR} = \alpha\sqrt{V}$;
 end q' loop;
5. output $\Phi \approx \Phi_L(-\infty, x; \Sigma)$ with error estimate ERR.

B.2 Properness of the gradient scoring rule

To derive the gradient score, we expand the divergence measure

$$\int_{\mathcal{A}_r(u)} \|\nabla_x \log \lambda_\theta(x) \otimes w(x) - \nabla_x \log \lambda_{\theta_0}(x) \otimes w(x)\|_2^2 \lambda_{\theta_0}(x) dx$$

in equation (26) of the paper to separate off the terms that do not depend on the parameters θ , giving

$$\begin{aligned} & \sum_{l=1}^L \int_{\mathcal{A}_r(u)} w_l(x)^2 \left\{ \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \right\}^2 \lambda_{\theta_0}(x) dx - \\ & 2 \sum_{l=1}^L \int_{\mathcal{A}_r(u)} w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \frac{\partial \log \lambda_{\theta_0}(x)}{\partial x_l} \lambda_{\theta_0}(x) dx + \\ & \sum_{l=1}^L \int_{\mathcal{A}_r(u)} w_l(x)^2 \left\{ \frac{\partial \log \lambda_{\theta_0}(x)}{\partial x_l} \right\}^2 \lambda_{\theta_0}(x) dx. \end{aligned}$$

We write this as $\Delta_1(\theta) + \Delta_2(\theta) + \Delta_3$, where Δ_3 does not depend on θ and can be omitted when minimizing the divergence.

Appendix B. Supplementary material for Chapter 2

An empirical estimate of integral $\Delta_1(\theta)$ is straightforward to obtain, but $\Delta_2(\theta)$ requires simplification using integration by parts:

$$\begin{aligned}\Delta_2(\theta) &= 2 \sum_{l=1}^L \int_{\mathcal{A}_r(u)} w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \frac{\partial \lambda_{\theta_0}(x)}{\partial x_l} dx \\ &= 2 \sum_{l=1}^L \int_{\partial_d \mathcal{A}_r(u)} w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \lambda_{\theta_0}(x) dx - \\ &\quad 2 \sum_{l=1}^L \int_{\mathcal{A}_r(u)} \frac{\partial}{\partial x_l} \left\{ w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \right\} \lambda_{\theta_0}(x) dx\end{aligned}$$

where

$$\partial_l \mathcal{A}_r(u)_l = \{x \in \mathcal{A}_r(u) : x_l \in \partial \mathcal{A}_r(u)\}$$

and $\partial \mathcal{A}_r(u)$ denotes the boundary of $\mathcal{A}_r(u)$.

Now suppose that the function w is chosen such that for all $x \in \mathcal{A}_r(u)$

$$\lim_{x_l \rightarrow \infty} w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \lambda_{\theta_0}(x) - \lim_{x_l \rightarrow a_l(x_{-l})} w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \lambda_{\theta_0}(x) = 0, \quad (\text{B.2})$$

where $a_l(x_{-l})$ is the lower bound of x_l on $\mathcal{A}_r(u)$ for fixed $x_{-l} = (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_L)$. Then the expression for $\Delta_2(\theta)$ simplifies to

$$\Delta_2(\theta) = -2 \sum_{l=1}^L \int_{\mathcal{A}_r(u)} \frac{\partial}{\partial x_l} \left\{ w_l(x)^2 \frac{\partial \log \lambda_\theta(x)}{\partial x_l} \right\} \lambda_{\theta_0}(x) dx.$$

Finally, suppose that X^1, \dots, X^N are independent replicates of the random variable X with density function h . Then the gradient scoring rule is obtained by replacing the integrals $\Delta_1(\theta)$ and $\Delta_2(\theta)$ by their empirical versions, giving

$$\delta_w(\theta, X) = \sum_{n=1}^N \sum_{l=1}^L w_l(X^n)^2 \left\{ \frac{\partial \log \lambda_\theta(X^n)}{\partial x_l} \right\}^2 + 2 \frac{\partial}{\partial x_l} \left\{ w_l(X^n)^2 \frac{\partial \log \lambda_\theta(X^n)}{\partial x_l} \right\}.$$

B.3 Gradient score for Brown–Resnick processes

Wadsworth and Tawn (2014) derive an alternative expression for the intensity function,

B.4. Average computation times of the fitting procedures

expression (15) of the paper:

$$\begin{aligned} \lambda_\theta(x) = & \frac{|\det \Sigma_\theta^*|^{-1/2} (1_L^\top \rho)^{-1/2}}{(2\pi)^{(L-1)/2} x_1 \cdots x_L} \exp \left(-\frac{1}{2} \left[\log x^\top \Gamma \log x + \log x^\top \left\{ \frac{2\rho}{1_L^\top \rho} + (\Sigma_\theta^*)^{-1} \sigma - \frac{\rho \rho^\top \sigma}{1_L^\top \rho} \right\} \right] \right) \\ & \times \exp \left[-\frac{1}{2} \left\{ \frac{1}{4} \sigma^\top (\Sigma_\theta^*)^{-1} \sigma - \frac{1}{4} \frac{\sigma^\top \rho \rho^\top \sigma}{1_L^\top \rho} + \frac{\sigma^\top \rho}{1_L^\top \rho} - \frac{1}{1_L^\top \rho} \right\} \right], \quad x \in \mathcal{A}_r(u), \end{aligned} \quad (\text{B.3})$$

where Σ_θ^* is the L -dimensional covariance matrix of a non-stationary Gaussian process with semi-variogram γ , $\rho = (\Sigma_\theta^*)^{-1} 1_L$, $\Gamma = (\Sigma_\theta^*)^{-1} - \rho \rho^\top / 1_L^\top \rho$ and $\sigma = \text{diag}(\Sigma_\theta^*)$. This expression is symmetric, so computing its gradient and Laplacian is relatively straightforward.

The gradient of $\log \lambda_{\theta,u}^r$ with respect to x and with the notation of equation (B.3) above is

$$\nabla_x \log \lambda_{\theta,u}^r(x) = -\Gamma \log x \otimes \frac{1}{x} - \frac{1}{2x} \otimes \left(\frac{2\rho}{1_L^\top \rho} + 2 + (\Sigma_\theta^*)^{-1} \sigma - \frac{\rho \rho^\top \sigma}{1_L^\top \rho} \right), \quad x \in \mathcal{A}_r(u), \quad u > 0, \quad (\text{B.4})$$

where \otimes is the Hadamard product, 1_L is a L -dimensional vector with unit components, Σ_θ^* is the covariance matrix of the non-stationary Gaussian process with semi-variogram γ_θ , $\rho = (\Sigma_\theta^*)^{-1} 1_L$, $\Gamma = (\Sigma_\theta^*)^{-1} - \rho \rho^\top / 1_L^\top \rho$ and $\sigma = \text{diag}(\Sigma_\theta^*)$. The Laplacian of $\Delta_x \log \lambda_{\theta,u}^r(x)$ equals

$$-\text{diag}(\Gamma)^\top \left(\frac{1 - \log x}{x^2} \right) + \left\| \left\{ \Gamma - \text{diag}(\Gamma) \right\} \log x \otimes \frac{1}{x^2} \right\|_1 + \frac{1}{(2x^2)^\top} \left\{ \frac{2\rho}{1_L^\top \rho} + 2 + (\Sigma_\theta^*)^{-1} \sigma - \frac{\rho \rho^\top \sigma}{1_L^\top \rho} \right\}, \quad (\text{B.5})$$

where $x \in \mathcal{A}_r(u)$, $u > 0$ and $\|\cdot\|_1$ denotes the L_1 norm.

B.4 Average computation times of the fitting procedures

Table B.1 contains some timings for optimization of the different objective functions.

B.5 Detailed results for simulations in Section 2.4

Tables 2.2 and B.3 contain detailed results for the simulations in §4.2 of the paper.

Appendix B. Supplementary material for Chapter 2

Table B.1 – Average times (s) of the optimization for the different objective functions, when fitting a Brown–Resnick process applied to the three different semi-variogram models γ with $\alpha = \{0.5, 1, 1.3, 1.8\}$ and the three grids 10×10 , 20×10 and 20×15 . Random starting points are used for fair comparison.

Grid size	α	Spectral likelihood	Censored log-likelihood	Gradient score
10×10	0.5	4	135	6.2
	1	4	140	4.9
	1.3	4.5	129	4.8
	1.8	5.1	235	6
20×10	0.5	6.2	486	10
	1	6	492	9.7
	1.3	6.7	483	9.8
	1.8	10	1636	15
20×15	0.5	14	1190	18
	1	14	1217	16.4
	1.3	14.6	1236	18.8
	1.8	20	4043	29

B.6 Proof of Proposition 2.1

Let $(Y^n)_{n=1, \dots, N}$ be independent replicates of a L -dimensional regularly-varying random vector Y with marginal distributions normalized as mentioned after equation (4) of the paper and limiting measure Λ_{θ_0} . Let $N_u = N_u(N)$ be a sequence of integers such that $N_u(N) \rightarrow \infty$ and $N_u(N)/N \rightarrow 0$ as $N \rightarrow \infty$ and suppose we only keep those Y^n for which $r(Y^n) > N/N_u$, and consider the subset

$$\mathcal{A}_r(1) = \{y \in \mathbb{R}_+^L : r(y) > 1\}.$$

For any $\mathcal{A} \subset \mathbb{R}_+^L$, we first establish the asymptotic normality of the empirical measure estimator

$$\tilde{\Lambda}_{N_u}(\mathcal{A}) = \frac{1}{N_u} \sum_{n=1}^N \mathbf{1}(N_u Y^n / N \in \mathcal{A})$$

Since Y lies in the max-domain of attraction of a multivariate extreme-value distribution P , Proposition 2.1 in de Haan and Resnick (1993) gives the convergence in probability

$$\tilde{\Lambda}_{N_u}(\mathcal{A}) \rightarrow \Lambda(\mathcal{A}), \quad \mathcal{A} \subset \mathbb{R}_+^L, \quad N \rightarrow \infty, \quad (\text{B.6})$$

B.6. Proof of Proposition 2.1

Table B.2 – Estimated bias for the spectral likelihood (left), censored log-likelihood (second) and the gradient score with weights w^1 (third) and w^2 (right). Inference is based on the top 1% of 10000 simulated Brown–Resnick processes with semi-variogram $\gamma(s, s') = (\|s - s'\|/\tau)^\alpha / 2$. In each case the scale parameter equals $\tau = 2.5$ and grids are regular of sizes 10×10 , 20×10 and 20×15 . All the values were multiplied by -10 for α and 10 for τ . “NC” means that optimization does not converge.

Shape α			
Grid size	10×10	20×10	20×15
$\alpha = 0.5$	0.45/0.034/0.33/0.52	0.39/0.11/0.28/0.38	0.38/0.14/0.28/0.06
$\alpha = 1$	4.5/0.63/2.4/4.2	3.5/0.29/2.3/3.0	3.3/0.47/2.5/3.0
$\alpha = 1.3$	9.6/4.7/5.6/9.1	6.9/1.8/3.8/5.0	6.3/1.6/3.7/4.5
$\alpha = 1.8$	NC	14.3/9.3/8.5/11.7	13.4/7.5/7.8/9.8

Scale τ			
Grid size	10×10	20×10	20×15
$\alpha = 0.5$	7.3/6.8/5.4/6.0	8.2/7.8/6.1/6.9	8.3/8.0/5.9/6.5
$\alpha = 1$	5.9/0.8/15.7/17.4	15.8/7.1/16.9/20.0	17.3/9.5/17.1/19.4
$\alpha = 1.3$	-15.1/ -13.6/6.4/ -1.3	7.6/ -0.031/11.6/14.3	10.6/2.5/11.7/13.6
$\alpha = 1.8$	NC	-11.9/ -10.7/3.5/4.4	-6.4/ -5.2/3.0/4.5

where Λ is the exponent measure associated to P . Moreover, following Propositions 3.1 and 3.2 in de Haan and Resnick (1993), define the random field

$$Z_N(x) = N_u^{1/2} [\tilde{\Lambda}_{N_u} \{(0, x]\} - \Lambda \{(0, x]^c\}], \quad x \in (0, \infty]^L.$$

There exists a zero-mean Gaussian random field $Z(x)$, $x \in (0, \infty]^L$, with continuous sample paths and covariance function

$$\text{cov}\{Z_N(x^1), Z_N(x^2)\} = \Lambda \{(0, x^1]^c \cap (0, x^2]^c\}, \quad x^1, x^2 \in (0, \infty]^L,$$

such that $Z_N(x)$ converges weakly to $Z(x)$ in the space of cadlag functions defined on $(0, \infty]^L$ equipped with the Skorohod topology.

Now let δ be a proper scoring rule satisfying the regularity conditions of Theorem 4.1 of Dawid et al. (2016). The maximum scoring rule estimator $\hat{\theta}_{k_u}^\delta$ is defined by

$$\sum_{\{n: N_u Y^n / N \in A_r(1)\}} \nabla_\theta \delta \left(\hat{\theta}_{\delta, N_u}^r, N_u Y^n / N \right) = 0,$$

Appendix B. Supplementary material for Chapter 2

Table B.3 – Estimated standard deviation for the spectral likelihood (left), censored log-likelihood (second) and the gradient score with weights w^1 (third) and w^2 (right). Inference is based on the top 1% of 10000 simulated Brown–Resnick processes with semi-variogram $\gamma(s, s') = (\|s - s'\|/\tau)^\alpha / 2$. In each case the scale parameter equals $\tau = 2.5$ and grids are regular of sizes 10×10 , 20×10 and 20×15 . All the values were multiplied by 10^2 . “NC” means that optimization does not converge.

Shape α			
Grid size	10×10	20×10	20×15
$\alpha = 0.5$	1.1/3.0/2.6/2.6	0.78/2.0/1.7/1.7	0.70/2.4/1.6/1.6
$\alpha = 1$	1.7/7.2/3.7/1.1	2.7/7.9/2.5/2.4	4.1/7.5/2.2/2.2
$\alpha = 1.3$	3.7/15/5.3/3.7	1.1/13/2.6/2.8	3.6/11/2.7/2.8
$\alpha = 1.8$	NC	1.4/33/3.5/3.7	1.3/34/2.6/3.3

Scale τ			
Grid size	10×10	20×10	20×15
$\alpha = 0.5$	16/17/23/24	12/13/14/15	12/13/13/14
$\alpha = 1$	14/22/18/23	12/30/10/12	18/33/10/11
$\alpha = 1.3$	3.6/381/19/32	7.4/24/6.2/8.1	1.3/27/6.4/7.5
$\alpha = 1.8$	NC	9.1/38/4.9/8.7	7.0/34/4.1/5.6

where $\delta(\theta, y) = \delta(\lambda_{\theta, u}^r, y)$, and which is equivalent to

$$\int_{A_r(1)} \nabla_{\theta} \delta(\hat{\theta}_{\delta, N_u}^r, y) \tilde{\Lambda}_{N_u}(dy) = 0.$$

The second-order condition in the hypotheses of Theorem 4.1 in Dawid et al. (2016) allows us to use a Taylor expansion around θ_0 , yielding

$$0 = \int_{A_r(1)} \nabla_{\theta} \delta(\theta_0, y) \tilde{\Lambda}_{N_u}(dy) + (\hat{\theta}_{\delta, N_u}^r - \theta_0) \int_{A_r(1)} \nabla_{\theta}^2 \delta(\theta_0, y) \tilde{\Lambda}_{N_u}(dy) + o(\hat{\theta}_{\delta, N_u}^r - \theta_0).$$

Also equation (B.6) ensures the convergence in probability

$$\int_{A_r(1)} \nabla_{\theta}^2 \delta(\theta_0, y) \tilde{\Lambda}_{N_u}(dy) \rightarrow \mathbb{E}_P \left\{ \frac{\partial^2 \delta}{\partial \theta^2}(\theta_0) \right\} = K, \quad N \rightarrow \infty,$$

and using the convergence of Z_N , we get the convergence in distribution

$$N_u^{1/2} \int_{A_r(1)} \nabla_{\theta} \delta(\theta_0, y) \tilde{\Lambda}_{N_u}(dy) \rightarrow \mathcal{N} \left[0, \mathbb{E}_P \left\{ \frac{\partial \delta}{\partial \theta}(\theta_0) \frac{\partial \delta}{\partial \theta^T}(\theta_0) \right\} \right], \quad N \rightarrow \infty.$$

Then it is straightforward to establish the convergence in distribution

$$N_u^{-1/2} \left(\hat{\theta}_{\delta, N_u}^r - \theta_0 \right) \rightarrow \mathcal{N} \left\{ 0, K^{-1} J (K^{-1})^T \right\}, \quad N \rightarrow \infty,$$

with $J = \mathbb{E}_P \left\{ \partial \delta(\theta_0) / \partial \theta \partial \delta(\theta_0) / \partial \theta^T \right\}$.

B.7 Pareto process simulation

The comparison of the performances of our estimators in Section 4 of the paper requires the simulation of an r -Pareto process P . We perform this for L locations over $S = [0, 100]^2$ with semi-variogram γ and risk function $r(x) = L^{-1} \sum_{l=1}^L x(s_l)$ as follows:

- for locations $\{s_1, \dots, s_L\} \in S$, choose $l \in \{1, \dots, L\}$ uniformly at random;
- for a given semi-variogram $\gamma(s, s')$, $s, s' \in [0, 100]^2$, generate an $(L-1)$ -dimensional Gaussian vector Z with covariance matrix

$$\Sigma = \{ \gamma(s_i, s_l) + \gamma(s_j, s_l) - \gamma(s_i, s_j) \}_{i, j \in \{1, \dots, L\} \setminus \{l\}}$$

and mean $\mu = \{-\gamma(s_i, s_l)\}_{i \in \{1, \dots, L\} \setminus \{l\}}$, i.e., conditional on the value at s_l ;

- set $W_l = 1$ and $W_1 = \exp(Z_1), \dots, W_{l-1} = \exp(Z_{l-1}), W_{l+1} = \exp(Z_l), \dots, W_L = \exp(Z_{L-1})$;
- generate a Pareto random variable U with distribution function $1 - 1/x$ ($x > 1$) and set $P = UW / \{(L-1) \|W\|_1\}$;
- return P .

C Supplementary material for Chapter 3

C.1 Proof of Theorem 3.2

Condition (3.3) implies that the exponent measure ν of Z , defined by

$$\nu(E) = \mathbb{E} \left(\int_0^\infty u^{-2} \mathbf{1}\{uW(s) \in E\} du \right), \quad E \in C_+(S), \quad (\text{C.1})$$

where $C_+(S)$ and $C_+(S)$ denote the analogues to $C(S)$ and $C(S)$ for non-negative functions, satisfies

$$\nu(E) = \begin{cases} \lim_{t \rightarrow \infty} t \Pr \left(\left[\left\{ \operatorname{sgn}(\xi) \frac{X(s) - b_s(t)}{a_s(t)} \right\}_+^{1/\xi}, s \in S \right] \in E \right), & \xi \neq 0, \\ \lim_{t \rightarrow \infty} t \Pr \left(\left[\exp \left\{ \frac{X(s) - b_s(t)}{a_s(t)} \right\}, s \in S \right] \in E \right), & \xi = 0. \end{cases} \quad (\text{C.2})$$

Closely related to the process W , the measure ν incorporates the extremal dependence structure of X .

For the Fréchet case, by Proposition 1.11 in Resnick (1987), $b_s(t) \equiv 0$ and $a_t(s) = \inf\{x \in \mathbb{R} : \Pr\{X(s) \leq x\} \geq 1 - 1/t\}$ are valid choices for the norming constants. In particular, $a(t) \rightarrow \infty$ as $t \rightarrow \infty$. Thus, for any $\varepsilon > 0$, $a(t)^{-1} |\inf_{s \in S} x_*(s)| < \varepsilon$ for sufficiently large t . As the continuous function A is strictly positive and thus bounded away from zero on the compact domain S , by (3.6), we also have $|a(t)^{-1} a_s(t) - A(s)| < \varepsilon A(s)$ for all $s \in S$ and sufficiently large t . We obtain the uniform bounds

$$\begin{aligned} \frac{X(s)}{a(t)} &= \frac{a_s(t)}{a(t)} \left\{ \frac{X(s)}{a_s(t)} \right\}_+ - \left\{ \frac{X(s)}{a(t)} \right\}_- \\ &\geq \frac{a_s(t)}{a(t)} \left\{ \frac{X(s)}{a_s(t)} \right\}_+ - \left| \inf_{s \in S} \frac{x_*(s)}{a(t)} \right| \end{aligned}$$

Appendix C. Supplementary material for Chapter 3

$$\geq (1 - \varepsilon)A(s) \left\{ \frac{X(s)}{a_s(t)} \right\}_+ - \varepsilon,$$

and

$$\frac{X(s)}{a(t)} \leq (1 + \varepsilon)A(s) \left\{ \frac{X(s)}{a_s(t)} \right\}_+.$$

As each r_j ($j = 1, \dots, L$) is uniformly continuous, there exists a function $h : (0, \infty) \rightarrow (0, \infty)$, $\lim_{\varepsilon \searrow 0} h(\varepsilon) = 0$, such that $\sup_{j=1, \dots, L} |r_j(x) - r_j(x')| \leq h(\varepsilon)$ for all $x, x' \in C_+(S)$ such that $\|x - x'\|_\infty \leq \varepsilon$. The monotonicity and homogeneity of each r_j ($j = 1, \dots, L$) entail

$$(1 - \varepsilon)r_j \left[\left\{ \frac{X}{a(t)} \right\}_+ A \right] - h(\varepsilon) \leq \frac{r_j(X)}{a(t)} \leq (1 + \varepsilon)r_j \left[\left\{ \frac{X}{a(t)} \right\}_+ A \right].$$

With $\varepsilon \searrow 0$, for $x_1, \dots, x_L > 0$, we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} t \Pr \left[\bigcup_{j=1}^L \left\{ \frac{r_j(X)}{a(t)} > x_j \right\} \right] &= \lim_{t \rightarrow \infty} t \Pr \left\{ \bigcup_{j=1}^L \left(r_j \left[\left\{ \frac{X}{a(t)} \right\}_+ A \right] > x_j \right) \right\} \\ &= \lim_{t \rightarrow \infty} t \Pr \left[\bigcup_{j=1}^L \left\{ r_j \left(\left[\left\{ \frac{X}{a(t)} \right\}_+^{1/\xi} \right]^\xi A \right) > x_j \right\} \right] \\ &= \nu \left(\bigcup_{j=1}^L \left\{ x \in C_+(S) : r_j(x^\xi A) > x_j \right\} \right) \\ &= \mathbb{E} \left\{ \int_0^\infty u^{-2} \mathbf{1} \left(\bigcup_{j=1}^L \left[r_j \{ (uW)^\xi A \} > x_j \right] \right) du \right\} \\ &= \mathbb{E} \left[\bigvee_{j=1}^L \left\{ \frac{r_j(W^\xi A)}{x_j} \right\}^{1/\xi} \right], \end{aligned}$$

where we used (C.2) and (C.1). In the Weibull case, by Proposition 1.13 in Resnick (1987), we may choose $b_s(t) = x^*(s)$. Then, by the linearity of each r_j , we have

$$\frac{r_j(X) - r_j(x^*)}{a(t)} = r_j \left\{ \frac{a(t)}{a(t)} \frac{X - x^*}{a(t)} \right\} = r_j \left[-\frac{a(t)}{a(t)} \left\{ -\frac{X - x^*}{a(t)} \right\}_+ \right].$$

The rest follows analogously to the Fréchet case. In the Gumbel case, the integral in the proof of Theorem 2.1 in Ferreira et al. (2012) can just be replaced by the linear functionals r_1, \dots, r_L to obtain that

$$\lim_{t \rightarrow \infty} t \Pr \left(\bigcup_{j=1}^L \left[\frac{r_j(X) - r_j\{b(t)\}}{a(t)} > x_j \right] \right)$$

C.2. Background and formula related to Hüsler–Reiss distributions

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} t \Pr \left(\bigcup_{j=1}^L \left[r_j \left\{ \frac{X - b.(t)}{a.(t)} A \right\} > x_j \right] \right) \\
&= \lim_{n \rightarrow \infty} t \Pr \left[\bigcup_{j=1}^L \left\{ r_j \left(\log \left[\exp \left\{ \frac{X - b.(t)}{a.(t)} \right\} \right] A \right) > x_j \right\} \right] \\
&= \nu \left[\bigcup_{j=1}^L \{x \in C_+(S) : r_j(A \log x) > x_j\} \right],
\end{aligned}$$

for $x_1, \dots, x_L \in \mathbb{R}$. Using its definition in (C.1), the exponent measure can be calculated yielding

$$\begin{aligned}
&\nu \left[\bigcup_{j=1}^L \{x \in C_+(S) : r_j(A \log x) > x_j\} \right] \\
&= \mathbb{E} \left[\int_0^\infty u^{-2} \mathbb{1} \left[\bigcup_{j=1}^L \left\{ \log u > \frac{x_j - r_j(A \log W)}{r_j(A)} \right\} \right] du \right] \\
&= \mathbb{E} \left[\bigvee_{j=1}^L \exp \left\{ \frac{x_j - r_j(A \log W)}{r_j(A)} \right\} \right].
\end{aligned}$$

Replacing x_j by $x_j r_j(A)$ closes the proof.

C.2 Background and formula related to Hüsler–Reiss distributions

C.2.1 Hüsler–Reiss distributions

The class of Brown–Resnick processes has a similar role in spatial extreme value statistics as the class of Gaussian processes in classical geostatistics. In order to specify their finite-dimensional distributions, we recall a popular model in multivariate extreme value theory, namely the Hüsler–Reiss distribution (Hüsler and Reiss, 1989). An m -dimensional max-stable random vector (Z_1, \dots, Z_m) with distribution function $F_Z(x_1, \dots, x_m) = \exp\{-V(x_1, \dots, x_m)\}$ is Hüsler–Reiss distributed with Gumbel margins and strictly conditionally negative definite parameter matrix $\Gamma \in [0, \infty)^{m \times m}$ if its exponent measure has the form

$$V(x_1, \dots, x_m) = \mathbb{E} \left[\max_{j=1, \dots, m} \exp \left\{ -x_j + Y_j - \frac{1}{2} \text{var}(Y_j) \right\} \right], \quad (\text{C.3})$$

Appendix C. Supplementary material for Chapter 3

for a centred Gaussian random vector (Y_1, \dots, Y_m) with variogram matrix $\Gamma_{jk} = \mathbb{E}\{(Y_j - Y_k)^2\}$, $j, k = 1, \dots, m$. In this case, one possible choice for the covariance matrix of Y is

$$\Sigma = \frac{1}{2} (\Gamma_{j1} + \Gamma_{k1} - \Gamma_{jk})_{1 \leq j, k \leq m}. \quad (\text{C.4})$$

The exponent measure V is normalized in the sense that $V(\infty, \dots, x_j, \dots, \infty) = \exp(-x_j)$, for any $j = 1, \dots, m$. If Z is a Brown–Resnick process associated to the variogram γ , then the distribution of $\{Z(s_1), \dots, Z(s_m)\}$ is Hüsler–Reiss with parameter matrix $\Gamma = \{\gamma(s_j, s_k)\}_{j, k=1, \dots, m}$.

For censored likelihood estimation of models with Hüsler–Reiss limit, we require the partial derivatives $V_{\mathcal{K}}$ of V in (C.3) with respect to any non-empty subset of variables $\mathcal{K} \subset \{1, \dots, m\}$. Let $b \in \{1, \dots, m\}$ be the number of components that exceed their thresholds, and, without loss of generality, let $\mathcal{K} = \{1, \dots, b\}$. Based on the results in Engelke et al. (2015), Wadsworth and Tawn (2014) and Asadi et al. (2015, Section 4.3.2), we obtain the representation

$$-V_{\mathcal{K}}(z) = \exp(-z_1) \varphi_{b-1}(\tilde{z}_{2:b}; \Sigma_{2:b}) \Phi_{m-b}\{\mu_C, \Sigma_C\}, \quad (\text{C.5})$$

where $\tilde{z} = \{(z_j - z_1) + \Gamma_{1j}/2\}_{1 \leq j \leq m}$, Σ is as in (C.4) and $\varphi_k(\cdot, \Psi)$ and $\Phi_k(\cdot, \Psi)$ are the multivariate density and distribution function of a k -variate normal distribution with mean zero and covariance Ψ . We use the convention that $\varphi_0 \equiv 1$ if $b = 1$ and $\Phi_0 \equiv 1$ if $b = m$. The mean and covariance matrix are

$$\begin{aligned} \mu_C &= \tilde{z}_{(b+1):m} - \Sigma_{(b+1):m, 2:b} \Sigma_{2:b, 2:b}^{-1} \tilde{z}_{2:b}, \\ \Sigma_C &= \Sigma_{(b+1):m, (b+1):m} - \Sigma_{(b+1):m, 2:b} \Sigma_{2:b, 2:b}^{-1} \Sigma_{2:b, (b+1):m}. \end{aligned}$$

C.2.2 Explicit formulas for extremes of aggregated data

When the underlying process X is in the domain of attraction of a Brown–Resnick process with Gumbel margins, we can obtain explicit formula for the r -extremal coefficient and the multivariate limits for certain aggregation functionals.

Throughout this section we work with the general assumptions and notation in Section 3.2 of the paper, and concentrate on the case where $\xi = 0$ and the limiting process Z is Brown–Resnick on a compact region $S \subset \mathbb{R}^D$. We further assume that the variogram γ as defined in Example 3.1 depends on the spatial lag $s - t$ only and we therefore write $\gamma(s - t)$ for $\gamma(s, t)$. Then, without loss of generality, we may assume

C.2. Background and formula related to Hüsler–Reiss distributions

that $G(0) = 0$ and the spectral function simplifies to

$$W(s) = \exp\{G(s) - \gamma(s)/2\}, \quad s \in S.$$

We first prove the closed-form expression of the r -extremal coefficient θ_0^{avg} , where r is a spatial average over the region S ; see Example 3.6. With $\bar{A} = \int_S A(s) \, ds$, it follows from Theorem 3.1 that

$$\theta_0^{\text{avg}} = \mathbb{E} \left(\exp \left[\frac{1}{\bar{A}} \int_S \{G(s) - \gamma(s)/2\} A(s) \, ds \right] \right) = \exp \left\{ \frac{\sigma_{\text{avg}}^2}{2} - \frac{1}{2\bar{A}} \int_S A(s) \gamma(s) \, ds \right\}, \quad (\text{C.6})$$

since the integral over a Gaussian process is normally distributed with variance

$$\sigma_{\text{avg}}^2 = \text{var} \left\{ \frac{\int_S A(s) G(s) \, ds}{\bar{A}} \right\} = \frac{1}{\bar{A}} \int_S A(s) \gamma(s) \, ds - \frac{1}{2\bar{A}^2} \int_S \int_S A(s) A(t) \gamma(s-t) \, ds \, dt,$$

which is a simple extension of Wackernagel (2003, p. 67–69). Plugging this into (C.6) yields formula (3.9).

For censored likelihood inference in Section 3.3.3 and conditional or unconditional simulation described in Section 3.4 in the paper, the multivariate limit behaviour of different functions is required. We consider here the case that is used in the application, namely that the aggregation functionals are either spatial averages over compact regions $S_l \subset S$, ($l = 1, \dots, L$), or point evaluations at locations $s_k \in S$, ($k = 1, \dots, K$), i.e.,

$$r_j(X) = \begin{cases} \frac{1}{|S_j|} \int_{S_j} X(s) \, ds, & j = 1, \dots, L, \\ X(s_{j-L}), & j = L+1, \dots, L+K. \end{cases} \quad (\text{C.7})$$

The vector $(r_1(X), \dots, r_{L+K}(X))$ then satisfies the assumptions of Theorem 3.2 and its limiting exponent measure \tilde{V} is the right-hand side of (3.11). This exponent measure is not normalized, since by Theorem 3.1, $\tilde{V}(\infty, \dots, x_j, \dots, \infty) = \exp(-x_j + \log \theta_0^{r_j})$, and $\log \theta_0^{r_j}$ is given by (3.9) for $j = 1, \dots, L$, and is equal to 0 for $j = L+1, \dots, L+K$. We therefore define the corresponding normalized exponent measure by

$$\begin{aligned} V(x_1, \dots, x_{L+K}) &= \mathbb{E} \left\{ \max_{j=1, \dots, L+K} \exp \left(-x_j + \frac{r_j\{(G - \gamma/2)A\}}{r_j(A)} - \log \theta_0^{r_j} \right) \right\} \\ &= \mathbb{E} \left\{ \max_{j=1, \dots, L+K} \exp \left(-x_j + \frac{r_j(GA)}{r_j(A)} - \frac{1}{2} \text{var} \left[\frac{r_j(GA)}{r_j(A)} \right] \right) \right\}, \end{aligned} \quad (\text{C.8})$$

where the second equality follows from (C.6). Since all aggregation functionals are either spatial averages or point evaluations, and the vector (Y_1, \dots, Y_{L+K}) with $Y_j = r_j(GA)/r_j(A)$, $j = 1, \dots, L+K$, is multivariate Gaussian, we recognize in (C.8) the

Appendix C. Supplementary material for Chapter 3

exponent measure of a Hüsler–Reiss distribution with parameter matrix Γ where $\Gamma_{jk} = E(Y_j - Y_k)^2$, $j, k = 1, \dots, L + K$. We can separate Γ into different blocks such that

$$\Gamma = \begin{pmatrix} \{\Gamma_{jk}\}_{j,k} & \{\Gamma_{jq}\}_{j,q} \\ \{\Gamma_{pk}\}_{p,k} & \{\Gamma_{pq}\}_{p,q} \end{pmatrix}, \quad \begin{cases} j, k = 1, \dots, L, \\ p, q = L + 1, \dots, L + K. \end{cases} \quad (\text{C.9})$$

We directly see that $\Gamma_{pq} = \gamma(s_{p-L} - s_{q-L})$ for $p, q = L + 1, \dots, L + K$. Since Γ is symmetric, letting $\bar{A}_j = \int_{S_j} A(s) ds$, $j = 1, \dots, L$, it suffices to compute

(i) for $j, k = 1, \dots, L$,

$$\begin{aligned} \Gamma_{jk} &= \frac{1}{\bar{A}_j \bar{A}_k} \int_{S_j} \int_{S_k} A(s) A(t) \gamma(s - t) ds dt - \frac{1}{2\bar{A}_j^2} \int_{S_j} \int_{S_j} A(s) A(t) \gamma(s - t) ds dt \\ &\quad - \frac{1}{2\bar{A}_k^2} \int_{S_k} \int_{S_k} A(s) A(t) \gamma(s - t) ds dt; \end{aligned}$$

(ii) for $j = 1, \dots, L$, $q = L + 1, \dots, L + K$,

$$\Gamma_{jq} = \frac{1}{\bar{A}_j} \int_{S_j} A(s) \gamma(s - s_{q-L}) ds - \frac{1}{2\bar{A}_j^2} \int_{S_j} \int_{S_j} A(s) A(t) \gamma(s - t) ds dt.$$

In order to show (i), we note that for $s, t \in S$,

$$\text{var} \left\{ \frac{A(s)G(s)}{\bar{A}_j} - \frac{A(t)G(t)}{\bar{A}_k} \right\} = \frac{A(s)^2}{\bar{A}_j^2} \gamma(s) + \frac{A(t)^2}{\bar{A}_k^2} \gamma(t) - \frac{A(s)A(t)}{\bar{A}_j \bar{A}_k} \{ \gamma(s) + \gamma(t) - \gamma(s - t) \}, \quad (\text{C.10})$$

since $E\{G(s)^2\} = \gamma(s)$ and $E\{G(s)G(t)\} = 1/2 \{ \gamma(s) + \gamma(t) - \gamma(s - t) \}$ and we use the following formula (Wackernagel, 2003, p. 67–69)

$$\begin{aligned} \Gamma_{jk} &= \text{var} \left\{ \frac{1}{\bar{A}_j} \int_{S_j} A(s)G(s) ds - \frac{1}{\bar{A}_k} \int_{S_k} A(t)G(t) dt \right\} \\ &= \int_{S_j} \int_{S_k} \text{var} \left\{ \frac{A(s)G(s)}{\bar{A}_j} - \frac{A(t)G(t)}{\bar{A}_k} \right\} ds dt - \\ &\quad \frac{1}{2} \int_{S_j} \int_{S_j} \text{var} \left\{ \frac{A(s)G(s)}{\bar{A}_j} - \frac{A(t)G(t)}{\bar{A}_j} \right\} ds dt - \\ &\quad \frac{1}{2} \int_{S_k} \int_{S_k} \text{var} \left\{ \frac{A(s)G(s)}{\bar{A}_k} - \frac{A(t)G(t)}{\bar{A}_k} \right\} ds dt. \end{aligned} \quad (\text{C.11})$$

Using (C.10), the first term in the last equation equals

$$\begin{aligned} & \frac{|S_k|}{\bar{A}_j^2} \int_{S_j} A(s)^2 \gamma(s) \, ds - \frac{|S_k|}{\bar{A}_j} \int_{S_j} A(s) \gamma(s) \, ds + \\ & \frac{|S_j|}{\bar{A}_k^2} \int_{S_k} A(s)^2 \gamma(s) \, ds - \frac{|S_k|}{\bar{A}_k} \int_{S_k} A(s) \gamma(s) \, ds + \\ & \frac{1}{\bar{A}_j \bar{A}_k} \int_{S_j} \int_{S_k} A(s) A(t) \gamma(s-t) \, ds \, dt. \end{aligned}$$

Similarly, the second term in (C.11) is

$$\frac{2|S_j|}{\bar{A}_j^2} \int_{S_j} A(s)^2 \gamma(s) \, ds - \frac{2|S_j|}{\bar{A}_j} \int_{S_j} A(s) \gamma(s) \, ds + \frac{1}{\bar{A}_j^2} \int_{S_j} \int_{S_j} A(s) A(t) \gamma(s-t) \, ds \, dt,$$

and analogously for the third term. Putting this together, we obtain the formula in (i). Very similar calculations yield the result in (ii).

The above calculation shows that the $(L+K)$ -dimensional vector in (C.7) is in the max-domain of attraction of a Hüsler–Reiss distribution with known parameter matrix, and we can use the inference and simulation methodology described in Section 3.3 and Section 3.4 in the paper and in the literature.

C.3 Simulation study

C.3.1 Gumbel case

We now apply our downscaling approach to a simple model that resembles the setup in the application in Section 3.5 of the paper. We suppose that we observe independent data X_1, \dots, X_n from a process X on $S = [0, 5]^2$, but only through aggregating functionals r_j ($j = 1, \dots, L$) with $L = 25$, which we will take to be spatial averages. The observations are thus 25-dimensional and of the form

$$\left(\frac{1}{|S_1|} \int_{S_1} X_i(s) \, ds, \dots, \frac{1}{|S_{25}|} \int_{S_{25}} X_i(s) \, ds \right), \quad i = 1, \dots, n,$$

where $S_j = [s_1^j, s_1^j + 1] \times [s_2^j, s_2^j + 1]$, with $s_1^j, s_2^j \in \{0, \dots, 4\}$, i.e., a regular grid of 1×1 squares. We consider X in the Gumbel ($\xi = 0$) max-domain of attraction of a Brown–Resnick process associated to the semi-variogram model

$$\gamma(s, t) = \left(\frac{\|s - t\|_2}{\lambda} \right)^\alpha, \quad \alpha = 1.5, \lambda = 1.$$

Appendix C. Supplementary material for Chapter 3

We impose a linear structure on the unknown functions A and B of the margins appearing in the setting described in Section 3.3.1,

$$\begin{aligned} A(s_1, s_2) &= a_0 + a_1 s_1 = 0.8 + 0.4 s_1, \\ B(s_1, s_2) &= b_0 + b_2 s_2 = -0.4 + 0.8 s_2, \end{aligned} \quad (s_1, s_2) \in [0, 5]^2,$$

where the parameters were chosen such that $r_1(A) = 1$ and $r_1(B) = 0$.

By Theorem 3.2 and Example 3.7, the vector of aggregated data $(r_1(X), \dots, r_L(X))$ is in the max-domain of attraction of a multivariate Hüsler–Reiss distribution with dependence matrix Γ described in Section C.2.2 and normalizing vectors $\{\mu_{j,t}\}_{j=1}^L$ and $\{\sigma_{j,t}\}_{j=1}^L$ as given in Equations (3.15) and (3.16), respectively. Such a vector of aggregated data can be simulated as follows.

1. Randomly select $j_0 \in \{1, \dots, L\}$.
2. Generate an exponential variable $U \sim \text{Exp}(1)$.
3. Generate a $(L-1)$ -dimensional Gaussian vector G with covariance matrix $\Sigma = (1/2)\{\Gamma_{jj_0} + \Gamma_{kk_0} - \Gamma_{jk}\}_{j,k \neq j_0}$ and mean $\mu = -\{\Gamma_{jj_0}/2\}_{j \neq j_0}$.
4. Set $\tilde{G}_{j_0} = 0$, $\tilde{G}_{-j_0} = G$ and $\tilde{Y} = \{U + \tilde{G} - \log \|\exp(\tilde{G})\|_1\} + \log L$.
5. Set $Y_j = r_j(A) \left(\tilde{Y}_j + \log \theta_0^{r_j} \right) + r_j(B)$ for $j = 1, \dots, L$.
6. Return $Y = (Y_1, \dots, Y_L)$.

We simulate $n = 10^4$ samples Y_1, \dots, Y_n of the random vector Y , which are then used to estimate the model parameters via the independence and full censored likelihood procedures. We could also have simulated the process X on a fine grid and then aggregated it over the squares. This approach, used in Section C.3.2 and Section C.3.3, is computationally less efficient and gives essentially the same results as those presented below.

For the independence likelihood procedure, we use the peaks-over-threshold representation described in Section 3.3.2 of the paper: for each grid cell, observations below the local 0.99 empirical quantile are censored, yielding exactly 100 exceedances with parameters $a(t) = 1$ and $b(t) = \log 10000$ with $t = 10000$.

For the censored likelihood method, we choose a threshold vector $u \in \mathbb{R}^L$, also based on the local 0.99 empirical quantiles, such that we keep the N_u observations satisfying $\max(Y_{i,1}/u_1, \dots, Y_{i,L}/u_L) \geq 1$ ($i = 1, \dots, n$). In this setting, the theoretical value of $a(t)$

Table C.1 – Relative root mean square error (%) with respect to the true parameters value for estimates based on censored and independence likelihood methods for the Gumbel domain of attraction. Inference is performed based on $n = 10^4$ simulated data, from which only vectors with at least one component exceeding its local empirical 0.99 quantile are used.

	$a(t)$	a_0	a_1	$b(t)$	b_0	b_2	α	λ	Mean
Full censored LLH	1.2	0.6	2.2	0.3	2.9	2.9	1.3	2.0	1.7
Independence LLH	1.8	1.0	3.9	0.3	3.3	3.3	2.0	2.7	2.3

and $b(t)$ are the same as for the independence likelihood case. We use these N_u exceedances in the full censored likelihood procedure described by Equation (3.19).

By repeating inference a 100 times, we confirm that both methods are unbiased and the results in Table C.1 show that all parameters can be estimated accurately. The censored likelihood approach, which makes use of the multivariate tail distribution, outperforms the independence likelihood procedure by about 30%.

C.3.2 Weibull case

We reproduce the simulation study described in Section C.3.1, but we now consider X in the Weibull domain of attraction ($\xi = -0.3$) of the same Brown–Resnick process. In this case, ξ is considered as unknown and needs to be estimated. A closed form for the joint tail distribution of the aggregated process is not available, so only the independence likelihood procedure is possible. Consequently, simulation of the aggregated vector $(r_1(X), \dots, r_L(X))$ cannot be performed exactly, and thus is approximated by aggregating samples of X on a regularly spaced dense grid with $D \gg 25$ locations.

1. Randomly select $j_0 \in \{1, \dots, D\}$.
2. Generate a univariate exponential variable $U \sim \text{Pareto}(1, 1)$.
3. Generate a $D - 1$ -dimensional Gaussian vector G with covariance matrix $\Sigma = (1/2)\{\Gamma_{jj_0} + \Gamma_{kk_0} - \Gamma_{jk}\}_{j,k \neq j_0}$ and mean $\mu = -\{\Gamma_{jj_0}/2\}_{j \neq j_0}$.
4. Set $\tilde{G}_{j_0} = 0$, $\tilde{G}_{-j_0} = G$ and $\tilde{X} = \text{sng}(\xi)[U \exp(G) / \{D^{-1} \|\exp(\tilde{G})\|_1\}]^\xi$.
5. Compute $\tilde{Y} = (r_1(\tilde{X}), \dots, r_L(\tilde{X}))$.
6. Set $Y_j = r_j(A)\{\tilde{Y}_j + \log \hat{\theta}_0^{r_j}\} + r_j(B)$ for $j = 1, \dots, L$.

Appendix C. Supplementary material for Chapter 3

7. Return $Y = (Y_1, \dots, Y_L)$.

In step 6, numerical evaluation of r -extremal coefficient is required. Following the representation of equation (3.8), $\hat{\theta}_0^{rj}$ is estimated by replacing the expectation with an empirical mean of replicates W_i , $i = 1, \dots, 1000$, of the spectral function sampled on a dense grid.

Inference is performed on $n = 10^4$ replicates and using a 0.99 empirical quantile for censoring. Similarly to the Gumbel case, the estimator is unbiased and the results in Table C.2 show that all parameters can be estimated accurately. We emphasize the good performance for the shape parameter that is a consequence of the exploitation of the multivariate structure of the problem.

Table C.2 – Relative root mean square error (%) with respect to the true parameters value for estimates based on censored and independence likelihood methods for the Weibull domain of attraction ($\xi = -0.3$). Inference is performed based on $n = 10^4$ simulated data, from which only vectors with at least one component exceeding its local empirical 0.99 quantile are used.

ξ	$a(t)$	a_0	a_1	$b(t)$	b_0	b_2	α	λ	Mean
4.7	2.7	0.4	1.7	0.6	0.3	0.3	0.5	0.6	1.6

C.3.3 Fréchet case

In this Section, we consider X in the Fréchet domain of attraction ($\xi = 0.1$) of the Brown–Resnick process described in Section C.3.1. As explained in Theorem 3.1, we suppose that $b(n)B(t) \equiv 0$, and the parametric model for the scale

$$A(s_1, s_2) = a_0 + a_1 s_1 + a_2 s_2 = 0.8 + 0.4 s_1 + 0.2 s_2, \quad (s_1, s_2) \in [0, 5]^2.$$

For positive tail index, the aggregation functional r does not need to be linear, thus to illustrate the flexibility of our results, we consider the functional in Example 3.5, i.e., the observations are of the form

$$Y = \left(\left\{ \frac{1}{|S_1|} \int_{S_1} X_i^2(s) ds \right\}^{1/2}, \dots, \left\{ \frac{1}{|S_{25}|} \int_{S_{25}} X_i^2(s) ds \right\}^{1/2} \right), \quad i = 1, \dots, n,$$

where $S_j = [s_1^j, s_1^j + 1] \times [s_2^j, s_2^j + 1]$, with $s_1^j, s_2^j \in \{0, \dots, 4\}$. Simulation of the vector Y is done with the algorithm described in Section C.3.2. Marginal likelihood inference is

C.3. Simulation study

Table C.3 – Relative root mean square error (%) with respect to the true parameters value for estimates based on censored and independence likelihood methods for the Fréchet domain of attraction ($\xi = 0.1$). Inference is performed based on $n = 10^4$ simulated data, from which only vectors with at least one component exceeding its local empirical 0.99 quantile are used.

ξ	$a(t)$	a_0	a_1	a_2	α	λ	Mean
2.4	0.2	3.1	7.0	9.8	2.8	5.2	4.4

performed 100 times with $n = 10^4$ and using the 0.99 quantile threshold. The estimator is unbiased and its performance is shown in Table C.3.

C.4 Model assessment for the downscaling application

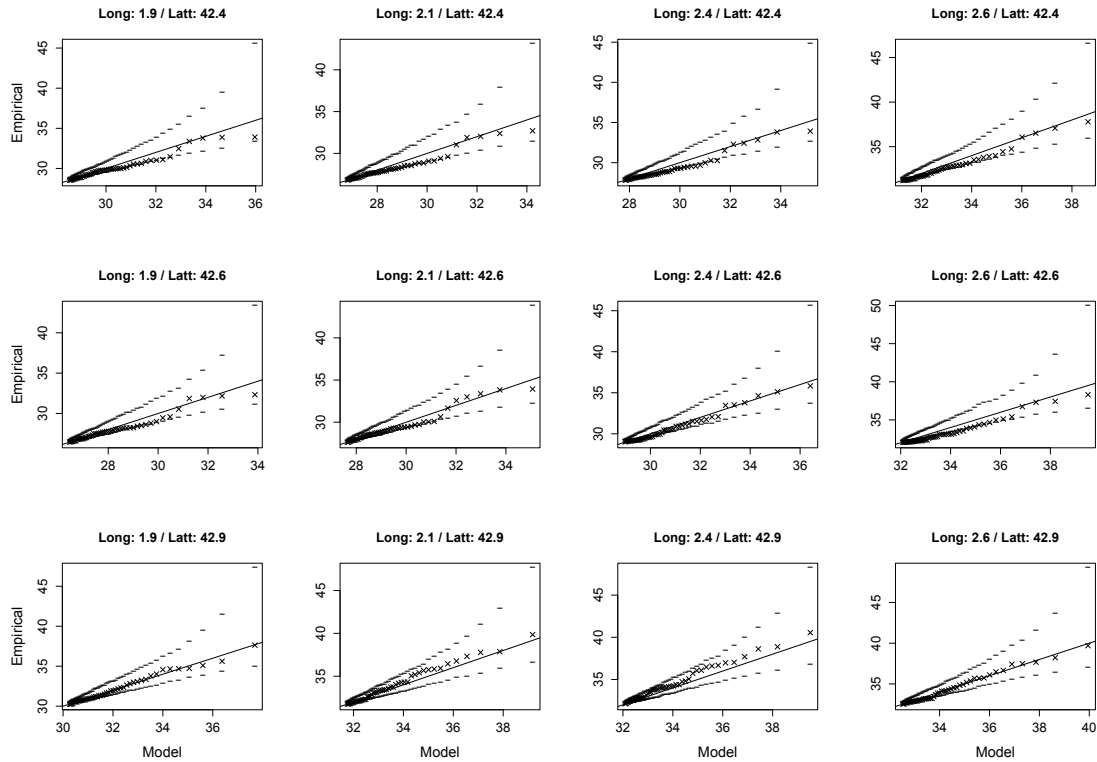


Figure C.1 – Quantile-quantile plots comparing the observations and the fitted marginal distribution for every grid cell. Pointwise confidence intervals are obtained by parametric bootstrap taking into account the uncertainty of the parameter estimates.

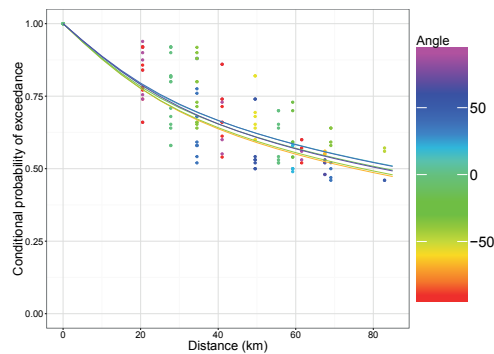


Figure C.2 – Estimated pairwise extremogram (dots) as function of the distance (km) between the centres of the grid cells and direction ($^{\circ}$). The solid lines represent the theoretical extremogram for the estimated anisotropic power variogram.

D Proofs of theoretical results in Chapter 4

D.1 Equation (4.5)

Let $\mathcal{F}_0 = C\{S, [0, \infty)\} \setminus \{0\}$. Take $u_0 \geq 0$ such that $\{y \in \mathcal{F}_0 : y(s_0) > u_0\} \subset \{y \in \mathcal{F}_0 : r(y) \geq 1\}$. Homogeneity of the limiting measure Λ gives

$$\begin{aligned} \Pr\{P(s_0) > r'\} &= \frac{\Lambda\{y \in \mathcal{F}_0 : r(y) \geq 1, y(s_0) \geq r'\}}{\Lambda\{y \in \mathcal{F}_0 : r(y) \geq 1\}} \\ &= \frac{\Lambda\{x \in \mathcal{F}_0 : r(y) \geq 1, y(s_0) \geq r' u_0^{-1} u_0\}}{\Lambda\{y \in \mathcal{F}_0 : r(y) \geq 1\}} \\ &= \left(\frac{r'}{u_0}\right)^{-1/\xi} \frac{\Lambda\{y \in \mathcal{F}_0 : r(y) \geq 1, y(s_0) \geq u_0\}}{\Lambda\{y \in \mathcal{F}_0 : r(y) \geq 1\}}, \end{aligned}$$

because for any $r' \geq 1$

$$\{y \in \mathcal{F}_0 : y(s_0) > r' u_0, r(y) \geq 1\} = r' \times \{y \in \mathcal{F}_0 : y(s_0) > u_0, r(y) \geq 1\}.$$

□

D.2 Theorem 4.1

Let $\mathcal{F}_0 = C\{S, [0, \infty)\} \setminus \{0\}$. For $\xi > 0$, let $r' \in \mathcal{U}_r^\xi$ and $\mathcal{W} \in \mathcal{S}_r^\xi$. The continuous function A is strictly positive and thus bounded away from zero on the compact set S . Hence, for any $\varepsilon > 0$, equation (4.8) gives $|r(a_n)^{-1} a_n(s) - A(s)| < \varepsilon A(s)$ and $|\{b_n(s) - r(b_n)\} / r(a_n) -$

Appendix D. Proofs of theoretical results in Chapter 4

$B(s) < \varepsilon$ for all $s \in S$ and sufficiently large n . Thus, for $\xi \neq 0$,

$$\frac{X - r(b_n)}{r(a_n)} = \frac{a_n}{r(a_n)} \frac{X - b_n}{a_n} + \frac{\{b_n - r(b_n)\}}{r(a_n)} \geq (1 - \varepsilon) A \left(\frac{X - b_n}{a_n} \right)_+ + B - \varepsilon,$$

and analogously

$$\frac{X - r(b_n)}{r(a_n)} \leq (1 + \varepsilon) A \left(\frac{X - b_n}{a_n} \right)_+ + B + \varepsilon.$$

With $\varepsilon \searrow 0$, equation (4.6) leads to

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr \left[\frac{X - r(b_n)}{r(a_n)} \in \cdot \right] &= \lim_{n \rightarrow \infty} n \Pr \left\{ A \frac{X - b_n}{a_n} + B \in \cdot \right\} \\ &= \Lambda \left\{ y \in \mathcal{F}_0 : A \frac{y^\xi - 1}{\xi} + B \in \cdot \right\}. \end{aligned}$$

To check that the sets $\{y \in \mathcal{F}_0 : r(A \frac{y^\xi - 1}{\xi} + B) \geq r'\}$ are bounded away from the singleton $\{0\}$, note that if $\xi > 0$,

$$d_\infty \left[\left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq r' \right\}, \{0\} \right] \geq d_\infty \left[\left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq 0 \right\}, \{0\} \right],$$

by definition of a valid risk functional $r\{B - A\xi^{-1}\} < 0$, and its continuity at $B - A\xi^{-1}$ ensures that

$$d_\infty \left[\left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq 0 \right\}, \{0\} \right] > 0.$$

We can then apply \hat{w} -convergence on these sets, yielding

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr \left\{ \frac{X - r(b_n)}{r(a_n)} \in \left\{ x \in \mathcal{F}^{\xi, A, B} : r(x) \geq r', \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W} \right\} \right\} &= \\ \Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^\xi - 1}{\xi} + B \right) \geq r, \frac{A(y^\xi - 1) + \xi B}{\|A(y^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}. \end{aligned}$$

For $\xi \leq 0$, the hypothesis $r(x) \rightarrow -\infty$ as $x \rightarrow -\infty$ ensures that $\{y \in \mathcal{F}_0 : r(A \log y + B) \geq 0\}$ and $\{y \in \mathcal{F}_0 : r\{A\xi^{-1}(y^\xi - 1) + B\} \geq 0\}$ are bounded away from $\{0\}$. The case $\xi < 0$ is analogous to the Fréchet domain of attraction and for $\xi = 0$, we obtain:

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr \left[\frac{X - r(b_n)}{r(a_n)} \in \left\{ x \in \mathcal{F}^{0, A, B} : r(x) \geq r', x - \|x\|_{\text{ang}} \in \mathcal{W} \right\} \right] &= \\ \Lambda \left\{ y \in \mathcal{F}_0 : r(A \log y + B) \geq r', A \log y + B - \|A \log y + B\|_{\text{ang}} \in \mathcal{W} \right\}. \end{aligned}$$

□

D.3 Corollary 4.1

Recall that $\mathcal{F}_0 = C\{S, [0, \infty)\} \setminus \{0\}$. We start with the results of Theorem 4.1. For $\xi \neq 0$, $r' \in \mathcal{U}^\xi$, and $\mathcal{W} \subset \mathcal{S}_r^\xi$, using the linearity of r_{lin} ,

$$\Lambda\{(r', \mathcal{W})\} = \Lambda\left\{y \in \mathcal{F}_0 : r_{lin}(Ay^\xi) \geq 1 + \xi r', \frac{A(y^\xi - 1) + \xi B}{\|A(y^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W}\right\}.$$

The homogeneity of Λ yields

$$\begin{aligned} \Lambda\{(r', \mathcal{W})\} &= (1 + \xi r')^{-1/\xi} \Lambda\left\{y \in \mathcal{F}_0 : r_{lin}(Ay^\xi) \geq 1, \frac{A(y^\xi - 1) + \xi B}{\|(Ay^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W}\right\} \\ &= (1 + \xi r')^{-1/\xi} \Lambda\{y \in \mathcal{F}_0 : r_{lin}(Ay^\xi) \geq 1\} \times \sigma_r^\xi(\mathcal{W}), \end{aligned}$$

where we define

$$\sigma_{r_{lin}}^\xi(\mathcal{W}) = \frac{\Lambda\left\{y \in \mathcal{F}_0 : r_{lin}(Ay^\xi) \geq 1, \frac{A(y^\xi - 1) + \xi B}{\|A(y^\xi - 1) + \xi B\|_{\text{ang}}} \in \mathcal{W}\right\}}{\Lambda\{y \in \mathcal{F}_0 : r_{lin}(Ay^\xi) \geq 1\}}.$$

Similarly, for $\xi = 0$, the condition $\exp\{r_{lin}(\log x)\} = r_{lin}(x)$ on r_{lin} ensures that

$$\Lambda\{(r', \mathcal{W})\} = \Lambda\{y \in \mathcal{F}_0 : r_{lin}(y \exp A) \geq \exp r', A \log y + B - r_{lin}(A \log y + B) \in \mathcal{W}\},$$

and the homogeneity of Λ yields

$$\begin{aligned} \Lambda\{(r', \mathcal{W})\} &= \exp(-r') \Lambda\{y \in \mathcal{F}_0 : r_{lin}(y \exp A) \geq 1, A \log y + B - r(A \log y + B) \in \mathcal{W}\} \\ &= \exp(-r') \Lambda\{y \in \mathcal{F}_0 : r_{lin}(y \exp A) \geq 1\} \times \sigma_r^\xi(\mathcal{W}), \end{aligned}$$

with

$$\sigma_r^\xi(\mathcal{W}) = \frac{\Lambda\{y \in \mathcal{F}_0 : r_{lin}(y \exp A) \geq 1, A \log y + B - r(A \log y + B) \in \mathcal{W}\}}{\Lambda\{y \in \mathcal{F}_0 : r_{lin}(y \exp A) \geq 1\}}.$$

□

D.4 Derivation of (4.13)

Let $\mathcal{F}_0 = C\{S, [0, \infty)\} \setminus \{0\}$ and $X \in \text{GRV}(\mathcal{F}_0, \xi, a_n, b_n, \Lambda)$. Suppose that for $n \gg 1$, we have $u_n = a_n(s_0)u + b_n(s_0) = u$ with $u > 0$. For $\xi \neq 0$ and $r' \geq 0$, we have

$$\begin{aligned} \Pr\{r_{s_0}(X) - u_n > r' \mid r_{s_0}(X) > u_n\} &= \frac{\Pr\{r_{s_0}(X) > r' + u_n\}}{\Pr\{r_{s_0}(X) > u_n\}} \\ &= \frac{n \Pr\left\{\frac{r_{s_0}(X) - b_n(s_0)}{a_n(s_0)} > \frac{r'}{a_n(s_0)} + u\right\}}{n \Pr\left\{\frac{r_{s_0}(X) - b_n(s_0)}{a_n(s_0)} > u\right\}}. \end{aligned}$$

Then because n is large, we can approximate the probability ratios by their limits, yielding

$$\begin{aligned} \Pr\{r_{s_0}(X) - u_n > r' \mid r_{s_0}(X) > u_n\} &\approx \frac{\left[1 + \xi \left\{\frac{r'}{a_n(s_0)} + u\right\}\right]^{-1/\xi} \times \Lambda\left[\{y \in \mathcal{F}_0 : r(Ay^\xi) \geq 1\}\right]}{\left[1 + \xi u\right]^{-1/\xi} \Lambda\left[\{y \in \mathcal{F}_0 : r(Ay^\xi) \geq 1\}\right]} \\ &\approx \left(1 + \xi \frac{r'}{a_n(s_0)(1 + \xi u)}\right)^{-1/\xi}. \end{aligned}$$

Setting $\sigma(s_0) = a_n(s_0)(1 + \xi u)$ gives the desired expression. The derivation for the Gumbel case is similar. \square

D.5 Derivation of (4.22)

We use similar calculations as for equation (4.5):

$$\begin{aligned} \Pr\{P(s_0) > u'\} &= \Pr\left[P \in \left\{x \in \mathcal{F}^{\xi, A, B} : x(s_0) > u'\right\}\right] \\ &= \frac{\Lambda\left[\left\{1 + \xi A^{-1}(x - B)\right\}^{1/\xi} \in \mathcal{F}^{\xi, A, B} : x(s_0) \geq u', r(x) \geq u\right]}{\Lambda\left\{y \in \mathcal{F}_0 : r\left(A \frac{y^\xi - 1}{\xi} + B\right) \geq u\right\}} \\ &= \frac{\Lambda\left\{y \in \mathcal{F}_0 : A(s_0) \frac{y(s_0)^\xi - 1}{\xi} + B(s_0) \geq u', r\left\{A \frac{y^\xi - 1}{\xi} + B\right\} \geq u\right\}}{\Lambda\left\{y \in \mathcal{F}_0 : r\left(A \frac{y^\xi - 1}{\xi} + B\right) \geq u\right\}} \\ &= \left[1 + \xi \frac{(u' - u_0)}{A(s_0) + \xi\{u_0 - B(s_0)\}}\right]^{-1/\xi} \times \\ &\quad \frac{\Lambda\left\{y \in \mathcal{F}_0 : y(s_0) \geq 1 + \xi A(s_0)^{-1}\{u_0 - B(s_0)\}, r\left\{A \frac{y^\xi - 1}{\xi} + B\right\} \geq u\right\}}{\Lambda\left\{y \in \mathcal{F}_0 : r\left(A \frac{y^\xi - 1}{\xi} + B\right) \geq u\right\}}, \end{aligned}$$

so for any $r' \geq 0$

$$\begin{aligned} \Pr [P(s_0) > r' + u' | P(s_0) > u'] &= \frac{\Pr\{P(s_0) > r' + u'\}}{\Pr\{P(s_0) > u'\}} \\ &= \frac{[1 + \xi A(s_0)^{-1}\{r' + u' - B(s_0)\}]^{-1/\xi}}{[1 + \xi A(s_0)^{-1}\{u' - B(s_0)\}]^{-1/\xi}} \\ &= H_{\xi, \sigma(u')} (r'), \end{aligned}$$

where $\sigma(u') = A(s_0) + \xi\{u' - B(s_0)\}$. For the linear case, we similarly obtain

$$\Pr [P(s_0) > r' + u' | P(s_0) > u'] = H_{\xi^{-1}, \sigma(u')} (r'), \quad r' \geq 0,$$

where $\sigma(u') = \sigma A(s_0) + \xi\{u' - A(s_0)u - B(s_0)\}$ with $u' > u_0$. \square

D.6 Derivation of (4.27)

For $u \geq 0$ and any $r' \geq u$, we simplify the measure

$$\begin{aligned} &\Pr \left[P \in \left\{ x \in \mathcal{F}^{\xi, A, B} : r(x) \geq r', \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W} \right\} \right] \\ &= \frac{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq r', w = \frac{A(y^{\xi-1}) + \xi B}{\|A(y^{\xi-1}) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\}} \\ &= \frac{\Lambda \left[\left\{ 1 + \xi A^{-1}(x - B) \right\}^{1/\xi} \in \mathcal{F}_0 : r(x) \geq r', w = \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W}, x \in \mathcal{F}^{\xi, A, B} \right]}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\}}, \end{aligned}$$

and thus using a chain rule to compute partial derivatives with respect to the elements of the vector x , we get

$$\frac{\partial \Pr(P \in \cdot)}{\partial x} = \frac{\lambda \left\{ \left(1 + \xi \frac{x-B}{A} \right)^{1/\xi} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq u \right\}} \prod_{l=1}^L A(s_l)^{-1} \left(1 + \xi \frac{x - B(s_l)}{A(s_l)} \right)^{1/\xi - 1},$$

which gives (4.27).

For the linear case, let $r' \geq 0$ and $\mathcal{W} \subset \mathcal{S}_r$, then

$$\begin{aligned} &\Pr \left\{ r(P) \geq r', \frac{P}{\|P\|_{\text{ang}}} \in \mathcal{W} \right\} \\ &= H_{\xi^{-1}, \sigma, u} (r') \times \sigma_r(\mathcal{W}), \end{aligned}$$

Appendix D. Proofs of theoretical results in Chapter 4

$$= \left\{ 1 + \xi \frac{r' - u}{\sigma} \right\}^{-1/\xi} \frac{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0, w = \frac{A(y^{\xi-1}) + \xi B}{\|A(y^{\xi-1}) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + \xi B \right) \geq 0 \right\}}.$$

The hypothesis $r(A) = 1$ ensures that

$$\left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\} = \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi}}{\xi} + B \right) \geq \xi^{-1} \right\}.$$

Then the linearity of r and the measure's homogeneity give

$$\begin{aligned} & \Pr \left\{ r(P) \geq r', \frac{P}{\|P\|_{\text{ang}}} \in \mathcal{W} \right\} \\ &= \frac{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi}}{\xi} + B \right) \geq \xi^{-1} + \frac{r' - u}{\sigma}, w = \frac{A(y^{\xi-1}) + \xi B}{\|A(y^{\xi-1}) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\}} \\ &= \frac{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \sigma \frac{y^{\xi-1}}{\xi} + B + Au \right) \geq r', w = \frac{A(y^{\xi-1}) + \xi B}{\|A(y^{\xi-1}) + \xi B\|_{\text{ang}}} \in \mathcal{W} \right\}}{\Lambda^{\xi} \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\}} \\ &= \frac{\Lambda \left[\left\{ 1 + \xi (\sigma A)^{-1} (x - Au - B) \right\}^{1/\xi} \in \mathcal{F}_0 : r(x) \geq r', w = \frac{x}{\|x\|_{\text{ang}}} \in \mathcal{W}, x \in \mathcal{F}^{\xi, A, B} \right]}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(A \frac{y^{\xi-1}}{\xi} + B \right) \geq 0 \right\}}. \end{aligned}$$

Then we proceed similarly as for equation (4.27), yielding

$$\begin{aligned} f_u^r(x) &= \frac{\lambda \left\{ \left(1 + \xi \frac{x - A(s_i)u - B(s_i)}{\sigma A(s_i)} \right)^{1/\xi} \right\}}{\Lambda \left\{ y \in \mathcal{F}_0 : r \left(\sigma A \frac{y^{\xi-1}}{\xi} + Bs \right) \geq u \right\}} \times \\ & \prod_{i=1}^L (A(s_i)\sigma)^{-1} \left(1 + \xi \frac{x - A(s_i)u - B(s_i)}{\sigma A(s_i)} \right)^{1/\xi-1}, \quad x \in \mathcal{A}_r(u). \quad \square \end{aligned}$$

D.7 Poisson process and binomial representation equivalence

Let N_n be the Poisson process

$$\left[A \frac{\{R_n X_n(s)\}^{\xi} - 1}{\xi} + B \right]_{n=1, \dots}$$

D.7. Poisson process and binomial representation equivalence

defined on $\mathcal{F}^{\xi, A, B}$ and where $\{R_n = (1/i), X_n\}_{n=1, \dots}$ is also a Poisson process on $(0, 1] \times \{y \in \mathcal{F}_0 : |S|^{-1} \|y\|_1 \geq 1\}$ with intensity $r^{-2} dr \times \lambda(dx)$. Then $N_n \rightarrow N$ as $n \rightarrow \infty$, and N is a Poisson process on $\mathcal{F}^{\xi, A, B}$ with intensity measure λ given in (4.27).

The Poisson point process N_n lies the set $\{x \in \mathcal{F}^{\xi, A, B} : |S|^{-1} \|(x - B)\xi / A\|^{1/\xi} \|_1 \geq n^{-1}\}$, which is also the set $\mathcal{A}_{\|\cdot\|_1}(n^{-1}) = \{y \in \mathcal{F}_0 : |S|^{-1} \|y\|_1 \geq n^{-1}\}$, so the probability that exactly n points lie in $\mathcal{A}_{\|\cdot\|_1}(n^{-1})$ is

$$\Pr(N = n) = \frac{\Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1})\}^n}{n!} \exp[-\Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1})\}].$$

Similarly, for a functional r and threshold $u \in \mathcal{U}^\xi$, the probability that the number of exceedances N_u equals n_u is

$$\Pr(N_u = n_u) = \frac{\Lambda \{\mathcal{A}_r(u)\}^{n_u}}{n_u!} \exp[-\Lambda \{\mathcal{A}_r(u)\}],$$

where $\mathcal{A}_r(u) = \{y \in \mathcal{F}_0 : r[A(y^\xi - 1)/\xi + B] \geq u\}$. Thus with N_u^c denoting the number of points lying in $\mathcal{A}_{L_1}(n^{-1})$ but not in $\mathcal{A}_r(u)$, we get

$$\begin{aligned} \Pr(N_u = n_u | N = n) &= \frac{\Pr\{N_u = n_u, N_u^c = n - n_u\}}{\Pr\{N = n\}} \\ &= \frac{n!}{n_u!(n - n_u)!} \frac{\Lambda \{\mathcal{A}_r(u)\}^{n_u} \Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1}) \setminus \mathcal{A}_r(u)\}^{n - n_u}}{\Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1})\}^n} \\ &= \binom{n}{n_u} \left(\frac{\Lambda \{\mathcal{A}_r(u)\}}{\Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1})\}} \right)^{n_u} \left(1 - \frac{\Lambda \{\mathcal{A}_r(u)\}}{\Lambda \{\mathcal{A}_{\|\cdot\|_1}(n^{-1})\}} \right)^{n - n_u}, \end{aligned}$$

Finally, we have $\Lambda \{\mathcal{A}_{\|\cdot\|_1}(\frac{1}{n})\} = n \Lambda \{\mathcal{A}_{\|\cdot\|_1}(1)\} = n$, so

$$\Pr(N_u = n_u | N = n) = \binom{n}{n_u} \left[\frac{\Lambda \{\mathcal{A}_r(u)\}}{n} \right]^{n_u} \left[1 - \frac{\Lambda \{\mathcal{A}_r(u)\}}{n} \right]^{n - n_u},$$

and thus we obtain the representation (4.29).

E Supplementary material for Chapter 6

This Appendix gives the detailed plots of the logistic regression modelling the distribution of $1\{r(x) \geq u\}$, the probability of storm occurrence in Europe. The North Atlantic Oscillation (NAO) index and the first and third eigenvalues of the temperature anomaly, shown in Figures E.1, E.2, and E.3, have a significant influence on the occurrence of winter storms at the 0.1% confidence level.

E.1 Diagnostic plots for modelling the frequency of wind storms

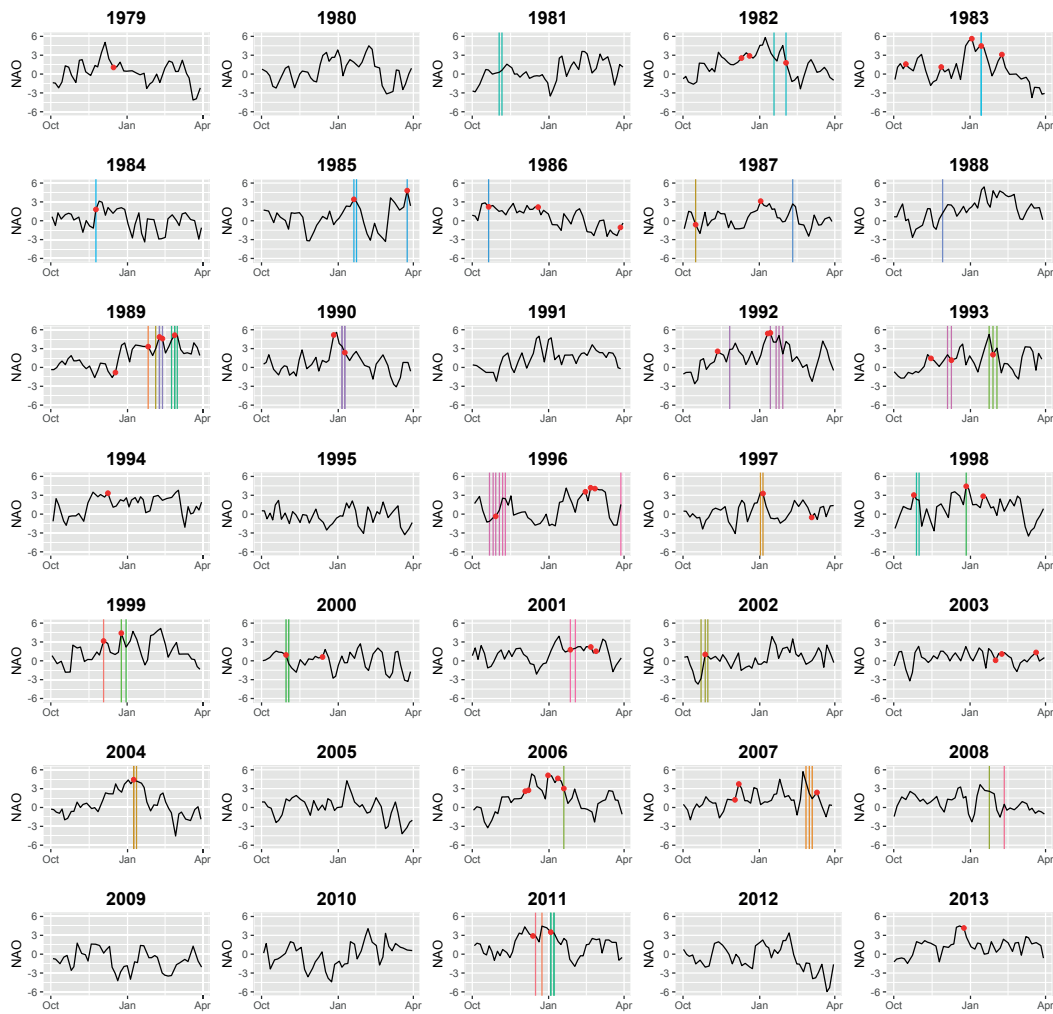


Figure E.1 – Three-hourly North Atlantic Oscillation (NAO) index computed on the ERA–Interim data set for each winter. r -exceedances above the 0.96 empirical quantile are represented by red dots and wind storms from XWS catalogue are represented by vertical lines coloured by dates.

E.1. Diagnostic plots for modelling the frequency of wind storms

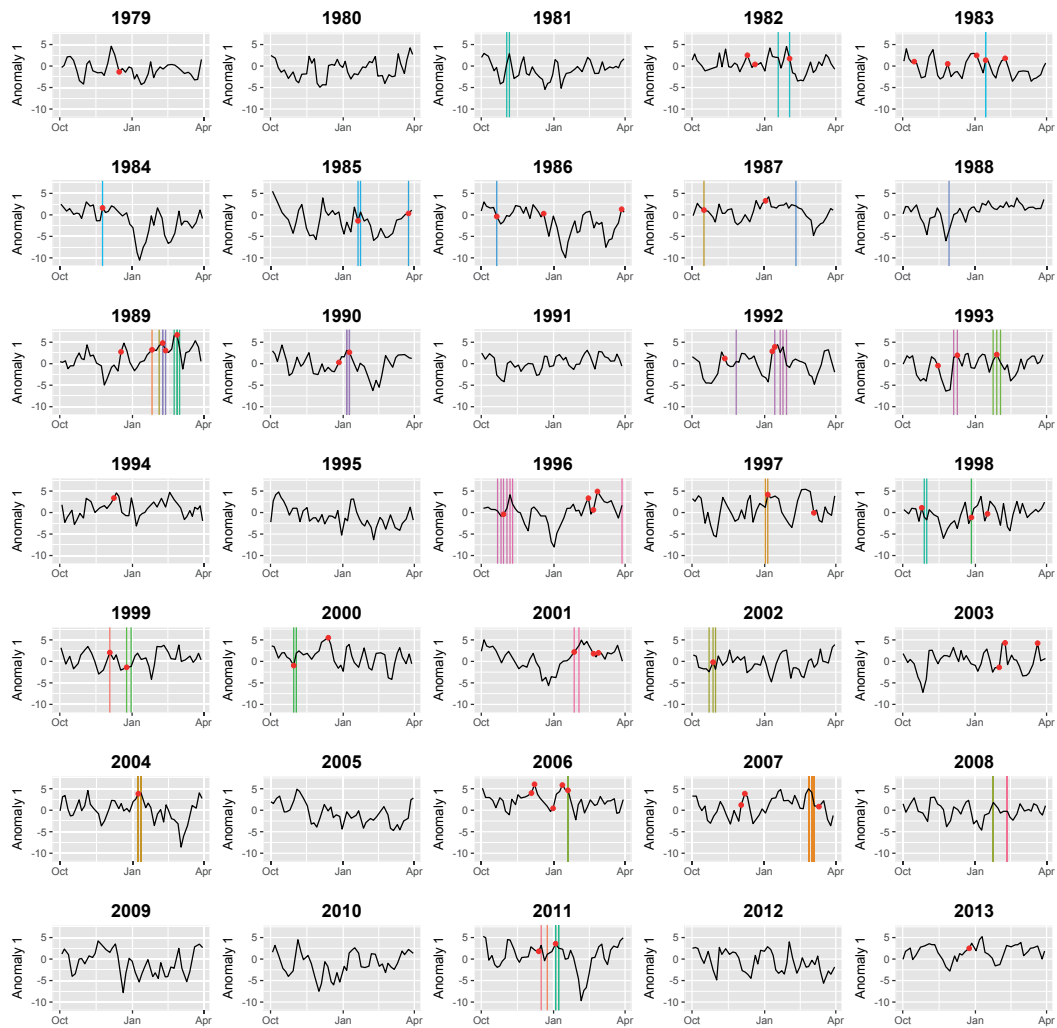


Figure E.2 – Three-hourly first eigenvalue of the temperature anomaly index computed on the ERA–Interim data set for each winter. r -exceedances above the 0.96 empirical quantile are represented by red dots and wind storms from XWS catalogue are represented by vertical lines coloured by dates.

Appendix E. Supplementary material for Chapter 6

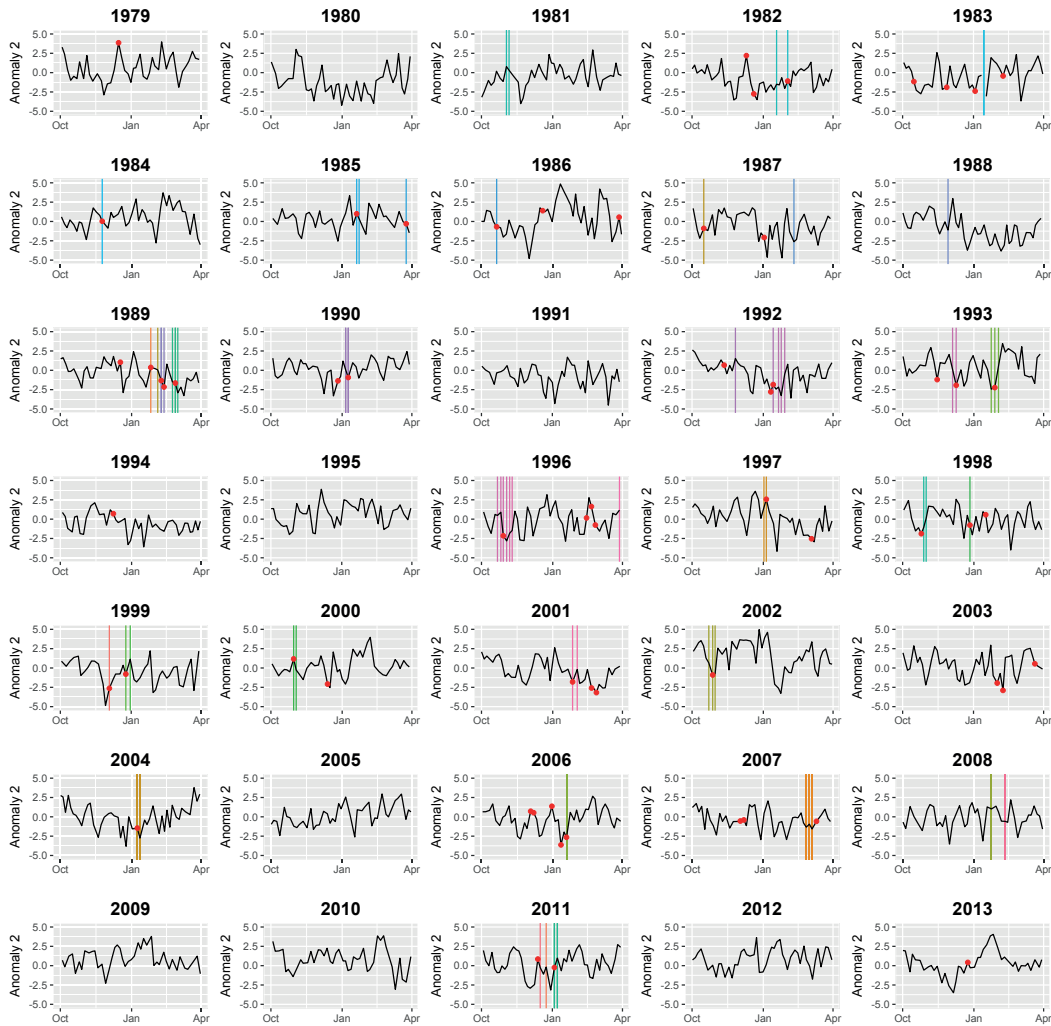


Figure E.3 – Three-hourly third eigenvalue of the temperature anomaly index computed on the ERA–Interim data set for each winter. r -exceedances above the 0.96 empirical quantile are represented by red dots and wind storms from XWS catalogue are represented by vertical lines coloured by dates.

E.1. Diagnostic plots for modelling the frequency of wind storms



Figure E.4 – Modelled three-hourly probability of r -exceedance for each winter from a logistic regression with the NAO index and the first and third temperature anomaly eigenvalues as covariates. Observed r -exceedances are represented by red points and the vertical lines coloured by dates correspond to the storms from the XWS catalogue.

Bibliography

- Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on River Networks. *Annals of Applied Statistics*, 9(4):2023–2050.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall/CRC Press, London.
- Bechler, A., Bel, L., and Vrac, M. (2015a). Conditional Simulations of the Extremal t process: Application to Fields of Extreme Precipitation. *Spatial Statistics*, 12(1):109–127.
- Bechler, A., Vrac, M., and Bel, L. (2015b). A Spatial Hybrid Approach for Downscaling of Extreme Precipitation Fields. *Journal of Geophysical Research: Atmospheres*, 120(10):4534–4550.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester.
- Berg, C., Christensen, J. P. R., and Ressel, P. (1984). *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. Springer, New York.
- Blessing, S., Fraedrich, K., Junge, M., Kunz, T., and Lunkeit, F. (2005). Daily North-Atlantic Oscillation (NAO) index: Statistics and its Stratospheric Polar Vortex Dependence. *Meteorologische Zeitschrift*, 14(6):763–769.
- Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. University of California Press, Berkeley.
- Brown, B. M. and Resnick, S. I. (1977). Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14(4):732–739.
- Buhl, S. and Kluppelberg, C. (2016). Anisotropic Brown–Resnick Space-time Processes: Estimation and Model Assessment. *Extremes*, to appear.

Bibliography

- Buhl, S. and Klüppelberg, C. (2018). Limit Theory for the Empirical Extremogram of Random Fields. *Stochastic Processes and their Applications*, 128(6):2060–2082.
- Buishand, T. A., de Haan, L., and Zhou, C. (2008). On Spatial Extremes: With Application to a Rainfall Problem. *Annals of Applied Statistics*, 2(2):624–642.
- Castruccio, S., Huser, R., and Genton, M. G. (2016). High-Order Composite Likelihood Inference for Max-Stable Distributions and Processes. *Journal of Computational and Graphical Statistics*, 25(4):1212–1229.
- Chandler, R. E. and Bate, S. (2007). Inference for Clustered Data Using the Independence Loglikelihood. *Biometrika*, 94(1):167–183.
- Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(1):207–222.
- Chavez-Demoulin, V. and Davison, A. C. (2012). Modelling Time Series Extremes. *REVSTAT*, 10(1):109–133.
- Chiles, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- Christakos, G., Hristopulos, D. T., and Bogaert, P. (2000). On the Physical Geometry Concept at the Basis of Space/time Geostatistical Hydrology. *Advances in Water Resources*, 23(8):799–810.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes*, 2(4):339–365.
- Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):377–392.
- Coles, S. G. and Tawn, J. A. (1994). Statistical Methods for Multivariate to Structural Design Extremes: an Application to Structural Design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):1–48.
- Coles, S. G. and Tawn, J. A. (1996). Modelling Extremes of the Areal Rainfall Process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(2):329–347.
- Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for Spatial Max-stable Random Fields. *Dependence in probability and statistics*.

- Cox, D. R. and Isham, V. (1988). A Simple Spatial-Temporal Model of Rainfall. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 415(1849):317–328.
- Cressie, N. and Huang, H. C. (1999). Classes of Nonseparable Spatio-Temporal Stationary Covariance Functions. *Journal of the American Statistical Association*, 94(448):1330–1340.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.
- Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken.
- Davis, R. A., Kluppelberg, C., and Steinkohl, C. (2013a). Max-stable Processes for Modelling Extremes Observed in Space and Time. *Journal of the Korean Statistical Society*, 42(3):399–414.
- Davis, R. A. and Mikosch, T. (2009). The Extremogram: a Correlogram for Extreme Events. *Bernoulli*, 15(4):977–1009.
- Davis, R. A., Mikosch, T., and Zhao, Y. (2013b). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications*, 123(7):2575–2602.
- Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of Extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2138):581–608.
- Davison, A. C. and Smith, R. L. (1990). Models for Exceedances over High Thresholds (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.
- Dawid, P. A., Musio, M., and Ventura, L. (2016). Minimum Scoring Rule Inference. *Scandinavian Journal of Statistics*, 43(1):123–138.
- de Fondeville, R. (2016). mvPot — R package version 0.1.1.
- de Fondeville, R. and Davison, A. C. (2018). High-dimensional Peaks-over-threshold Inference. *Biometrika*, 105(3):575–592.
- de Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194–1204.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York, USA.

Bibliography

- de Haan, L. and Lin, T. (2001). On Convergence Toward an Extreme Value Distribution in $C[0, 1]$. *The Annals of Probability*, 29(1):467–483.
- de Haan, L. and Resnick, S. I. (1993). Estimating the Limit Distribution of Multivariate Extremes. *Communications in Statistics. Stochastic Models*, 9(2):275–309.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.
- Della Marta, P. and Mathis, H. (2008). The return period of wind storms over Europe. *International Journal of Climatology*, 29(3):437–459.
- Della-Marta, P. M., Liniger, M. A., Appenzeller, C., Bresch, D. N., Köllner-Heck, P., and Muccione, V. (2010). Improved estimates of the European winter windstorm climate and the risk of reinsurance loss using climate model data. *Journal of Applied Meteorology and Climatology*, 49(10):2092–2120.
- Dick, J. and Pillichshammer, F. (2010). *Digital Nets and Sequences*. Cambridge University Press, Cambridge.
- Dieker, A. B. and Mikosch, T. (2015). Exact Simulation of Brown–Resnick Random Fields at a Finite Number of Locations. *Extremes*, 18(2):304–315.
- Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer, New York.
- Dombry, C., Engelke, S., and Oesting, M. (2016). Exact Simulation of Max-stable Processes. *Biometrika*, 103(2):303–317.
- Dombry, C., Éyi-Minko, E., and Ribatet, M. (2013). Conditional Simulation of Max-stable Processes. *Biometrika*, 100(1):111–124.
- Dombry, C. and Ribatet, M. (2015). Functional Regular Variations, Pareto Processes and Peaks Over Thresholds. *Statistics and Its Interface*, 8(1):9–17.
- Donat, M., Leckebusch, G., Pinto, J., and Ulbrich, U. (2010). European Storminess and Associated Circulation Weather Types: Future Changes Deduced from a Multi-model Ensemble of GCM Simulations. *Climate Research*, 42(1):27–43.

- Donat, M. G., Leckebusch, G. C., Wild, S., and Ulbrich, U. (2011). Future Changes in European Winter Storm Losses and Extreme Wind Speeds Inferred from GCM and RCM Multi-model Simulations. *Natural Hazards and Earth System Science*, 11(5):1351–1370.
- Economou, T. and David, B. (2014). Spatio-temporal Modelling of Extreme Storms. *Annals of Applied Statistics*, 8(4):2223–2246.
- Einmahl, J. H. J., Kiriliouk, A., Krajina, A., and Segers, J. (2016a). An M-estimator of Spatial Tail Dependence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 78(1):275–298.
- Einmahl, J. H. J., Kiriliouk, A., and Segers, J. (2016b). A Continuous Updating Weighted Least Squares Estimator of Tail Dependence in High Dimensions. *Journal of Statistical Planning and Inference*, 169(12):22–33.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.
- Engelke, S., de Fondeville, R., and Oesting, M. (2019). Extremal Behaviour of Aggregated Data with an Application to Downscaling. *Biometrika (to appear)*.
- Engelke, S., Kabluchko, Z., and Schlather, M. (2011). An equivalent Representation of the Brown-Resnick Process. *Statistics and Probability Letters*, 81(8):1150–1154.
- Engelke, S. and Malinowski, A. (2014). Statistical Inference for Max-stable Processes by Conditioning on Extreme Events. *Advances in Applied Probability*, 46(2):478–495.
- Engelke, S., Malinowski, A., Kabluchko, Z., and 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.
- Falk, M., Hüsler, J., and Reiss, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events*. Birkhäuser, Basel.
- Ferreira, A. and de Haan, L. (2014). The Generalized Pareto Process; with a View Towards Application and Simulation. *Bernoulli*, 20(4):1717–1737.
- Ferreira, A., de Haan, L., and Zhou, C. (2012). Exceedance Probability of the Integral of a Stochastic Process. *Journal of Multivariate Analysis*, 105(1):241–257.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Bibliography

- Fouedjio, F., Desassis, N., and Rivoirard, J. (2016). A generalized Convolution Model and Estimation for Non-stationary Random Functions. *Spatial Statistics*, 16(1):35–52.
- Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2013). Exploring a New Class of Non-stationary Spatial Gaussian Random Fields with Varying Local Anisotropy. (1):1–31.
- Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015a). Exploring a New Class of Non-stationary Spatial Gaussian Random Fields with Varying Local Anisotropy. *Statistica Sinica*, 25(1):115–133.
- Fuglstad, G. A., Simpson, D., Lindgren, F., and Rue, H. (2015b). Does Non-stationary Spatial Data Always Require Non-stationary Random Fields? *Spatial Statistics*, 14(1):505–531.
- Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010). *Handbook of Spatial Statistics*. Chapman & Hall/CRC Press, New York, USA.
- Genz, A. (2013). QSILATMVNV, Matlab program.
- Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer, Dordrecht.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2014). mvtnorm: Multivariate Normal and t-Distributions. R package version 1.0-2.
- Geyer, C. J. (1994). On the Convergence of Monte Carlo Maximum Likelihood Calculations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 56(1):261–274.
- Giné, E., Hahn, M. G., and 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 Série Aléatoire. *Annals of Mathematics*, 44(3):423–453.
- Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378.
- Gneiting, T. and Schlather, M. (2004). Separate Fractal Dimension and the Hurst Effect. *SIAM Journal on Numerical Analysis*, 46(2):269–282.
- Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

- Hall, T. M. and Jewson, S. (2008). Comparison of local and basinwide methods for risk assessment of tropical cyclone landfall. *Journal of Applied Meteorology and Climatology*, 47(2):361–367.
- Haylock, M. (2011). European Extra-tropical Storm Damage Risk from a Multi-model Ensemble of Dynamically-downscaled Global Climate Models. *Natural Hazards and Earth System Science*, 11(10):2847–2857.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European Daily High-resolution Gridded Data Set of Surface Temperature and Precipitation for 1950-2006. *Journal of Geophysical Research Atmospheres*, 113(20).
- Hill, B. M. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163–1174.
- Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and Quantile Estimation for Generalized Pareto Distribution. *Technometrics*, 29(3):339–349.
- Hsing, T., Hüsler, J., and Leadbetter, M. R. (1988). On the Exceedance Point Process for a Stationary Sequence. *Probability Theory and Related Fields*, 78(1):97–112.
- Huber, P. J. (1967). The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Berkeley Symposium on Mathematical Statistics and Probability*, 1(1):221–233.
- Hult, H. and Lindskog, F. (2005). Extremal Behavior of Regularly Varying Stochastic Processes. *Stochastic Processes and their Applications*, 115(2):249–274.
- Huser, R. and Davison, A. C. (2013). Composite Likelihood Estimation for the Brown-Resnick Process. *Biometrika*, 100(2):511–518.
- Huser, R. and Davison, A. C. (2014). Space-time Modelling of Extreme Events. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(2):439–461.
- Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood Estimators for Multivariate Extremes. *Extremes*, 19(1):79–103.
- Hüsler, J. and Reiss, R.-D. (1989). Maxima of Normal Random Vectors: Between Independence and Complete Dependence. *Statistics and Probability Letters*, 7(4):283–286.
- Hyvärinen, A. (2005). Estimation of Non-normalized Statistical Models by Score Matching. *Journal of Machine Learning Research*, 6(4):695–708.

Bibliography

- Hyvärinen, A. (2007). Some Extensions of Score Matching. *Computational Statistics & Data Analysis*, 51(5):2499–2512.
- Jenkinson, A. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements. *Quarterly Journal of Royal Meteorological Society*, 81(348):158–171.
- Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate Threshold Methods for Extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(1):171–183.
- Kabluchko, Z. (2011). Extremes of Independent Gaussian Processes. *Extremes*, 14(3):285–310.
- Kabluchko, Z. and Schlather, M. (2010). Ergodic Properties of Max-infinitely Divisible Processes. *Stochastic Processes and their Applications*, 120(3):281–295.
- Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary Max-stable Fields Associated to Negative Definite Functions. *Annals of Probability*, 37(5):2042–2065.
- Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A. (2011). Nonstationary Probabilistic Downscaling of Extreme Precipitation. *Journal of Geophysical Research: Atmospheres*, 116(5).
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of Extremes in Climatology. *Advances in Water Resources*, 25(8-12):1287–1304.
- Klüppelberg, C. and Resnick, S. I. (2008). The Pareto Copula, Aggregation of Risks, and the Emperor's Socks. *Journal of Applied Probability*, 45(1):67–84.
- Lajaunie, C. (1993). L'Estimation Géostatistique Non Linéaire. Technical report, C-152, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- Leadbetter, M. R. (1983). Extremes and Local Dependence in Stationary Sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*, 83(1):169–187.
- Lindgren, F., Rue, H., and Lindström, J. (2011). An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach (with discussions). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(4):423–498.

- Lindskog, F., Resnick, S. I., and Roy, J. (2014). Regularly Varying Measures on Metric Spaces: Hidden Regular Variation and Hidden Jumps. *Probability Surveys*, 11(1):270–314.
- Madsen, H., Rasmussen, P. F., and Rosbjerg, D. (1997). Comparison of Annual Maximum Series and Partial Duration Series Methods for Modeling Extreme Hydrologic Events. *Water Resources Research*, 33(4):747–757.
- Mainik, G. and Embrechts, P. (2013). Diversification in Heavy-tailed Portfolios: Properties and Pitfalls. *Annals of Actuarial Science*, 7(1):26–45.
- Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). Downscaling Extremes: A Comparison of Extreme Value Distributions in Point-source and Gridded Precipitation Data. *Annals of Applied Statistics*, 4(1):484–502.
- Matern, B. (1960). *Spatial Variation*. Springer, New York.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management : Concepts, Techniques and Tools*. Princeton University Press, Princeton.
- Mikosch, T. (2006). Copulas: Tales and Facts. *Extremes*, 9(1):3–20.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, USA.
- Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme Value Properties of Multivariate t Copulas. *Extremes*, 12(2):129–148.
- Northrop, P. J. and Coleman, C. L. (2014). Improved Threshold Diagnostic Plots for Extreme Value Analyses. *Extremes*, 17(2):289–303.
- Nuyens, D. and Cools, R. (2004). Fast Component-by-Component Construction, a Reprise for Different Kernels. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 373–387. Springer Berlin.
- Oesting, M., Bel, L., and Lantuéjoul, C. (2017a). Sampling from a Max-Stable Process Conditional on a Homogeneous Functional with an Application for Downscaling Climate Data. *Scandinavian Journal of Statistics*, 45(2):382–404.
- Oesting, M., Kabluchko, Z., and Schlather, M. (2011). Simulation of Brown–Resnick processes. *Extremes*, 15(1):89–107.

Bibliography

- Oesting, M. and Schlather, M. (2013). Conditional Sampling for Max-stable Processes with a Mixed Moving Maxima Representation. *Extremes*, 17(1):157–192.
- Oesting, M., Schlather, M., and Friederichs, P. (2017b). Statistical Post-Processing of Forecasts for Extremes Using Bivariate Brown-Resnick Processes with an Application to Wind Gusts. *Extremes*, 20(2):309–332.
- Opitz, T. (2013a). Extremal t -processes: Elliptical Domain of Attraction and a Spectral Representation. *Journal of Multivariate Analysis*, 122(1):409–413.
- Opitz, T. (2013b). *Extrêmes Multivariés et Spatiaux : Approches Spectrales et Modèles Elliptiques*. PhD thesis, Université Montpellier II.
- Pachauri, R. K., Meyer, L., Van Ypersele, J.-P., Brinkman, S., Van Kesteren, L., Leprince-Ringuet, N., and Van Boxmeer, F. (2014). *IPCC, 2014: Climate Change 2014: Synthesis Report*. Geneva, Switzerland.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Pebesma, E. and Graeler, B. (2013). gstat — R package version 1.1.6.
- Penrose, M. D. (1992). Semi-min-stable Processes. *The Annals of Probability*, 20(3):1450–1463.
- Pfahl, S. (2014). Characterising the Relationship Between Weather Extremes in Europe and Synoptic Circulation Features. *Natural Hazards and Earth System Science*, 14(6):1461–1475.
- Pinto, J., Karremann, M., Born, K., Della-Marta, P., and Klawa, M. (2012). Loss Potentials Associated with European Windstorms under Future Climate Conditions. *Climate Research*, 54(1):1–20.
- Powell, M. D. and Reinhold, T. A. (2007). Tropical Cyclone Destructive Potential by Integrated Kinetic Energy. *Bulletin of the American Meteorological Society*, 88(1):513–526.
- Renard, D., Bez, N., Desassis, N., Beucher, H., Ors, F., and Freulon, X. (2017). RGeostats: The Geostatistical R package. R package.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York, USA.

- Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- Ribatet, M. and Sedki, M. (2013). Extreme Value Copulas and Max-stable Processes. *Journal de la Société Française de Statistique*, 154(1):138–150.
- Roberts, J. F., Champion, A. J., Dawkins, L. C., Hodges, K. I., Shaffrey, L. C., Stephenson, D. B., Stringer, M. A., Thornton, H. E., and Youngman, B. D. (2014). The XWS Open Access Catalogue of Extreme European Windstorms from 1979 to 2012. *Natural Hazards and Earth System Sciences*, 14(9):2487–2501.
- Rootzén, H., Segers, J., and Wadsworth, J. L. (2018a). Multivariate Generalized Pareto Distributions: Parametrizations, Representations, and Properties. *Journal of Multivariate Analysis*, 165(1):117–131.
- Rootzén, H., Segers, J., and Wadsworth, J. L. (2018b). Multivariate Peaks-over-thresholds Models. *Extremes*, 21(1):115–145.
- Rootzén, H. and Tajvidi, N. (2006). Multivariate Generalized Pareto Distributions. *Bernoulli*, 12(5):917–930.
- Rue, H., Lindgren, F., Simpson, D., Riebler, A., Krainski, E. T., and Fuglstad, G.-A. (2017). R-INLA — R package version 18.07.12.
- Scarrott, C. and MacDonald, A. (2012). A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *Revstat – Statistical Journal*, 10(1):33–60.
- Schilling, R. L., Song, R., and Vondraček, Z. (2012). *Bernstein Functions: Theory and Applications*. De Gruyter, Berlin.
- Schlather, M. (2002). Models for Stationary Max-Stable Random Fields. *Extremes*, 5(1):33–44.
- Schlather, M. and Moreva, O. (2017). A Parametric Model Bridging Between Bounded and Unbounded Variograms. *Stat*, 6(1):47–52.
- Schlather, M. and Tawn, J. A. (2003). A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference. *Biometrika*, 90(1):139–156.
- Smith, R. L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science*, 4(4):367–377.
- Smith, R. L. (1990). Max-stable processes and spatial extremes. Technical report.

Bibliography

- Stephenson, A. G. and Tawn, J. A. (2005). Exploiting Occurrence Times in Likelihood Inference for Componentwise Maxima. *Biometrika*, 92(1):213–227.
- Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian Inference for the Brown–Resnick Process, with an Application to Extreme Low Temperatures. *Annals of Applied Statistics*, 10(4):2303–2324.
- Thibaud, E. and Opitz, T. (2015). Efficient Inference and Simulation for Elliptical Pareto Processes. *Biometrika*, 102(4):855–870.
- Thorarinsdottir, T. L., Gneiting, T., and Gissibl, N. (2013). Using Proper Divergence Functions to Evaluate Climate Models. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):522–534.
- Towe, R., Eastoe, E., Tawn, J. A., and Jonathan, P. (2017). Statistical Downscaling for Future Extreme Wave Heights in the North Sea. *The Annals of Applied Statistics*, 11(4):2375–2403.
- Varin, C., Reid, N., and Firth, D. (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21(1):5—42.
- von Mises, R. (1964). La Distribution de la Plus Grande de n Valeurs. In Ph. Frank, Goldstein, S., Kac, M., Prager, W., Szegő, G., and Birkhoff, G., editors, *Selected Papers of Richard von Mises: Volume II. Probability and Statistics, General*, volume 2, pages 271–294. American Mathematical Society, Providence.
- Wackernagel, H. (2003). *Multivariate Geostatistics*. Springer-Verlag, Berlin.
- Wadsworth, J. L. (2015). On the Occurrence Times of Componentwise Maxima and Bias in Likelihood Inference for Multivariate Max-stable Distributions. *Biometrika*, 102(3):705–711.
- Wadsworth, J. L. (2016). Exploiting Structure of Maximum Likelihood Estimators for Extreme Value Threshold Selection. *Technometrics*, 58(1):116–126.
- Wadsworth, J. L. and Tawn, J. A. (2014). Efficient Inference for Spatial Extreme Value Processes Associated to Log-Gaussian Random Functions. *Biometrika*, 101(1):1–15.
- Wan, P. and Davis, R. A. (2018). Threshold Selection for Multivariate Heavy-Tailed Data. *Extremes*, To appear.
- Weller, G. B., Cooley, D., Sain, S. R., Bukovsky, M. S., and Mearns, L. O. (2013). Two Case Studies on NARCCAP Precipitation Extremes. *Journal of Geophysical Research Atmospheres*, 118(18):10475–10489.

- Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika*, 41(3):434–449.
- Whittle, P. (1963). Stochastic Processes in Several Dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994.
- Yiou, P. (2014). AnaWEGE: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543.
- Zhou, C. (2010). Dependence Structure of Risk Factors and Diversification Effects. *Insurance: Mathematics and Economics*, 46(3):531–540.

Raphaël de Fondeville

Avenue Louis-Ruchonnet, 35
Lausanne, 1003, Switzerland
☎ +41 78 823 8933
✉ raphael.defondeville@gmail.com

Research Interests

Extreme Value Theory, Spatio-Temporal Statistics, Dependence Modeling for Stochastic Processes, Computational Statistics, Environmental Statistics, Proper Scoring Rules.

Education

- Ph.D. **Statistics**, *École Polytechnique Fédérale de Lausanne (EPFL)*, Suisse.
2014 - 2018 Topic : Statistical modeling of complex extreme events.
Advisor : Anthony Davison.
- M.Sc. **Science and Executive Engineering**, *Mines ParisTech*, France.
2010 - 2014 Topic : Scoring and multivariate extremes, assessing climate forecasts of extremes.
Awarded with highest distinction.
- Aug 2011 - Jan 2012 **Exchange student**, *École Polytechnique de Montréal*, Canada.
- Preparatory School **Mathematics and Physics**, *Lycée Sainte Geneviève*, France.
2008 - 2010 Awarded with highest distinction.

Publications

Papers

- In press **de Fondeville, R.** and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.
- Accepted **de Fondeville, R.**, Engleke, S. and Oesting, M. (2018). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, to appear.
- In preparation **de Fondeville, R.** and Davison, A.C. (2018). Functional peaks-over-threshold analysis and generalized r-Pareto process.

Reports

- Master thesis **de Fondeville, R.** (2014), Scoring and multivariate extremes : Assessing climate forecast of extremes.

Teaching

- 2014 - **Teaching Assistant**, *École Polytechnique Fédérale de Lausanne*.
Courses assignments include Analysis, Linear Algebra, Probabilities and Statistics, Modern Regression at both Bachelor and Master levels.
- 2007 - 2014 **Private teacher**.
Tutor in Mathematics, Physics and Chemistry for high school students.

Conferences

- Novembre 2018 **Poster**, *Swiss Statistics Seminars*, Bern, Switzerland.
- Novembre 2018 **Invited seminar**, *Colloquium in Climatology, Climate Impact and Remote Sensing*, Bern, Switzerland.
- September 2018 **Poster**, *Data, Models and Statistical Inference*, Lausanne, Switzerland.
- August 2018 **Invited speaker**, *Journées MAS*, Dijon, France.
- June 2018 **Poster**, *Data Science summer school (DS³)*, Palaiseau, France.
- December 2017 **Invited speaker**, *10th ERCIM Conference*, London, England.
- December 2017 **Invited seminar**, *BIRS workshop on Inferential Challenges for Large Spatio-Temporal Data Structures*, Banff, Canada.
- September 2017 **Invited seminar**, *Centre de géosciences, Mines Paristech*, Fontainebleau, France.
- June 2017 **Invited speaker**, *10th Extreme Value Analysis Conference*, Delft, Netherlands.
- February 2017 **Contributed speaker**, *Workshop on Risk Quantification and Extreme Values in Applications*, Lausanne, Switzerland.
- June 2016 **Contributed speaker**, *13th International Meeting on Statistical Climatology*, Canmore, Canada.
- January 2016 **Contributed speaker**, *EPFL statistics workshop*, Davos, Switzerland.
- June 2016 **Contributed speaker**, *Young Researcher's Conference in Applied Probability and Statistics (CUSO)*, Lausanne, Switzerland.
- June 2015 **Poster**, *9th international conference on Extreme Value Analysis*, Ann Arbor, USA.
- January 2015 **Contributed speaker**, *EPFL statistics workshop*, Davos, Switzerland.

Professional Experience

- Jan - Jun 2014 **CNRS**, Gif-sur-Yvette, France.
Colorado State University, Fort Collins, USA.
Scoring and multivariate extremes : assessing climate forecasts of extremes.
- Dec 2013 - Feb 2014 **AXA**, Paris, France.
Data accessibility in climatology.
- Jan - Jul 2013 **TOTAL E&P, Geoscience Research Center**, Aberdeen, United Kingdom.
Stochastic inversion in geophysics for uncertainty quantification and reservoir modeling.
- Dec 2012 **'My Studies'**, Paris, France.
Chemistry textbook of exercises for students preparing the French baccalauréat.
- Jun - Nov 2012 **Ernst & Young GmbH, Advisory Services**, Frankfurt am Main, Germany.
Innovation management, business model evaluation and risk analysis.

Software

de Fondeville, R. (2016). mvPot : Multivariate Peaks-over-Threshold modeling for Spatial Extreme Events. *R package available on the Comprehensive R Archive Network*.

Volunteering and Services

- Nov 2017 **My Thesis in 180 seconds (MT180)**, Presentation of the Ph.D. research topic in three minutes to a general public..
Finalist of the EPFL competition.
- Feb 2017 **Workshop on Risk Quantification and Extreme Values in Applications.**
In charge of the poster session.
- Jun 2016 **Conférence universitaire de Suisse occidentale**, Doctoral association in statistics for the french speaking universities in Switzerland.
Organisation of the Young Researcher's Conference in Applied Probability and Statistics.
- 2011 - 2012 **Member of the Students' Union**, *Association in charge of the student life at MINES ParisTech.*
Event manager of monthly events for more than 200 students.
- 2010 - 2011 **Association Forum Trium**, *Association in charge of a forum gathering no fewer than 150 companies and more than 3500 students.*
In charge of enterprises solicitation and 'in school' advertising.

Programming Skills

R, Java, Python, C++, Bash, L^AT_EX, Office. Notions of CUDA.

Languages

- French Native speaker.
English Fluent.
German Professional proficiency (speaking, reading); advanced (writing).

Personal Interests

I practice road and mountain running competitively at distances ranging from 10 to 40km. I like hiking, snowboarding and other outdoor activities. I practiced ice-skating in team and solo for 14 years.