Statistical modelling and inference for covariate-dependent extremal dependence

Mhalla, Linda

Statistical Modelling and Inference for Covariate-dependent Extremal Dependence

by

Linda Mhalla

A thesis submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,
in fulfillment of the requirements for the degree of
PhD in Statistics

Members of the thesis committee:
Prof. Sebastian Engelke, Chair, University of Geneva
Prof. Valérie Chavez-Demoulin, Co-adviser, University of Lausanne
Prof. Elvezio Ronchetti, Co-adviser, University of Geneva
Prof. Jonathan Tawn, University of Lancaster

Thesis No. 57
June 2018

Acknowledgements

I have been blessed to be surrounded by people without whom none of the work presented in this thesis would have been possible, and I would like to express my deep gratitude to them.

First and foremost, I would like to thank my supervisors Prof. Valérie Chavez-Demoulin and Prof. Elvezio Ronchetti. I am thankful to Prof. Elvezio Ronchetti for giving me the opportunity to do this PhD, for allowing me to choose the research topics on which this thesis is built, and for his support whenever I asked to participate to a conference. I am extremely (!) lucky to have had Prof. Valérie Chavez-Demoulin as a supervisor. I have found a friend in her, a person that not only have supported me unconditionally from the beginning, but that also shared my passion for pastries, chocolate, and raclette. Valérie has always been available when I needed her and I have learnt a lot from her, both at the academic level and at the personal level. I thank her very much for everything she has done for me.

I further thank Sebastian Engelke and Jonathan Tawn who honoured me by accepting to be part of my thesis committee and whose instructive comments improved the contents of this thesis.

My sincere thanks also go to Miguel de Carvalho, Philippe Naveau, and Thomas Opitz with whom I enjoyed working and who kindly invited me to very interesting seminars, workshops, and conferences. A special thanks to Anthony Davison who introduced me to the fascinating world of extremes, and whose passion for research inspired me and got me interested in pursuing this PhD.

Many thanks to all my colleagues at the Geneva School of Economics and Management. Arguably, I wasn't always in my office but I enjoyed the happy moments we shared during the CUSO schools, the seminars, the informal discussions, and the coffee breaks. I am also grateful to my "non-official" colleagues at EPFL, where I spent most of my time, for considering me as part of their lab and for allowing me to use their super computers to perform intensive simulations.

Finally, I would like to thank my family, those who are with us and those who life took away from us, for their never-ending confidence and support, as well as my family-in-law for their constant encouragement. Most of all, I would like to thank my husband Paul, or as I prefer calling him Poufpouf, without whom none of this would have been possible. You have been my source of motivation since day one, through your constant love, patience, joy, and laughter. These past months spent in the office with you have been some of the best days in my life, and I look forward to our next adventure, wherever life may take us.

Abstract

Over the past four decades, we have witnessed an increasing interest in developing and using multivariate extreme value theory to deal with the growing concern for risk assessment of rare phenomena, due to their large socio-economic impacts. The extreme value community have realized the prominence of considering the multivariate nature of extreme events to take into account the risk component emerging from the dependence between these events, but have also recognized the additional complexity of this task compared to the well-understood univariate extreme value theory. Moreover, the non-stationary extent of most extreme events have focused the attention of many researchers over the past few years, owing to the auxiliary information that a set of covariates might contain, thereby allowing data pooling, improved inference, and a better understanding of the behaviour of extreme events in different environments and settings.

In this thesis, we develop new models for covariate-varying tail dependence structures based on asymptotically justified arguments, and propose novel techniques for fitting these models to both block maxima and threshold exceedances data, under the assumptions of asymptotic dependence and asymptotic independence. Our proposals for the flexible incorporation of covariate influence on the extremal dependence rely on the (vector) generalized additive modelling infrastructure, and are established in a parametric setting where we extend the standard approach of modelling non-stationary univariate extremes to the multivariate framework through spectral density modelling, as well as a non-parametric setting where we develop projection techniques enabling the reduction of the problem of characterizing joint tail dependences to the modelling of univariate random variables. Inference is performed by penalized maximum likelihood estimation combined, when applicable, with censored likelihood techniques. The performance of the resulting estimators is assessed either through simulation studies or based on asymptotic distributions, when the parametric approach to the extremal dependence modelling is undertaken.

The developed methodologies are illustrated on environmental datasets where dependence between large events is linked to a set of covariates describing time as well as characteristics of the measurement sites.

Keywords: Angular density; Asymptotic dependence; Asymptotic independence; Covariate-adjustment; Extreme value theory; Generalized additive models; Max-stable random vectors; Multivariate extreme values; Penalized log-likelihood; Threshold exceedances; Vector generalized additive models.

Résumé

Durant les quatre dernières décennies, on a observé un intérêt croissant pour l'utilisation et le développement de la théorie des valeurs extrêmes multivariées dans le but de répondre à la préoccupation, de plus en plus importante, pour l'évaluation des risques liés aux phénomènes rares, du fait de leurs impacts socio-économiques considérables. La communauté des valeurs extrêmes a répondu à la nécessité de considérer la nature multivariée des événements extrêmes afin de prendre en compte la composante du risque émanant de la dépendance entre ces événements, mais a aussi reconnu la complexité additionnelle de cette tâche, comparée à la théorie des valeurs extrêmes univariées qui est bien connue. En outre, au cours des dernières années, le caractère non stationnaire de la plupart des événements extrêmes a attiré l'attention de scientifiques portant un intérêt à la pratique, dû à l'information auxiliaire qu'un ensemble de variables explicatives peut contenir ainsi qu'à l'utilisation profitable de cette information, en vue de regrouper les données, améliorer l'inférence, et mieux comprendre le comportement des événements extrêmes dans différents environnements.

Dans cette thèse, nous développons de nouveaux modèles, justifiés par des arguments asymptotiques, pour des structures de dépendance extrême qui varient en fonction de variables explicatives, et nous proposons des techniques novatrices pour ajuster ces modèles à des maxima de blocs ainsi qu'à des excès de seuils élevés, et ceci dans le cas de dépendance asymptotique comme dans le cas d'indépendance asymptotique. Nos propositions d'incorporer, d'une manière flexible, l'influence de variables explicatives sur la dépendance extrémale reposent sur l'infrastructure des modèles (vecteurs) additifs généralisés. Ces propositions sont établies dans un cadre paramétrique à travers une extension de l'approche standard pour la modélisation d'extrêmes univariés non stationnaires, qui repose sur la modélisation de densités spectrales, ainsi que dans un cadre non paramètrique via le développement de techniques de projection qui permettent de réduire le problème de caractérisation de la dépendance dans les queues jointes à un problème de modélisation d'une variable aléatoire univariée. L'inférence s'appuie sur l'estimation par maximum de vraisemblance pénalisée combinée, dans le cas échéant, à des techniques de vraisemblance censurée. La performance des estimateurs proposés est établie soit au travers d'études de données simulées, soit en se basant sur leur distribution asymptotique, ceci étant sous l'approche paramétrique pour la modélisation de la dépendance extrémale.

Les méthodologies développées sont illustrées au travers d'applications à des données environnementales où la dépendance entre les événements extrêmes est décrite par un ensemble de variables explicatives comprenant le temps ainsi que des caractéristiques des stations de mesure.

Mots clés: Ajustement par variables explicatives; Densité angulaire; Dépendance asymptotique; Excès de seuil; Indépendance asymptotique; Log-vraisemblance pénalisée; Modèles additifs généralisés; Modèles vecteurs additifs généralisés; Théorie des valeurs extrêmes; Valeurs extrêmes multivariées; Vecteurs aléatoires max-stable.

Contents

Introduction

Motivation

In recent years, there has been a great deal of interest in multivariate extreme value modelling, driven by impetus for dependence structure modelling. This is especially due to the inherent multivariate nature of most of the phenomena resulting in devastating losses. Two major areas that have witnessed an upsurge of research on extreme value modelling and risk assessment are the environmental sciences (climatology, hydrology, oceanography) and finance. Examples of applications include the analysis of flood events resulting from extreme discharges at different upstream flows and the modelling of extreme price fluctuations of several stocks that can dramatically affect an institution's asset. In these settings, the motivation for extreme value modelling is the extrapolation of probabilities of events beyond the range of observed data. Without certain knowledge of the tail properties underlying the data, such extrapolations rely on asymptotic arguments where a strong belief in the adequacy of the asymptotic models outside the range of the observed data is required.

Fundamentally, the analysis of multivariate variables entails the analysis of two components: the marginal distributions and the dependence structure. While the first component, i.e., the margins, is dealt with using the univariate extreme value theory that goes back to the late 1920s, the treatment of the second component is based on the characterization of multivariate extreme value distributions that goes back to the late 1970s. Since this time, various efforts have been taken to model, in a flexible and parsimonious way, the dependence structure that cannot be represented by a finite parametric family. While some authors focus on parametric modelling of the limiting dependence structure, others deal with issues related to the validity of the asymptotic models and propose solutions to the modelling of the dependence either at penultimate levels or by considering different kinds of sparsity in its structure.

Although there is a plethora of approaches to extremal dependence modelling under numerous structural and limiting assumptions, little work is dedicated to incorporating covariate information to capture features of the tail behaviour. In the flood example mentioned above, the dependence between extreme river discharges might, and probably does, depend on the Euclidean or river distance between the locations of the flows. Although spatio-temporal max-stable models offer an elegant way to incorporate space and time effects on the extremal dependence, these models rely on isotropy and stationarity assumptions and more essentially, require a notion of distance between the observed components, which might be intricate in some cases, e.g,. the above-mentioned financial example. Furthermore, in many applications, covariates other than time are believed to affect the strength of the extremal dependence: the type of stock sectors, the business line in which losses occurred, the altitude or the type of station at which a process was observed, etc. We illustrate below the effect of the type of sector and time on the extremal dependence

between pairs of stocks. Figure 1 shows joint occurrences of extreme losses for two major players in the banking industry—Citibank and JP Morgan—during two different time periods as well as the joint extreme losses for two key players in the healthcare industry—Merk & Co and Patterson Companies. Care is needed when interpreting these plots as no

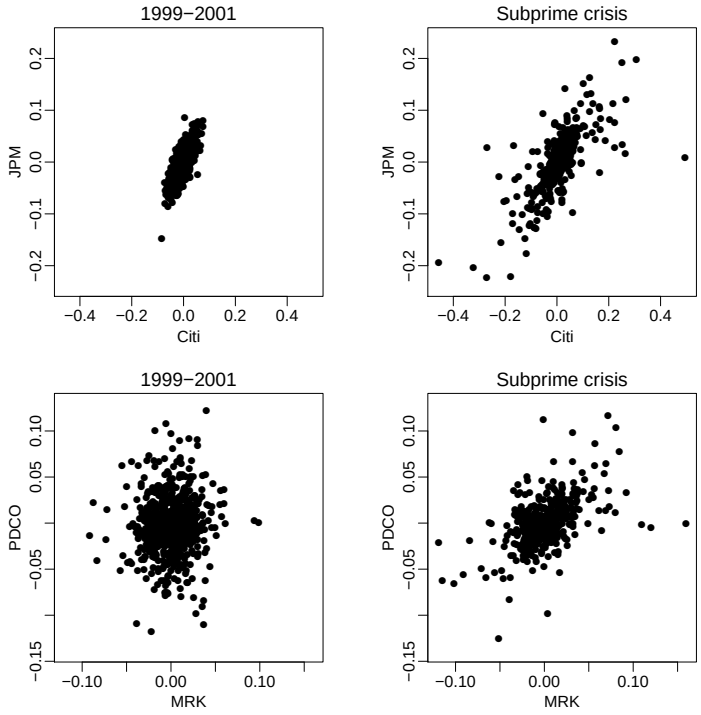

Figure 1: Scatterplots of negative daily log-returns stock prices of Citibank (Citi), JP Morgan (JPM), Merck & Co (MRK), and Patterson Companies (PDCO). Left: Time period of 1999–2001; Right: Time period corresponding to the subprime crisis (from December 1st, 2007 to June 30th, 2009).

pre-processing was performed to remove marginal non-stationarity and heteroskedasticity that may affect the tail dependence; see for instance Poon et al. [2003]. Nevertheless, during the same time period, 1999-2001 say, the strength of the tail dependence between pairs of stocks seems to differ between the two types of sectors with stronger dependence in the financial industry compared to the healthcare industry. Additionally, there is a noticeable change in the extremal dependence structure of the healthcare stocks, during the subprime crisis where data seem to exhibit more dependence in the tails.

This thesis contributes to the extreme value theory and the generalized additive modelling literature by developing new models for covariate-varying extremal dependence structures, as well as novel inference methods for such structures. The following paragraph delineates the content of the chapters and their contributions.

Outline and contributions of the thesis

In Chapter 1, classical extreme value theory is surveyed in the finite-dimensional case and an overview of generalized additive models is given. We tie together these two fields, in order to extend the available literature to the multivariate setting. This chapter serves as a background to the work developed in this thesis.

In Chapter 2, we discuss modelling of the extremal dependence in a parametric and covariate-varying framework. The main new contributions are to develop families of models for non-stationary extremal dependence structures based on a parametric assumption and the vector generalized additive modelling framework, and to devise a penalized maximum log-likelihood estimator for covariate-dependent angular densities. Consistency and asymptotic normality of this estimator are derived under regularity conditions. We illustrate the methodology with an application to winter extreme air temperatures in Switzerland, where the extremal dependence is linked to time and the North Atlantic Oscillation index. This chapter corresponds to the arXiv manuscript Mhalla et al. [2017b], that is under review for the Journal of the Royal Statistical Society, Series B.
In Chapter 3, we propose a semi-parametric method for the estimation of covariate-dependent multivariate Pickands' dependence functions, based on a novel projection technique of max-stable random vectors. The effect of the covariates is accounted for through the use of generalized additive models and the resulting Pickands' dependence function is regularized, in the bivariate case, using constrained median smoothing B-splines. An application of this methodology to temperature data in the U.S. is conducted and the effects of time and altitude on the extremal dependence are assessed. This chapter corresponds to the work published in Mhalla et al. [2017a].
In Chapter 4, we construct a threshold-based estimator of tail dependence measures, in a covariate-dependent framework. We extend the notion of projection of random vectors, developed in Chapter 3, to allow for the estimation of covariate-varying extremal dependence under asymptotic dependence and asymptotic independence settings. We also detail how inference based on the full likelihood in the asymptotic independence case, and the censored likelihood in the asymptotic dependence case, can be performed. The proposed techniques are illustrated on a large dataset of French air pollution where the influence of spatial distance, time, and type of station is determined. This chapter corresponds to the arXiv manuscript Mhalla et al. [2018], that is under review for Extremes.

Chapter 1

Background and literature review

This chapter is an overview of the cornerstones of this thesis: extreme value theory in the finite-dimensional case and generalized additive models. This chapter is intended to provide the necessary background for the rest of the thesis and to give an overview of the existing methods for modelling non-stationary extremes.

1.1 Extreme value theory

1.1.1 Univariate extreme value theory

1.1.1.1 Asymptotic distribution of normalized maxima

Let $(Y_i)_{i\geq 1}$ be a sequence of independent and identically distributed (IID) random variables with common distribution F. We are interested in the behaviour of the maximum of a sequence of n such random variables which we denote by M_n and define as $M_n = \max\{Y_1, \ldots, Y_n\}$. Based on the IID hypothesis, we can derive the cumulative distribution function of M_n:

$$\Pr(M_n \leq y) = \Pr(Y_1 \leq y, \ldots, Y_n \leq y) = \prod_{i=1}^{n} \Pr(Y_i \leq y) = \{F(y)\}^n,$$

and show that it converges, as $n \to \infty$, to a degenerate distribution function with a point mass on the upper end-point of the distribution F. By analogy with the central limit theorem, we can look for a suitable normalization of M_n that would lead to a non-degenerate limiting distribution function. The following theorem, derived heuristically by Fisher and Tippett [1928] and formalized by Gnedenko [1943], is the foundation of extreme value theory (EVT).

Theorem 1.1.1. *If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that the normalized random variable M_n converges in distribution to a random variable with a non–degenerate distribution function G, i.e.,*

$$\Pr\{(M_n - b_n)/a_n \leq y\} \to G(y), \quad n \to \infty,$$

then G belongs to the Generalized Extreme Value (GEV) family of distributions

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

defined on $\{y : 1+\xi(y-\mu)/\sigma > 0\}$, with $-\infty < \mu, \xi < \infty$, $\sigma > 0$, and $x_+ = \max(x, 0)$.

The parameters μ, σ, and ξ correspond to the location, scale, and shape parameters, respectively. Depending on the value of the shape parameter, we distinguish three limiting distributions: Fréchet distribution ($\xi > 0$) with a lower bounded support, reversed Weibull distribution ($\xi < 0$) with an upper bounded support, and Gumbel distribution ($\xi = 0$) with an unbounded support and an exponential decay in the upper tail. The characterization of the GEV distribution is convenient for inference as no assumption on the shape parameter should be made beforehand. We can therefore define the notion of the maximum domain of attraction of the extreme value distribution G as the class of underlying distributions F whose suitably normalized maxima converge in distribution to G. Examples of such distributions include the Beta distribution that belongs to the max-domain of attraction of the reversed Weibull, the Gamma distribution that belongs to the max-domain of attraction of the Gumbel, and the Pareto distribution that belongs to the max-domain of attraction of the Fréchet. Necessary and sufficient conditions for the existence of the max-domain of attraction along with a classification of numerous well-known distributions according to their max-domain of attraction can be found in Embrechts et al. [1997], Beirlant et al. [2004], de Haan and Ferreira [2006]. For instance, the following Gnedenko–de Haan theorem [Gnedenko, 1943, de Haan, 1970] provides necessary and sufficient conditions for a distribution F to belong to the max-domain of attraction of a GEV distribution $G_{(0,1,\xi)}$, denoted $F \in \mathrm{MDA}(G_\xi)$.

Theorem 1.1.2. *Let F be a distribution function with upper end-point y_F. Then*

- *$F \in MDA(G_\xi)$ with $\xi > 0$ if and only if $y_F = \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F(ty)}{1 - F(t)} = y^{-1/\xi}, \quad y > 0.$$

- *$F \in MDA(G_\xi)$ with $\xi < 0$ if and only if $y_f < \infty$ and*

$$\lim_{t\to\infty} \frac{1 - F\{y_F - (ty)^{-1}\}}{1 - F(y_F - t^{-1})} = y^{1/\xi}, \quad y > 0.$$

- *$F \in MDA(G_0)$ if and only if $\int_{t_0}^{y_f} \{1 - F(y)\}dy < \infty$ for some $t_0 < y_F$ and*

$$\lim_{y\to y_F} \frac{1 - F\{t + yR(t)\}}{1 - F(t)} = \exp(-y), \quad y \in \mathbb{R},$$

where $R(t) := \int_t^{y_F} \{1 - F(y)\}dy / \{1 - F(t)\}$, $t < y_F$.

Needless to say that the existence of the normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ in Theorem 1.1.1 is not guaranteed and one can find distributions that do not belong to any max-domain of attraction. Examples include the Poisson and the geometric distributions [Embrechts et al., 1997, p.118].

The assumption of the independence and identical distribution of the sequence of random variables $(Y_i)_{i\geq 1}$ is very likely to be violated in practice: the data may exhibit temporal dependence as it is the case for daily wind speeds or hourly rainfall, for example. Moreover, the limiting behaviour of the maxima may change over time as a consequence of a seasonal pattern or financial crises, for example. An extensive work on the theory of extreme values for strictly stationary time series with short-range dependence has been developed and summarized in Leadbetter et al. [1983]. The issue of non-stationarity is often handled by modelling the parameters of the GEV distribution using parametric and

semi-parametric models (see Section 1.3 for details) or by filtering the data which should, at least in theory, result in a stationary series of residuals on which "standard" EVT can be applied [McNeil and Frey, 2000, Eastoe and Tawn, 2009].

The extremal types theorem (1.1.1) justifies the use of the limiting GEV distribution as an approximate distribution for block maxima. The distribution $G_{(\mu,\sigma,\xi)}$ is typically fitted at sub-asymptotic levels, i.e, using finite-size blocks, based on likelihood methods. The log-likelihood is easily obtained and can be maximized numerically using standard optimization techniques. Care is needed when it comes to the classical asymptotic properties of the maximum likelihood estimates (MLE) as the end–points of the GEV are functions of the parameter values. Smith [1985] discussed the regularity conditions for likelihood estimation in EVT and showed that despite the standard regularity conditions being violated, classical asymptotic properties of the MLE hold when $\xi > -1/2$. This is the case for most of the environmental and financial applications. Alternatives to maximum likelihood estimators were proposed: Hosking et al. [1985], Hosking and Wallis [1987] advocated the use of probability weighted moments estimators for their good properties in small samples, although these estimators are defined only when $\xi < 1$ (finite mean cases), Dupuis and Field [1998] proposed the use of optimal bias-robust estimators for their bounded influence functions which limit the influence of outliers, and a Bayesian approach based on hierarchical models was adopted by Sang and Gelfand [2009, 2010]. Finally, it should be mentioned that, even though the choice of the block length results in a bias-variance trade-off, it is often chosen, in practice, according to the nature of the physical phenomenon underlying the data, e.g., seasonal or annual maxima in environmental applications [Coles, 2001, p.54].

1.1.1.2 Threshold-based models

The extreme events being scarce by definition, the block maxima approach may be qualified as wasteful since it discards extreme events that are not as extreme as the block maximum but that should be informative about the behaviour in the tails. In what follows, we modify the definition of rare events from the maximum over a block of observations to the observations exceeding a high fixed threshold, and focus on the asymptotic distribution of threshold exceedances. Mainly, two approaches have been considered: The Peaks Over Threshold (POT) method due to Balkema and de Haan [1974] and developed by Davison and Smith [1990], and the point process method studied by Pickands [1971], Smith [1989].

Let $(Y_i)_{i\geq 1}$ be a sequence of IID copies of a random variable Y with distribution function F, and suppose that there exist normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ such that Theorem 1.1.1 holds, i.e., F is in the max-domain of attraction of a $\text{GEV}_{(\mu,\sigma,\xi)}$. In the POT approach, we consider a finite high threshold u and model the limiting distribution of the exceedances $Y - u|Y > u$ with a Generalized Pareto Distribution (GPD) $G_{(\tilde{\sigma}_u,\xi)}$, as follows:

$$\Pr\left\{(Y_n - b_n)/a_n > u + y | (Y_n - b_n)/a_n > u\right\} \to \begin{cases} (1 + \xi y/\tilde{\sigma}_u)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-y/\tilde{\sigma}_u), & \xi = 0, \end{cases}$$

where $\tilde{\sigma}_u = \sigma + \xi(u - \mu)$ and the shape parameter ξ is the same as that of the corresponding GEV distribution. The distribution function $G_{(\tilde{\sigma}_u,\xi)}$ is defined on $\{y : y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}_u) > 0\}$, and the case of $\xi = 0$ is interpreted as the limit. Then, $\Pr(Y - u \leq y) \approx 1 - \zeta_u(1 + \xi y/\tilde{\sigma}_u)$ for exceedances of u, with $\zeta_u = \Pr(Y > u)$ and where we suppose that

the normalizing constants are absorbed by the location and scale parameters of the GEV distribution. Inference is performed using likelihood methods where we make use of the assumption of independence of the threshold exceedances. Alternative techniques such as the method of probability weighted moments or the elemental percentile method can be used to derive the parameters estimates; see e.g. Section 5.3 in Beirlant et al. [2004].
The second approach, based on two-dimensional point processes, formulates in an elegant way the limiting behaviour of observations exceeding a high threshold, while avoiding a characterization depending on the fixed threshold, as it is the case with the GPD modelling. We again suppose that Theorem 1.1.1 holds for the normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$, and define the set of points $\mathcal{P}_n$ as

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

The sequence of points $\mathcal{P}_n$ defines a point process on $(0,1) \times \mathbb{R}$. Then, provided F is in the max-domain of attraction of a $\text{GEV}_{(\mu,\sigma,\xi)}$, the sequence of point processes $\mathcal{P}_n$ converges on sets of the form $(0,1) \times [u, +\infty)$, $u \in \mathbb{R}$ large enough, to a Poisson process with intensity measure

$$\Lambda([t_1, t_2] \times [y, \infty)) = (t_2 - t_1) \left(1 + \xi \frac{y - \mu}{\sigma} \right)_+^{-1/\xi}, \quad y > u.$$

This result states that, for large n, non-extreme observations are absorbed towards the lower end-point of the $\text{GEV}_{(\mu,\sigma,\xi)}$ while observations belonging to the extremal set $[u, \infty)$ are approximated by a Poisson process whose intensity measure depends on the parameters of the underlying GEV. In addition to its nice mathematical derivation (see Embrechts et al. [1997] for details), the point process approach allows us to perform a likelihood-based inference where the parameters do not depend on the fixed threshold, as opposed to the POT approach.

As with the block maxima approach, the point process approach can be extended to stationary time series under a mixing condition imposing long-range independence [Hsing et al., 1988]. Short-range dependence being allowed, exceedances at high levels tend to occur in clusters and the limiting Poisson process for the exceedances is then modified to take into account the distribution of the cluster size; see [Beirlant et al., 2004, Section 10.3] for details. Another limiting result obtained under stationarity, concerns the limiting distribution of the cluster maxima that are shown to be asymptotically distributed as a Poisson process with intensity measure depending on the mean cluster size. This result is often used in practice to perform inference in presence of temporal dependence and where declustering methods, i.e., identification of the clusters in the data, are used [Davison and Smith, 1990, Smith and Weissman, 1994, Ferro and Segers, 2003]. The temporal dependence can also be accounted for by explicit formulation based on Markov chains [Fawcett and Walshaw, 2006].

An important caveat should be attached to the use of threshold methods and is related to the choice of the threshold. The choice of the threshold translates to a bias-variance trade-off as a low threshold, resulting in a higher number of exceedances, will minimize the variance but increases the bias as the asymptotic approximation for the tail of the distribution will be poor. Several graphical tools have been developed and rely on the threshold stability property of the GPD, that is if the validity of the GPD holds for some high threshold u, then it holds above any threshold $\tilde{u} > u$; see Coles [2001, Section 4.3]. Another solution that is more practical, in the sense that it is less subjective, is to set the threshold at various high quantile levels, which might be pre-determined in some

applications, e.g. Value-at-Risk computations, and to assess the stability of the GPD. See Scarrott and MacDonald [2012] for an overview of the literature regarding threshold selection.

1.1.2 Multivariate extreme value theory

We now extend the notions developed in Section 1.1.1 to the multivariate framework where not only the extremes of multiple series of observations are of interest, but also the dependence structure in their joint tails. This is of utmost importance for aggregate risk quantification, for example, where the interest is in co-occurrences of extreme events.

1.1.2.1 Multivariate extreme value distributions

Let $(\mathbf{Y}_i)_{i\geq 1}$ be a sequence of IID copies of a d-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ with marginal distributions F_j, $j = 1, \ldots, d$, and joint distribution F. We denote by $\mathbf{M}_n = (M_{n,1}, \ldots, M_{n,d})$ the vector of componentwise maxima, where $M_{n,j} = \max_{i=1}^{n} Y_{i,j}$ the sample maximum of the j-th component. Note that $\mathbf{M}_n$ is not necessarily observed as the componentwise maxima may occur at different times. Similarly to the univariate EVT, the vector $\mathbf{M}_n$ needs to be suitably normalized to avoid degeneracy of its limit law as $n \to \infty$. We suppose that there exist sequences $\{\mathbf{a}_n\} \subset \mathbb{R}^d_+$ and $\{\mathbf{b}_n\} \subset \mathbb{R}^d$ such that the normalized vector $(\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n$ converges in distribution to a random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)$ with joint distribution G and non-degenerate margins G_j, $j = 1, \ldots, d$. Then, the limiting distribution G is called the multivariate extreme value distribution (MEVD) [Resnick, 1987, Chapter 5.3] and the distribution function F is said to be in the maximum domain of attraction of G. The class of MEVDs has the nice property of coinciding with the class of max-stable distribution functions with non-degenerate margins [Beirlant et al., 2004, Chapter 8.2.1], where a d-variate distribution G is said to be max-stable if there exist vectors $\boldsymbol{\alpha}_k > \mathbf{0}$ and $\boldsymbol{\beta}_k$ such that

$$G(\boldsymbol{\alpha}_k \mathbf{z} + \boldsymbol{\beta}_k)^k = G(\mathbf{z}), \quad \mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}^d,$$

for any integer $k > 0$. Two subsequent results follow on from this property. The margins G_j must be max-stable, or equivalently $\mathrm{GEV}_{(\mu_j, \sigma_j, \xi_j)}$, which follows from the univariate EVT of Section 1.1.1 and the construction of $\mathbf{M}_n$, and the distribution function G is max-infinitely divisible, that is, $G^{1/k}$ is a distribution function, for any integer $k > 0$ [Balkema and Resnick, 1977]. In particular, there exists a measure μ^* defined on $[-\infty, \infty)^d$, such that

$$G(\mathbf{z}) = \exp\left\{-\mu^*\left(A_{\mathbf{z}}\right)\right\} = \exp\left\{-V(\mathbf{z})\right\}, \tag{1.1}$$

where $A_{\mathbf{z}} = \mathbb{R}^d \setminus ([-\infty, z_1] \times \cdots \times [-\infty, z_d])$. The measure μ^* is termed the exponent measure and contains the information regarding the structure of the dependence between the components of $\mathbf{Z}$. Equivalently, the function V is named the exponent function.

It is common practice in multivariate extreme value analysis to transform the marginal distributions G_j to a common scale and to study the dependence structure using max-stable distributions. A particularly convenient choice for the marginal distributions is the unit Fréchet distribution with distribution function $\Pr(Z_i \leq z) = \exp(-1/z)\mathbb{1}_{[0,\infty)}(z)$ and for which the normalizing sequences required for the limiting theorem to hold, are $\mathbf{a}_n = n$ and $\mathbf{b}_n = \mathbf{0}$. The marginal transformation is performed using probability integral transforms where the transformation

$$t_j(z) = -1/\log\{G_j(z)\} = \{1 + \xi_j(z - \mu_j)/\sigma_j\}_+^{1/\xi_j}, \quad z > 0,$$

is applied to the j-th margin, and the dependence structure of the transformed random vector is the same as the dependence structure of $\mathbf{Z}$. Without loss of generality, we will assume hereafter that $\mathbf{Y}$ has unit Fréchet margins and, consequently, that the exponent measure μ^* associated to G is concentrated on $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$. Then, G is said to be a simple max-stable distribution. The marginal assumption implies that $V(\infty, \ldots, z, \ldots, \infty) = 1/z$, whereas the max-stability of G implies the homogeneity of order -1, of V and μ^*, i.e., $V(t\mathbf{z}) = t^{-1}V(\mathbf{z})$ and $\mu^*(tB) = t^{-1}\mu^*(B)$, for $t > 0$ and B a Borel subset of $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$. In terms of its associated copula [Nelsen, 2006]

$$C(\mathbf{u}) = \exp\left[-V\left\{-1/\log(\mathbf{u})\right\}\right], \quad \mathbf{u} \in [0,1]^d,$$

the max-stability of G implies that its copula C belongs to the large class of max-stable copulas verifying

$$C(\mathbf{u}) = \left\{C\left(u_1^{1/k}, \ldots, u_d^{1/k}\right)\right\}^k, \quad \mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d, \quad k > 0.$$

Similarly to the univariate case (see Section 1.1.2), a point process characterization of the limiting behaviour of the componentwise maxima can be formulated. Define the point process $N_n(B) = \sum_{i=1}^n \mathbb{1}_{P_i \in B}$ on $\mathbb{R}_+^d$, where $P_i = \mathbf{Y}_i/n$. Then, N_n converges weakly to a non-homogeneous Poisson process N on $\mathbb{R}_+^d$ with mean measure μ^* [Resnick, 1987, Chapter 3.5]. Hence, by a computation of the void probability of the Poisson process N on the set $A_{\mathbf{z}}$, we recover the result (1.1) on the characterization of the limiting distribution G:

$$\Pr\left(\mathbf{M}_n/n \leq \mathbf{z}\right) = \Pr\left\{N_n(A_{\mathbf{z}}) = 0\right\} \to \exp\{-\mu^*(A_{\mathbf{z}})\} = \exp\{-V(\mathbf{z})\}, \quad \text{as } n \to \infty. \tag{1.2}$$

This result means that, by counting the stabilized observations that fall into a critical set where at least one component is large, one can estimate the dependence structure in the limiting MEVD.

Now that we have defined the exponent measure, we will rely on its homogeneity property to characterize, through a spectral decomposition, the class of multivariate extreme values distributions. We define the d-dimensional unit simplex S_d by

$$S_d = \{\boldsymbol{\omega} \in \mathbb{R}_+^d | \omega_1 + \cdots + \omega_d = 1\},$$

and map $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$ to $(0, \infty) \times S_d$ through the transformation

$$t(\mathbf{z}) = (r, \boldsymbol{\omega}) = \left(\sum_{j=1}^d z_j, \mathbf{z}/\sum_{j=1}^d z_j\right), \tag{1.3}$$

where r and $\boldsymbol{\omega}$ are the radial and angular parts of $\mathbf{z}$, respectively. With this parametrization, the intensity measure $d\mu^*$ factorizes into a product measure given by

$$d\mu^*(r, \boldsymbol{\omega}) = r^{-2} dr H(d\boldsymbol{\omega}), \tag{1.4}$$

where H is called the angular measure (or the spectral measure, for obvious reasons) and describes the dependence structure in the limiting MEVD G. The spectral representation (1.4) of the exponent measure μ^* is due to de Haan and Resnick [1977] and implies that, asymptotically, the radial part and the angular part of the rescaled observations $\mathbf{Y}_i/n$ are

independent. The link between the exponent function V defined in (1.1), and the angular measure H is easily recovered as follows

$$\mu^*(A_{\mathbf{z}}) = \int_{S_d} \int_{\min_{j=1,\dots,d}(z_j/\omega_j)}^{\infty} r^{-2} dr H(d\boldsymbol{\omega}) = \int_{S_d} \max_{j=1,\dots,d}(\omega_j/z_j) H(d\boldsymbol{\omega}) = V(\mathbf{z}). \tag{1.5}$$

This result, along with the marginal assumption on G, implies the following mean constraint on H

$$\int_{S_d} \omega_j H(d\boldsymbol{\omega}) = 1, \quad j = 1, \dots, d. \tag{1.6}$$

In what follows, we compute the expected number of points in the extreme set $E_{r_0} = \{\mathbf{z} \in (0,\infty)^d | \sum_{j=1}^d z_j/r_0 > 1\}$, for some $r_0 > 0$, which will turn out to be useful when performing inference on multivariate extremes,

$$\begin{aligned} \mu^*(E_{r_0}) &= \int_{S_d} \int_{\left(\sum_{j=1}^d \omega_j/r_0\right)^{-1}}^{\infty} r^{-2} dr H(d\boldsymbol{\omega}) \\ &= \int_{S_d} \left(\sum_{j=1}^d \omega_j/r_0\right) H(d\boldsymbol{\omega}) = \int_{S_d} r_0^{-1} H(d\boldsymbol{\omega}) = r_0^{-1}. \end{aligned} \tag{1.7}$$

As opposed to the univariate case where a parametric family of distributions characterizes all the possible limiting distributions of suitably normalized maxima, the class of multivariate extreme value distributions yields an infinite number of possible parametric representations. As a matter of fact, the validity of the MEVD G relies solely on its angular measure H satisfying the mean condition (1.6). Therefore, when it comes to extremal dependence modelling, we can either rely on flexible classes of parametric models, where the underlying features of the data at the extremal level (symmetry, pairwise dependence equality, etc) can be used to guide model choice, or go the non-parametric way.

1.1.2.2 Parametric models

Parametric representations of the exponent function V need to be flexible enough to capture the different features of the data, but parsimonious to endow the models with interpretability and inference feasibility, as things get quickly complicated when the dimension d increases. Along with the mean constraint (1.6) on H, the exponent function $V(\mathbf{z})$ needs to lie between the boundaries $\max(1/z_1, \dots, 1/z_d)$ and $1/z_1 + \dots + 1/z_d$ that correspond to the cases of perfect dependence and independence, respectively. In terms of the angular density, these boundaries imply that H places total mass 1 in the center of the simplex S_d, i.e., $(1/d, \dots, 1/d)$ (perfect dependence) and mass $1/d$ on the vertices of S_d (independence). Coles and Tawn [1991] proposed a method to construct flexible classes of parametric models for the angular density h of the angular measure H, when the MEVD G is absolutely continuous, and Boldi and Davison [2007] constructed a semi-parametric model for the angular density, based on mixtures of Dirichlet distributions. We describe below some of the most popular models for the extremal dependence structure. For an overview of the parametric models for MEVDs, refer to Kotz and Nadarajah [2000, Chapter 3] and Beirlant et al. [2004, Section 9.2.2]. Model selection is performed using likelihood-based model selection criteria such as the likelihood ratio test when the models are nested and the Akaike information criterion (AIC) when the models are non-nested, although these criteria do not measure the quality of the individual models.

The logistic model

Due to Gumbel [1961], the logistic model is one of the most famous models owing to its simplicity. The d-dimensional exponent function is

$$V(\mathbf{z}) = \left(\sum_{j=1}^{d} z_j^{-1/\alpha} \right)^{\alpha}, \quad \mathbf{z} > \mathbf{0},$$

where $\alpha \in (0, 1]$. This model is described by a unique dependence parameter with $\alpha = 1$ corresponding to independence and $\alpha \to 0$ corresponding to perfect extremal dependence. The logistic angular density is given by

$$h(\boldsymbol{\omega}; \alpha) = \prod_{j=1}^{d-1} (j/\alpha - 1) \prod_{j=1}^{d} \omega_j^{(-1/\alpha - 1)} \left(\sum_{j=1}^{d} \omega_j^{-1/\alpha} \right)^{\alpha - d}, \quad \boldsymbol{\omega} \in S_d.$$

As a result of the model being described by a unique parameter, the symmetric dependence structure at extremal levels lacks of flexibility and extensions were proposed.

Coles and Tawn [1991] proposed the asymmetric logistic model where asymmetry in the extremal dependence structure can be captured through an additional set of parameters as follows

$$V(\mathbf{z}) = \sum_{I \in \mathcal{P}_d} \left\{ \sum_{j \in I} (z_j / \beta_{I,j})^{-1/\alpha_I} \right\}^{\alpha_I}, \quad \mathbf{z} > \mathbf{0},$$

where $\mathcal{P}_d$ is the set of all non-empty subsets of $\{1, \ldots, d\}$, $\alpha_I \in (0, 1]$ for all $I \in \mathcal{P}_d$ and $\beta_{I,j} \in [0, 1]$ with $\sum_{I \in \mathcal{P}_d} \beta_{I,j} = 1$ for $j = 1, \ldots, d$ with the convention that $\beta_{I,j} = 0$ if $j \notin I$. Other extensions of the logistic model include the asymmetric negative logistic model [Joe, 1990] and the bilogistic model [Smith, 1990].

The Dirichlet model

A second model allowing for asymmetric extremal dependence structures is the (tilted) Dirichlet model proposed by Coles and Tawn [1991] and based on their general model construction of valid angular densities. The d-dimensional angular density is

$$h(\boldsymbol{\omega}; \boldsymbol{\alpha}) = \frac{\Gamma(\boldsymbol{\alpha}.\mathbf{1} + 1)}{(\boldsymbol{\alpha}.\boldsymbol{\omega})^{(d+1)}} \prod_{j=1}^{d} \frac{\alpha_j}{\Gamma(\alpha_j)} \left(\frac{\alpha_j w_j}{\boldsymbol{\alpha}.\boldsymbol{\omega}} \right)^{\alpha_j - 1}, \quad \boldsymbol{\omega} \in S_d,$$

where $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and $\mathbf{1}$ is the d-dimensional vector of ones. Symmetry is obtained when $\alpha_1 = \cdots = \alpha_d$, perfect dependence when $\alpha_j \to \infty$, and independence when $\alpha_j \to 0$, for all $j \in \{1, \ldots, d\}$. The exponent function V cannot, in this case, be computed in closed form.

The pairwise beta model

Applying similar construction arguments as for the Dirichlet model, Cooley et al. [2010] came up with the pairwise beta model with angular density

$$h(\boldsymbol{\omega}; \alpha, \boldsymbol{\beta}) = \frac{2(d-3)!\Gamma(d\alpha + 1)}{(d-1)\Gamma(2\alpha + 1)\Gamma\{(d-2)\alpha\}} \sum_{1 \le i < j \le d} h_{i,j}(\boldsymbol{\omega}; \alpha, \beta_{i,j}), \quad \boldsymbol{\omega} \in S_d,$$

where

$$h_{i,j}(\boldsymbol{\omega};\alpha,\beta_{i,j}) = (\omega_i+\omega_j)^{2\alpha-1}\left\{1-(\omega_i+\omega_j)\right\}^{(d-2)\alpha-d+2}\frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})}\left(\frac{\omega_i}{\omega_i+\omega_j}\right)^{\beta_{i,j}-1}\left(\frac{\omega_j}{\omega_i+\omega_j}\right)^{\beta_{i,j}-1}.$$

The parameter $\alpha > 0$ can be interpreted as a global dependence parameter, while each of the $\binom{d}{2}$ parameters, $\beta_{i,j} > 0$ describes the level of dependence between the i-th and the j-th components with larger values inducing stronger dependences. Again, the exponent function cannot be computed analytically but numerical evaluation is feasible.

The Hüsler–Reiss model

Another symmetric model due to Hüsler and Reiss [1989] is the Hüsler–Reiss model. This model corresponds to the limiting dependence structure of suitably normalized componentwise maxima of Gaussian variables with zero mean, unit variance, and pairwise correlations $\rho_{i,j}(n)$ increasing with the sample size n such that

$$\log n\{1-\rho_{i,j}(n)\} \underset{n\to\infty}{\longrightarrow} \lambda_{i,j}^2 \in [0,\infty), \quad \forall i,j \in \{1,\ldots,d\}, \quad i \neq j.$$

The exponent function associated to the Hüsler–Reiss model is

$$V(\mathbf{z}) = \sum_{j=1}^{d}\frac{1}{z_j}\Phi_{d-1}\left[\left\{\lambda_{i,j}+\frac{\log(z_i/z_j)}{2\lambda_{i,j}}\right\}_{\substack{i\in\{1,\ldots,d\}\\ i\neq j}};\tilde{\Lambda}_j\right], \quad \mathbf{z} > \mathbf{0},$$

where Φ_{d-1} is the $(d-1)$-dimensional Gaussian distribution function with correlation matrix $\tilde{\Lambda}_j$ whose elements are $(\lambda_{k,j}^2+\lambda_{i,j}^2-\lambda_{k,i}^2)/(2\lambda_{k,j}\lambda_{i,j})$, for $k, i \neq j$. The parameters $\lambda_{i,j}$ control the strength of dependence between the i-th and j-th components with complete dependence when $\lambda_{i,j} = 0$ and independence attained as $\lambda_{i,j} \to \infty$. Engelke et al. [2015, Proposition 3.5] derived the angular density of the Hüsler–Reiss model which is equal to

$$h(\boldsymbol{\omega};\boldsymbol{\lambda}) = \boldsymbol{\phi}_{d-1}\left[\left\{\lambda_{j,1}+\frac{\log(\omega_j/\omega_1)}{2\lambda_{j,1}}\right\}_{j\in\{2,\ldots,d\}};\tilde{\Lambda}_1\right]\left(\omega_1^2\prod_{j=2}^{d}2\lambda_{j,1}\omega_j\right)^{-1}, \quad \boldsymbol{\omega}\in S_d,$$

where $\boldsymbol{\lambda} = (\lambda_{i,j})_{i,j\in\{1,\ldots,d\}}$ and $\boldsymbol{\phi}_{d-1}$ is the $(d-1)$-dimensional Gaussian density with correlation $\tilde{\Lambda}_1$.

Figure 1.1 displays examples, in the trivariate case, of the angular densities $h(\boldsymbol{\omega};.)$ in the symmetric logistic case, the Dirichlet case, the pairwise beta case, and the Hüsler–Reiss case. Different values of the dependence parameters are considered to reflect different extremal dependence structures.

1.1.2.3 Inference

As discussed above, the characterization of the extremal dependence structure can be performed either parametrically, based on the different classes of parametric models presented in Section 1.1.2.2, or non-parametrically, where non-parametric estimation of the Pickands' dependence function [Pickands, 1981] (see Section 1.1.2.4) is performed. We describe in this section likelihood-based inference methods where multivariate extreme values are either defined as componentwise block maxima or as threshold exceedances, and we refer to Huser et al. [2016] for a comparison of the performances of these methods.

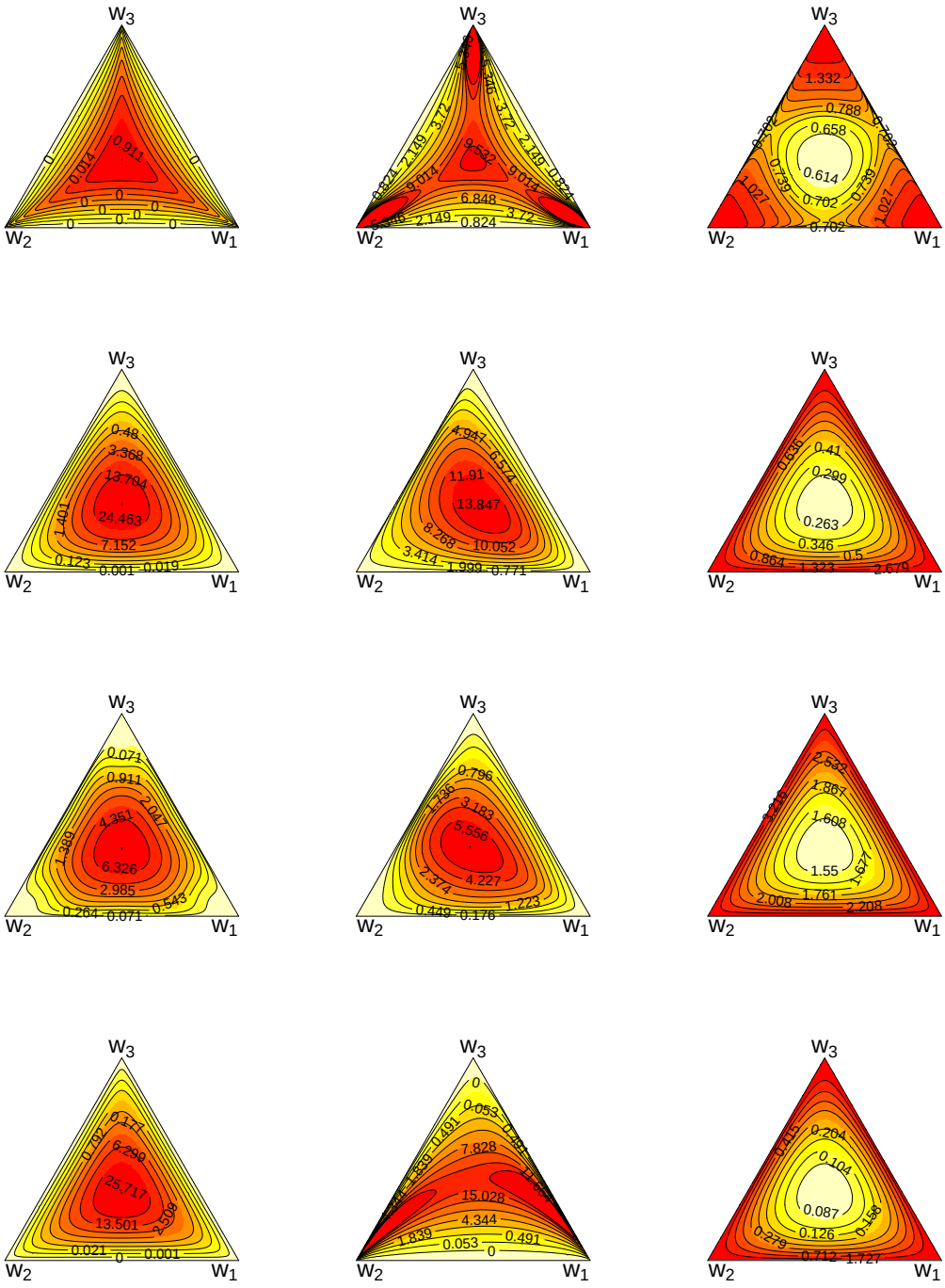

Figure 1.1: Examples of trivariate angular densities with different dependence parameters (from left to right) of the logistic model with $\alpha = 0.1, 0.6, 0.95$, the Dirichlet model with $\boldsymbol{\alpha} = (5,5,5), (1.5,2,3), (0.1,0.1,0.1)$, the pairwise beta model with $(\alpha, \boldsymbol{\beta}) = (6,2,2,2), (2,2,3,7), (0.5,1,1,1)$, and the Hüsler–Reiss model with $\boldsymbol{\lambda} = (0.3,0.3,0.3), (0.4,0.74,0.5), (2,2,2)$, from top to bottom.

The inference approaches rely on the validity of the asymptotic model (1.1) or the point process approximation, at subasymptotic levels and should hence be dealt with carefully as not only the rate of convergence of the univariate margins is of concern but also the

rate of convergence of the dependence structure; see Omey and Rachev [1991], Falk and Reiss [2002]. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be n realizations of a random vector $\mathbf{Y}$ with unit Fréchet margins and joint distribution F. We assume that F is in the max-domain of attraction of an MEVD G that depends on a vector of parameters $\boldsymbol{\theta}$, that is $\boldsymbol{\theta}$ characterizes the dependence structure at extremal levels.

Block maxima approach

Assume that N componentwise block maxima, denoted $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,d})$, are extracted from the data at hand. For an adequate parametric model for G, or equivalently (under the assumption of known margins) for V, the likelihood function is written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} g(\mathbf{z}_i),$$

where $g(\mathbf{z}) = \partial^d G(\mathbf{z})/\partial z_1 \ldots \partial z_d$. For $d = 2$, $g(z_1, z_2) = (V_1 V_2 - V_{12}) \exp\{-V(z_1, z_2)\}$, where $V_i(z_1, z_2) = \partial V(z_1, z_2)/\partial z_i$. When the dimension d increases, the likelihood function becomes intractable as it involves the partial derivatives of V with respect to the elements in all the sets included in the partition of $\{1, \ldots, d\}$, and involves hence B_d terms, where B_d is the Bell number. In order to overcome this issue, one possibility is to consider composite likelihoods [Padoan et al., 2010, Davison and Gholamrezaee, 2011] instead of the full likelihood. This solution results in a loss of efficiency [Varin et al., 2011] but has the advantage of simplifying the likelihood function, especially when pairwise likelihood is used, as only bivariate marginal densities are considered.

Threshold exceedances approach

We now describe three different threshold-based inference methods that rely on the Poisson point process approximation of observations belonging to an extremal set. Although the choice of the threshold (marginal or radial) results in a trade-off between bias and variance, it is often treated in practice as a fixed input with a small impact on the analysis procedure. Figure 1.2 illustrates, in the bivariate case, the extremal sets considered in these methods.

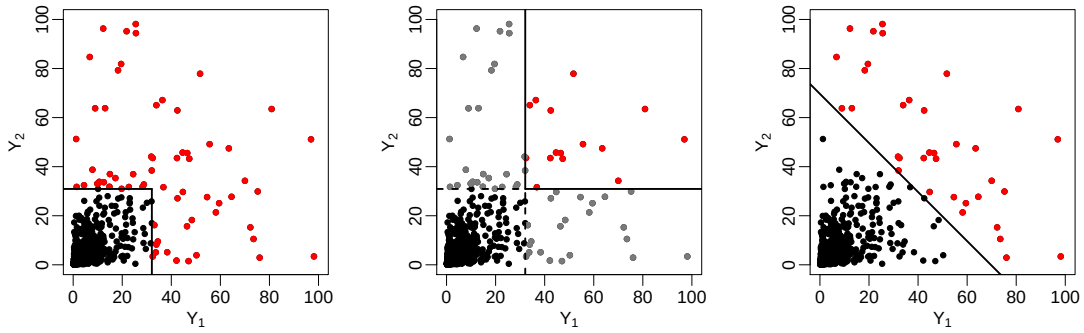

Figure 1.2: Examples of extremal sets in $\mathbb{R}_+^2$ used for threshold exceedances modelling. Red points correspond to the observations defined as extreme while grey points correspond to the censored observations.

- **Exceedances of a marginal threshold**
The joint distribution F being in the max-domain of attraction of an MEVD G, the approximation by a Poisson point process with intensity measure $d\mu^*$, of the rescaled points $\mathbf{Y}_i/n$, holds in the extreme set $A_{\mathbf{u}}$, where $\mathbf{u} = (u_1, \ldots, u_d)$ is a vector of high marginal thresholds; see left panel of Figure 1.2. We denote by $\mathbf{Y}^k = (Y_1^k, \ldots, Y_d^k)$ the $N_{\mathbf{u}}$ exceedances of $\mathbf{u}$. Inference for the exponent function V can then be performed based on the likelihood of the limiting Poisson process, i.e.,

$$\begin{aligned} L(\boldsymbol{\theta}) &= \exp\{-\mu^*(A_{\mathbf{u}/n})\} \prod_{k=1}^{N_{\mathbf{u}}} d\mu^*(\mathbf{Y}^k/n), \\ &= \exp\{-V(n^{-1}\mathbf{u})\} \prod_{k=1}^{N_{\mathbf{u}}} \{-V_{1:d}(\mathbf{Y}^k/n)\} \\ &\propto \exp\{-nV(\mathbf{u})\} \prod_{k=1}^{N_{\mathbf{u}}} \{-V_{1:d}(\mathbf{Y}^k)\}, \\ &\propto \exp\{-nV(\mathbf{u})\} \prod_{k=1}^{N_{\mathbf{u}}} h\left(Y_1^k / \sum_{j=1}^{d} Y_j^k, \ldots, Y_d^k / \sum_{j=1}^{d} Y_j^k\right), \end{aligned}$$

where $V_{1:d}$ is the derivative of V with respect to all its arguments and h is the angular density associated to V. The last proportionality sign is due to a result in Coles and Tawn [1991, Theorem 1] relating $V_{1:d}$ to h. Using this approach, extreme observations are defined as the observations that are extreme in at least one component and the convergence to the Poisson point process might be a strong assumption when not all the components are simultaneously extreme. This led Wadsworth and Tawn [2014] to consider a censoring scheme where, for each extreme observation (defined as above), the values of the components not exceeding the marginal threshold are replaced by the value of that threshold. Thus, these components contribute in the likelihood by the information that they do not exceed their corresponding marginal thresholds. This method can therefore suffer from a loss of efficiency as we consider less extreme observations contributing fully to the likelihood, i.e., only observations in the extreme set $(\mathbf{u}, \infty)$ are used fully; see middle panel of Figure 1.2 where red points contribute fully to the likelihood and grey points are censored and contribute partially. More precisely, censoring an extreme observation $\mathbf{Y}^k$ corresponds to replacing it by $\tilde{\mathbf{Y}}^k = (\tilde{Y}_1^k, \ldots, \tilde{Y}_d^k)$, where $\tilde{Y}_j^k = \max(Y_j^k, u_j)$. The Poisson process likelihood is now written as

$$L(\boldsymbol{\theta}) \propto \exp\{-nV(\mathbf{u})\} \prod_{k=1}^{N_{\mathbf{u}}} \{-V_{I_k}(\tilde{\mathbf{Y}}^k)\},$$

where I_k denotes the set of indices corresponding to the non-censored components of $\mathbf{Y}^k$ and V_{I_i} is the partial derivative of the exponent function V with respect to the elements in I_i.

- **Exceedances of a diagonal threshold**
As above, we rely on the Poisson point process approximation in extremal sets, to perform a likelihood-based inference. The extremal set is now defined as the observations $\mathbf{Y}_i$ with a radial component exceeding a high threshold u, i.e., $E_u = \{\mathbf{y} \in (0,\infty)^d | \sum_{j=1}^d y_j > u\}$; see right panel of Figure 1.2. We denote by $\mathbf{Y}^k$,

$k = 1, \ldots, N_u$ the extreme points in this region. The Poisson process likelihood over E_u is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \exp\{-\mu^*(E_{u/n})\} \prod_{k=1}^{N_u} d\mu^*(\mathbf{Y}^k/n), \\ &\propto \prod_{k=1}^{N_u} \{-V_{1:d}(\mathbf{Y}^k)\}, \\ &\propto \prod_{k=1}^{N_u} h\left(Y_1^k / \sum_{j=1}^{d} Y_j^k, \ldots, Y_d^k / \sum_{j=1}^{d} Y_j^k\right), \end{aligned}$$

where the first proportionality sign is due to the fact that the expected number of exceedances in extreme regions of the form E_u does not depend on the exponent function; see (1.7).

1.1.2.4 Measures of extremal dependence

The characterization of an MEVD relies on its marginal features, completely described by the univariate EVT, and on the structure of its dependence, determined by the exponent function. Various alternative measures, related to the exponent function, have been proposed in the literature, and we discuss in this section two of these measures. In the sequel, we assume that $\mathbf{Z} = (Z_1, \ldots, Z_d)$ is a simple max-stable random vector with exponent function V.

Pickands' dependence function

The Pickands' dependence function [Pickands, 1981] A is the restriction of the exponent function to the unit simplex S_d. In fact, the two functions are related through the relation

$$V(\mathbf{z}) = \left(\frac{1}{z_1} + \ldots + \frac{1}{z_d}\right) A\left(\frac{z_1}{z_1 + \ldots + z_d}, \ldots, \frac{z_d}{z_1 + \ldots + z_d}\right), \quad \mathbf{z} \in \mathbb{R}_+^d.$$

Thus, the function A measures the strength of extremal dependence along a direction $\boldsymbol{\omega} \in S_d$. The Pickands' dependence function A inherits of the boundary restrictions of the exponent function V, that is $\max(\omega_1, \ldots, \omega_d) \leq A(\boldsymbol{\omega}) \leq 1$ where the lower bound corresponds to the perfect extremal dependence case and the upper bound to the independence case. Moreover, A is a continuous and convex function on S_d and satisfies the border conditions $A(\mathbf{0}) = 1$ and $A(\mathbf{e}_j) = 1$, for $j = 1, \ldots, d$ and where $\mathbf{e}_j$ is the j-th unit vector in $\mathbb{R}^d$. These constraints are sufficient for the validity of A in the bivariate case but necessary and not sufficient when $d \geq 3$ (see Beirlant et al. [2004, p.257] for a counterexample). Thus, most of the proposed estimators of the Pickands' dependence function were restricted to the bivariate case where correction techniques based on the greatest convex minorant or constrained splines smoothing were considered [Deheuvels, 1991, Capéraà et al., 1997, Hall and Tajvidi, 2000a, Fils-Villetard et al., 2008, Zhang, 2008, Genest and Segers, 2009, Cormier et al., 2014]. Exceptions include Gudendorf and Segers [2012], Marcon et al. [2016] who considered the estimation of A for a general $d \geq 2$ and where corrections based on projection techniques and constrained Bernstein-polynomials representations were considered.
The original non-corrected Pickands' estimator of A in the bivariate case [Pickands, 1981], denoted $\hat{A}^{\mathrm{P}}$, relies on the random variable $\xi(\omega) = \min\{S/(1-\omega), T/\omega\}$, $\omega \in (0,1)$, where

$S = -\log\{F_1(Z_1)\}$, $T = -\log\{F_2(Z_2)\}$, and F_1 and F_2 are the marginal distributions[1]. As the distribution of $\xi(\omega)$ is exponential[2] with parameter $A(\omega)$, the estimator $\hat{A}^{\mathrm{P}}$ is defined as

$$\{\hat{A}^{\mathrm{P}}(\omega)\}^{-1} = N^{-1}\sum_{i=1}^{N}\xi_i(\omega), \quad \omega \in [0,1],$$

based on a sample $\{\mathbf{Z}_i = (Z_{i,1}, Z_{i,2})\}_{i=1}^{N}$, and where $\xi_i(\omega) = \min\{S_i/(1-\omega), T_i/\omega\}$, with $S_i = -\log\{F_1(Z_{i,1})\}$ and $T_i = -\log\{F_2(Z_{i,2})\}$, for $i = 1, \ldots, N$. The subsequent estimators of the Pickands' dependence function mentioned above are characterized by improved small-sample properties as well as asymptotic properties compared to the original Pickands' estimator.

Extremal coefficient

When the dimension d increases, visualization of the extremal dependence structure gets more complicated and simpler dependence measures are needed. The extremal coefficient [Smith, 1990, Coles, 1993] is a summary of the strength of extremal dependence and is defined as

$$\theta = V(1, \ldots, 1) = dA(1/d, \ldots, 1/d).$$

Loosely speaking, this coefficient represents the effective number of independent components in $\mathbf{Z}$ as

$$\begin{aligned}\Pr(\mathbf{Z} \le z) &= \Pr(Z_1 \le z, \ldots, Z_d \le z)\\ &= \exp\{-V(z\mathbf{1})\} = \{\exp(-1/z)\}^{V(1,\ldots,1)}\\ &= \{\exp(-1/z)\}^{\theta}. \end{aligned} \tag{1.8}$$

Therefore, the extremal coefficient lies between 1 (perfect dependence) and d (independence). The extremal coefficient can also be defined for a subset of the components, such as pairs of components, giving rise to the pairwise extremal coefficient

$$\theta_{i,j} = V(\mathbf{e}_{i,j}),$$

where $\mathbf{e}_{i,j}$ is the d-dimensional vector with components equal to ∞, except the i-th and j-th that are equal to 1. This coefficient is often used as a summary for spatial dependence as well as a validation tool for parametric modelling of multivariate extremes; see Davison et al. [2012] and the references therein. A self-consistency issue of the set of pairwise and higher orders extremal coefficients arises as the boundary condition on each of the $2^d - 1$ elements of this set is not sufficient to ensure that the extremal coefficients correspond to the same MEVD G. Schlather and Tawn [2002, 2003] discussed this issue and provided necessary and sufficient conditions on a set of extremal coefficients to be self-consistent along with two self-consistent extremal coefficient estimators.
Smith [1990], Coles and Dixon [1999] proposed a naive estimator of the extremal coefficient

[1] In our case where $\mathbf{Z}$ is a simple max-stable random vector, the marginal distributions are unit Fréchet.
[2]

$$\begin{aligned}\Pr\{\xi(\omega) \le x\} &= 1 - \Pr\{Z_1 < x^{-1}(1-\omega)^{-1}, Z_2 < x^{-1}\omega^{-1}\}\\ &= 1 - \exp[-V\{x^{-1}(1-\omega)^{-1}, x^{-1}\omega^{-1}\}]\\ &= 1 - \exp[-xV\{(1-\omega)^{-1}, \omega)^{-1}\}] = 1 - \exp\{-xA(\omega)\}.\end{aligned}$$

based on (1.8) and the fact that the inverse of a unit Fréchet random variable is unit exponentially distributed. As a matter of fact, in our case where $\mathbf{Z}$ has unit Fréchet margins, the relation (1.8) implies that $\{\max(Z_1, \ldots, Z_d)\}^{-1}$ is exponentially distributed with parameter θ. Therefore, given a sample $\{\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,d})\}_{i=1}^N$, the estimator $\hat{\theta}$ is defined as

$$\hat{\theta} = \frac{N}{\sum_{i=1}^N \{\max(Z_{i,1}, \ldots, Z_{i,d})\}^{-1}}.$$

Another estimator of the extremal coefficient based on the concept of madograms was proposed by Cooley et al. [2006], Naveau et al. [2009].

1.1.2.5 Asymptotic independence

In the previous sections, we discussed the main asymptotic results of multivariate extreme value theory under the assumption that the random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ is in the maximum domain of attraction of some MEVD G. We briefly mentioned the case of asymptotic independence, i.e., the case where the limiting distribution function G factorizes into the product of its marginal distributions and hence no information regarding the extremal dependence structure, at a penultimate level, is left. The asymptotic independence case implies the following equivalent statements

- $V(\mathbf{z}) = 1/z_1 + \ldots + 1/z_d$, for $\mathbf{z} \in \mathbb{R}_+^d$,
- H places mass $1/d$ on the d vertices of the unit simplex S_d, and
- $A(\boldsymbol{\omega}) = 1$, $\forall \boldsymbol{\omega} \in S_d$.

When $d = 2$, this situation arises when the following condition, due to Sibuya [1960], holds

$$\chi = \lim_{u \to 1} \Pr\{F_1(Y_1) > u | F_2(Y_2) > u)\} = 0, \tag{1.9}$$

where F_i is the marginal distribution of the i-th component of $\mathbf{Y}$. Loosely stated, this criterion translates to the variables Y_1 and Y_2 having isolated extremes in the limit but does not prevent these variables from being dependent at observable threshold levels. Ledford and Tawn [1996, 1997] addressed this issue by characterizing the rate of decay of the dependence (towards independence) in the joint tail of $\mathbf{Y}$ based on the concept of the coefficient of tail dependence $\eta \in (0, 1]$. Given that Y_1 and Y_2 are unit Fréchet distributed, Ledford and Tawn [1996] proposed to model the joint tail behaviour as

$$\Pr(Y_1 > y, Y_2 > y) \sim \mathcal{L}(y) y^{-1/\eta}, \quad y \to \infty,$$

where $\mathcal{L}$ is a slowly varying function at infinity, that is $\mathcal{L}(ay)/\mathcal{L}(y) \to 1$, as $y \to \infty$ and for all $a > 0$. This joint tail representation translates in terms of the survival copula $\bar{C}$ of the random vector (Y_1, Y_2) as

$$\bar{C}(u, u) \sim s(u) u^{1/\eta}, \quad u \to 0,$$

where s is a slowly varying function at 0, i.e., $s(au)/s(u) \to 1$, as $u \to 0$ and for all $a > 0$. Under this modelling assumption,

$$\Pr(Y_1 > y | Y_2 > y) \sim \mathcal{L}(y) y^{1-1/\eta}, \tag{1.10}$$

which implies that both situations where $\eta < 1$ or $\eta = 1$ and $\mathcal{L}(y) \to 0$ result in asymptotic independence; see Heffernan [2000] for a classification of the different dependence structures according to the value of η. However, when $\eta = 1$ and $\mathcal{L}(y) \nrightarrow 0$, the conditional upper tail probability in (1.10) does no longer converge to 0 when we move further into the joint tail and this situation corresponds to asymptotic dependence where the slowly varying function $\mathcal{L}$ depends on the exponent function V of the limiting MEVD. Ledford and Tawn [1996] proposed to estimate the tail dependence coefficient η based on the structure variable $T := \min(Y_1, Y_2)$, which conditional on exceeding a large threshold u, is approximately Generalized Pareto distributed[3] with scale parameter $u\eta$ and shape parameter η. Therefore, given a sample $\{\mathbf{Y}_i = (Y_{i,1}, Y_{i,2})\}_{i=1}^n$, one can fit a GPD to the threshold exceedances of $T_k = \min(Y_{k,1}, Y_{k,2})$, such that $T_k > u$, $k = 1, \ldots, N_u$, and derive the maximum likelihood estimate of η defined as

$$\hat{\eta} = N_u^{-1} \sum_{k=1}^{N_u} \log(T_k/u).$$

Resnick [2002], Maulik and Resnick [2004] extended the tail characterization by introducing the concept of hidden regular variation which is a property characterizing families of distributions possessing both multivariate regular variation [Basrak et al., 2002, Hult and Lindskog, 2002], a property equivalent to the convergence (1.2) of the componentwise maxima, and asymptotic independence. Wadsworth and Tawn [2012] extended the concept of the coefficient of tail dependence to the multivariate case by providing a flexible characterization of the tail behaviour based on the so-called angular dependence function, which can be seen as the analogue of the Pickands' dependence function but for asymptotically independent random vectors. Under this characterization, the authors introduced the class of inverted max-stable processes which have a one-to-one correspondence with the class of max-stable processes, as well as the class of max-mixture models that encompass both asymptotic dependence and independence. This idea of bridging asymptotic dependence and independence was extended by Huser et al. [2017], Huser and Wadsworth [2018], Huser et al. [2018].

1.2 Generalized additive models

A generalized additive model (GAM) [Hastie and Tibshirani, 1986, 1990] is a generalized linear model where the specification of the linear predictor, or equivalently a function of the conditional expectation of a response variable Y_i, involves smooth functions of subsets of a given vector of covariates $\mathbf{x} = (\tilde{x}_1, \ldots, \tilde{x}_p, x_1, \ldots, x_q)^\top$. For example,

$$g\{\mathrm{E}(Y_i|\mathbf{X} = \mathbf{x}_i)\} = g(\mu_i) = \tilde{\mathbf{X}}_i\tilde{\beta} + f_1(x_{1i}) + \ldots + f_q(x_{qi}), \tag{1.11}$$

[3]

$$\begin{aligned}
\Pr(T > t + u | T > u) &= \frac{\Pr(Y_1 > t + u, Y_2 > t + u)}{\Pr(Y_1 > u, Y_2 > u)} \\
&\sim \frac{\mathcal{L}\{(1 + t/u)u\}u^{-1/\eta}(1 + t/u)^{-1/\eta}}{\mathcal{L}(u)u^{-1/\eta}} \\
&\sim (1 + t/u)^{-1/\eta},
\end{aligned}$$

where we used the fact that $\mathcal{L}$ is slowly varying.

where g is a monotone link function ensuring that μ_i has the right range of values (between 0 and 1 if it is a probability, strictly positive in the Poisson setting, etc), $\tilde{\mathbf{X}}_i^\top \in \mathbb{R}^p$ corresponds to the i-th row of a model matrix for the parametric part of the model, and the f_j are smooth functions of the covariates. Here, smooth functions of multiple covariates can be considered in the model (1.11) and will briefly be described in Section 1.2.1. Model (1.11) allows for a considerable amount of flexibility, as opposed to parametric regressions. This flexibility comes at a cost as one needs to specify how the smooth functions are represented and how their smoothness is controlled.

1.2.1 Representation of the smooth functions

Suppose that $Y_1, \ldots, Y_n$ are independent observations of a response variable Y and that the covariates $x_{1,i}, \ldots, x_{q,i}$ are observed for each individual i. In a model fitting perspective, the smooth functions in (1.11) are represented as linear combinations of basis functions, whether they are or can be closely approximated by elements of the basis [Wood, 2017, Chapter 4]. Different smooth functions in the same model can be represented using different bases: natural cubic splines, penalized regression splines, thin plate splines, tensor product splines, etc. Once the basis is set, one needs to choose an appropriate set of (known) basis functions $B_{j\cdot}$ so that f_j can be represented as

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{jk} B_{jk}(x), \quad j = 1, \ldots, q,$$

where β_{jk} are the parameters to be estimated and m_j is the number of used basis functions, termed also the rank of the basis expansion. In a matrix notation, each smooth term $\mathbf{f}_j = (f_j(x_{j1}), \ldots, f_j(x_{jn}))^\top$ can be written as

$$\mathbf{f}_j = \mathbf{X}^j \boldsymbol{\beta}_j,$$

where $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jm_j})^\top$ and $X_i^j = (B_{j1}(x_{ji}), \ldots, B_{jm_j}(x_{ji}))^\top$ is the i-th row of $\mathbf{X}^j$. The identifiability constraint $\mathbf{1}^\top \mathbf{f}_j = 0$ is imposed on the smooth terms f_j so that they are orthogonal to the intercept, implicitly included in the parametric part. The model (1.11) can now be written in a matrix notation

$$g(\mu_i) = \mathbf{X}_i \boldsymbol{\beta},$$

with $\mathbf{X}_i$ is the i-th row of $\mathbf{X} = (\tilde{\mathbf{X}}, \mathbf{X}^1, \ldots, \mathbf{X}^q)$ and $\boldsymbol{\beta} = (\tilde{\boldsymbol{\beta}}^\top, \boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_q^\top)^\top \in \mathbf{B}$. The finite-basis representation of the smooth functions results in a linear representation of the GAM simplifying hence the estimation of its parameters. Nonetheless, as mentioned above, one needs to control the fidelity to the data of the smooth functions to avoid overfitting. A wiggliness measure is then defined for each smooth function and is expressed in terms of the integrated second derivative of the smooth. For the j-th smooth term, the penalty is defined as $\boldsymbol{\beta}^\top \mathbf{S}_j \boldsymbol{\beta}$, where $\mathbf{S}_j$ is a positive definite matrix that depends on the basis functions. The idea of the penalization approach is to allow the basis dimension (rank) to be quite high but to control the smoothness of the model via a penalty term. A low wiggliness penalty results in a very wiggly function f_j and a high wiggliness penalty in a rather smooth function f_j.

One way to represent regression splines of different orders is by use of the B-spline basis [de Boor, 1978]. The B-splines are appealing due to their smooth interpolation properties and to their minimal support leading to an efficient and stable inference. To

represent a spline of order $Q+1$ [Hastie et al., 2009] using an m parameter B-spline basis, one needs to define $m+Q+2$ knots $s_1 < \ldots < s_{m+Q+2}$, on which the B-spline functions are evaluated and such that, due to their strictly local support, each B-spline function $B_\cdot^Q$ takes non-zero values only between $Q+3$ adjacent knots[4]. The spline is then represented as

$$f(x) = \sum_{k=1}^{m} \beta_k B_k^Q(x), \tag{1.12}$$

where

$$\begin{aligned} B_k^Q(x) &= \frac{x - x_k}{x_{k+Q+1} - x_k} B_k^{Q-1}(x) + \frac{x_{k+Q+2} - x}{x_{k+Q+2} - x_{k+1}} B_{k+1}^{Q-1}(x), \quad k = 1, \ldots, m, \\ B_k^{-1}(x) &= \mathbb{1}_{\{s_k \leq x \leq s_{k+1}\}}. \end{aligned}$$

Figure 1.3 illustrates the representation of a smooth curve by rank 10 B-spline bases of different orders.

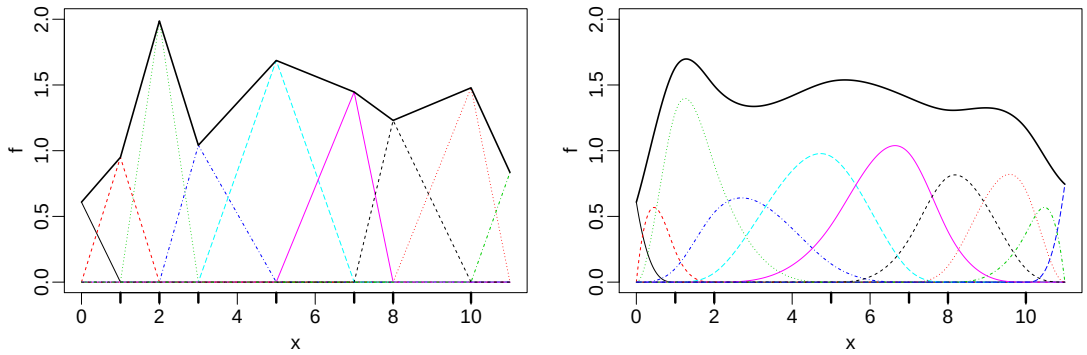

Figure 1.3: B-spline basis functions of order 1 (left plot) and order 3 (right plot). The thick curves show the smooth curves obtained by multiplication of the B-spline functions by their associated coefficients. The internal knots are displayed by thick tick-marks.

As the smoothness of f needs to be controlled, the basis representation (1.12) is modified and the B-splines are replaced by the P-splines [Eilers and Marx, 1996]. The P-splines are low rank smoothers defined as in (1.12) but with equidistant knots and a first-order difference penalty on adjacent coefficients. This results in a penalty matrix $\mathbf{S}$ that is band diagonal and easily determined based on the derivatives of the B-splines.

When the smooth is a function of multiple covariates, tensor product bases [de Boor, 1978] are needed to represent such functions. In the simplest setting with two covariates x_1 and x_2, the construction of the basis representation of $f_{x_1 x_2}$ starts by assuming that low rank basis representations of the univariate smooth functions f_{x_1} and f_{x_2} exist. Then, the coefficients of the basis representation of f_{x_1} are assumed to be smooth functions of x_2 and hence replaced by the basis representation of f_{x_2}. The construction is quite straightforward and the result is independent of the order in which the covariates are treated. The penalty term of such a construction is built in a similar fashion and has different penalties associated to each (re-parametrized) marginal smooth. See Wood [2017, Section 5.6] for technical details on tensor product constructions and penalties.

[4]Strictly speaking, the spline is evaluated on the interval $[s_{Q+2}, s_{m+1}]$ which contains the internal knots. The locations of the knots outside this interval are arbitrary.

1.2.2 Model fitting

When GAM originated in Hastie and Tibshirani [1986], the authors proposed a backfitting technique to fit the model. This technique estimates iteratively the smooth functions by smoothing the partial residuals of the model defined as the difference between the response and its expected value obtained by ignoring the contribution of the smooth to be estimated. Although the backfitting algorithm is very flexible as it allows the use of a wide variety of smoothers, and not only smoothers that can be represented using basis functions with an associated penalty term, the algorithm does not allow for an automatic smoothness selection as opposed to the penalized smoothing spline approach of Wood [2000] that we describe in this section.

Once the penalties associated to the smooth functions in (1.11) are set, the vector of parameters $\boldsymbol{\beta}$ is estimated by maximizing the penalized log-likelihood function [Green, 1987]

$$\ell_p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \ell(\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}^\top \mathbf{S}(\boldsymbol{\gamma})\boldsymbol{\beta},$$

where

$$\begin{aligned} \mathbf{S}(\boldsymbol{\gamma}) &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}(\boldsymbol{\gamma}) \end{pmatrix}, \\ \ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \ell(Y_i; \mu_i) = \sum_{i=1}^{n} \ell\{Y_i; g^{-1}(\mathbf{X}_i\boldsymbol{\beta})\} \end{aligned}$$

is the log-likelihood of the model, and $\bar{\mathbf{S}}(\boldsymbol{\gamma}) = \sum_{j=1}^{q} \gamma_j \mathbf{S}_j$ is the total wiggliness penalty. The first p-dimensional diagonal block in $\mathbf{S}(\boldsymbol{\gamma})$ corresponds to the zero-penalty related to the parametric part of the model (1.11). The vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ is the vector of smoothing parameters controlling the trade-off between goodness-of-fit and smoothness of the model. Given values for the smoothing parameters, $\boldsymbol{\beta}$ is estimated by a penalized iteratively re-weighted least squares (PIRLS) algorithm [Green, 1984, Wood, 2017] based on Fisher scoring or a full Newton–Raphson algorithm.

To avoid clutter in the notations, we suppose in what follows an identity link function. In order to perform a Newton–Raphson step, we compute the derivatives of first and second orders of the penalized log-likelihood with respect to the parameter of interest $\boldsymbol{\beta}$

$$\begin{aligned} \mathbf{u}_p(\boldsymbol{\beta}) &= \frac{\partial \ell_p(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\mu}} - \mathbf{S}(\boldsymbol{\gamma})\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{u}(\boldsymbol{\beta}) - \mathbf{S}(\boldsymbol{\gamma})\boldsymbol{\beta}, \\ \mathbf{W}_p(\boldsymbol{\beta}) &= -\frac{\partial^2 \ell_p(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^\top} = -\mathbf{X}^\top \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\mu}\boldsymbol{\mu}^\top}\mathbf{X} + \mathbf{S}(\boldsymbol{\gamma}) = \mathbf{X}^\top \mathbf{W}(\boldsymbol{\beta})\mathbf{X} + \mathbf{S}(\boldsymbol{\gamma}), \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$. We denote by $\mathbf{u}^{(a)} = \mathbf{u}(\boldsymbol{\beta}^{(a)})$ and $\mathbf{W}^{(a)} = \mathbf{W}(\boldsymbol{\beta}^{(a)})$ the first order and the (negative) second order derivatives of $\ell(\boldsymbol{\beta})$ obtained at the a-th step of the Newton–Raphson algorithm[5], which now has the form

$$\begin{aligned} \boldsymbol{\beta}^{(a)} &= \boldsymbol{\beta}^{(a-1)} + \left\{\mathbf{X}^\top \mathbf{W}^{(a-1)}\mathbf{X} + \mathbf{S}(\boldsymbol{\gamma})\right\}^{-1} \left\{\mathbf{X}^\top \mathbf{u}^{(a-1)} - \mathbf{S}(\boldsymbol{\gamma})\boldsymbol{\beta}^{(a-1)}\right\} \\ &= \left\{\mathbf{X}^\top \mathbf{W}^{(a-1)}\mathbf{X} + \mathbf{S}(\boldsymbol{\gamma})\right\}^{-1} \mathbf{X}^\top \mathbf{W}^{(a-1)}\mathbf{z}^{(a-1)}, \end{aligned} \tag{1.13}$$

[5]This results in a full Newton algorithm and a thorough inspection of the positive definiteness of $\mathbf{W}^{(a)}$, at each step, needs to be conducted. One could work instead with the expected information matrix if the latter is easily obtainable, resulting hence in a Fisher scoring algorithm.

where $\mathbf{z}^{(a-1)} = \mathbf{X}\boldsymbol{\beta}^{(a-1)} + \mathbf{W}^{(a-1)^{-1}}\mathbf{u}^{(a-1)}$ are the pseudo-responses. Hence, $\boldsymbol{\beta}^{(a)}$ is the minimizer of the penalized weighted least squares equation, i.e.,

$$\boldsymbol{\beta}^{(a)} = \underset{\boldsymbol{\beta}\in\mathbf{B}}{\operatorname{argmin}} \left\{ \left(\mathbf{z}^{(a-1)} - \mathbf{X}\boldsymbol{\beta}\right)^{\top} \mathbf{W}^{(a-1)} \left(\mathbf{z}^{(a-1)} - \mathbf{X}\boldsymbol{\beta}\right) + \boldsymbol{\beta}^{\top}\mathbf{S}(\boldsymbol{\gamma})\boldsymbol{\beta} \right\}.$$

The described PIRLS step is iterated until convergence which can be defined in terms of the difference between the penalized likelihoods or the difference between the resulting predictors.

We now focus on the selection of the smoothing parameters $\boldsymbol{\gamma}$ which were supposed fixed in the PIRLS algorithm. There are two main approaches to the selection of the smoothing parameters: the first one consists of minimizing prediction error criteria such as the Akaike's information criterion (AIC), cross validation or generalized cross validation (GCV) [Craven and Wahba, 1978, Hastie and Tibshirani, 1990], whereas the second approach is based on maximum marginal likelihood [Anderssen and Bloomfield, 1974] or restricted maximum likelihood methods [Wahba, 1985, Wood et al., 2016]. A comprehensive overview of smoothness selection criteria is given in Wood [2017, Section 6.2]. We now give details on the GCV criterion, used to select model smoothness throughout this thesis and implemented in many statistical software.

The estimation of the smoothness parameters is based on minimizing the squared prediction error but such that the objective function is invariant to orthogonal transformation. This is a desired property as the PIRLS algorithm, performed given fixed smoothing parameters, should result in identical estimates of $\boldsymbol{\beta}$ whether the data are rotated or not. Golub et al. [1979] suggested the use of the GCV as a suitable variant of the ordinary cross validation criterion, that satisfies the invariance property. Let $\mathbf{A}$ be the influence matrix of the weighted penalized linear model, that is

$$\mathbf{A}(\boldsymbol{\gamma}) = \mathbf{X}\left\{\mathbf{X}^{\top}\mathbf{W}(\boldsymbol{\beta})\mathbf{X} + \mathbf{S}(\boldsymbol{\gamma})\right\}^{-1}\mathbf{X}^{\top}\mathbf{W}(\boldsymbol{\beta}).$$

Then, the GCV score is defined as

$$\mathrm{GCV}(\boldsymbol{\gamma}) = \frac{n\left\{\mathbf{z} - \mathbf{A}(\boldsymbol{\gamma})\mathbf{z}\right\}^{\top}\mathbf{W}(\boldsymbol{\beta})\left\{\mathbf{z} - \mathbf{A}(\boldsymbol{\gamma})\mathbf{z}\right\}}{\left[n - \operatorname{tr}\left\{\mathbf{A}(\boldsymbol{\gamma})\right\}\right]^2}, \tag{1.14}$$

with $\mathbf{z}$ being the pseudo-responses defined in the PIRLS algorithm. Apart from the choice of the smoothness selection criterion, we need to define a strategy to combine the estimation of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ in an efficient and stable way. Two possible ways of tackling this problem are available. The first method known as "performance-oriented iteration" [Gu, 1992] consists of optimizing the GCV (1.14) by Newton's method, at each step a of the PIRLS algorithm, i.e., $\mathbf{W}(\boldsymbol{\beta})$ is replaced by $\mathbf{W}^{(a)}$ and $\mathbf{z}$ by $\mathbf{z}^{(a)}$ in (1.14). The final smoothing parameter estimates $\hat{\boldsymbol{\gamma}}$ are obtained by minimizing (1.14) evaluated at the model coefficient estimates $\hat{\boldsymbol{\beta}}$, obtained at convergence of the PIRLS algorithm. It is argued in Wood [2017, Section 6.5] that the number of PIRLS steps is around the same as would be needed with fixed smoothing parameters. The second method known as "outer-iteration" requires the minimization of the GCV score in a step "outer" to the PIRLS algorithm. This strategy can be described as a nested optimization procedure where at each step of the GCV minimization algorithm (Newton or quasi Newton's method), a PIRLS algorithm is performed with the proposed smoothing parameters, and results in the estimate $\hat{\boldsymbol{\beta}}$ that is used for the next step of the GCV minimization algorithm. The alternation of steps is performed until convergence of the GCV minimization algorithm

and results in the smoothing parameter estimates $\hat{\boldsymbol{\gamma}}$. This method requires however the first and second order derivatives of the GCV and hence the model coefficient estimates, with respect to the (logarithm of the) smoothing parameters. See Wood [2017, Section 6.6] for the technical details of this technique.

Finally, we define the effective degrees of freedom (EDF) of the GAM (1.11) as the trace of the influence matrix $\mathbf{A}(\boldsymbol{\gamma})$ obtained at convergence, i.e.,

$$\text{EDF} = \text{tr}\left\{\mathbf{A}(\hat{\boldsymbol{\gamma}})\right\}.$$

The effective degrees of freedom of a particular smoother f_j are obtained by summing the diagonal terms of the matrix $\left\{\mathbf{X}^\top\mathbf{W}(\hat{\boldsymbol{\beta}})\mathbf{X} + \mathbf{S}(\hat{\boldsymbol{\gamma}})\right\}^{-1}\mathbf{X}^\top\mathbf{W}(\hat{\boldsymbol{\beta}})\mathbf{X}$, corresponding to the coefficients $\boldsymbol{\beta}_j$.

Before closing this section, the existence of shape constrained GAM [Pya and Wood, 2015] and vector GAM (VGAM) [Yee, 2015] should be highlighted. The first extension allows the construction of shape constrained splines (monotonicity, convexity/concavity) by formulating these constraints on the P-splines coefficients scale. The second extension allows the generalized additive modelling of multiple predictors where the response variable can be either univariate or multivariate. Moreover, VGAMs offer an infrastructure where the effects of the covariates can be restricted on different predictors while tuned to be the same for others (spicing up a bit the fitting algorithm).

1.3 Non-stationary extremes

As mentioned in Section 1.1.1, non-stationarity in univariate extremes can be handled in various ways. Fully parametric modelling of the dependence of the location, scale, and shape parameters on a set of covariates can be formulated and likelihood-based inference performed [Smith, 1990, Katz et al., 2002, Northrop and Jonathan, 2011]. Such models are straightforward to fit, provided that the number of parameters to estimate is reasonable. However, unless a physical explanation and justification for the chosen parametric form exists, these models might describe the data poorly. Non- or semi-parametric models do not suffer from this limitation as they allow flexible forms of dependence on the covariates while including the parametric form as a submodel. There are two main approaches to non- or semi-parametric modelling of extremes: the local likelihood approach advocated by Davison and Ramesh [2000], Hall and Tajvidi [2000b], Ramesh and Davison [2002] and the penalized smoothing spline approach advocated by Pauli and Coles [2001], Chavez-Demoulin and Davison [2005], Yee and Stephenson [2007]. The local likelihood approach relies on a bandwidth parameter that confers to the model its smoothness, whereas the penalized spline approach relies on the notion of penalized likelihood, as seen in the general case of generalized additive modelling in Section 1.2.

While it is relatively straightforward to take into account non-stationarity at extremal levels in the univariate framework, the difficulty of this task quickly escalates in the multivariate framework as the additional non-stationarity in the extremal dependence structure needs to be properly modelled. Moreover, as discussed in Section 1.1.2.1, the class of extremal dependence structures cannot be fully parametrized and whether the angular measure or the Pickands' function is to be modelled, shape or mean constraints such as (1.6) need to be accounted for. Few attempts towards filling this gap in MEVT literature have been proposed: de Carvalho and Davison [2014] proposed a semi-parametric approach linking angular densities from different populations through exponential tilting, Jonathan

et al. [2014a] extended the conditional dependence model of Heffernan and Tawn [2004] to include the influence of covariates on both the marginal and the dependence parameters, de Carvalho et al. [2018] constructed kernel-based estimators for bivariate angular densities depending on a continuous covariate, and Huser and Genton [2016] developed max-stable models with a non-stationary correlation function depending parametrically on a set of covariates. Two limitations of these available techniques are worth highlighting. The first is a computational limitation of the non-parametric methods that either necessitate intensive constrained optimization techniques when multiple (continuous) covariates are included, or are restricted to the inclusion of the effect of one unique covariate when the above mentioned constraints are embedded in the model. The second limitation is a practical one and comes within the scope of the application of ultimate models at penultimate levels and to settings where (full) asymptotic independence may arise. In that sense, the existing literature on non-stationary multivariate extremes relies on the assumption of asymptotic dependence to model limiting covariate-varying dependence structures whereas in many applications where this assumption does not hold (rainfall data for example), modelling non-stationarity in the decay rates towards independence is more informative and essential for accurate extrapolation to unobserved extreme levels.

Chapter 2

Regression type models for extremal dependence

The starting point for the work presented in this chapter is the extension of the work of Chavez-Demoulin and Davison [2005] to the multivariate setting. While Chavez-Demoulin and Davison [2005] adapt the framework of generalized additive modelling to the parameters of a peaks-over-threshold model to take into account the effect of covariates on the tail behaviour, we adapt the framework of vector generalized additive modelling to the parameters of a parametric model for the angular density to take into account the effect of covariates on the dependence structure of the joint tails in multidimensional distributions; see Section 1.1.2.2. We rely on the Poisson point process approximation in the extremal set described by the exceedances of a diagonal threshold (Section 1.1.2.3) to formulate a likelihood for the angular density which we modify adequately when the parameters are linked through splines to a set of predictors. We devise a maximum penalized log-likelihood estimator, discuss details of the estimation procedure, and derive its consistency and asymptotic normality under mild regularity conditions. We perform a simulation study suggesting that the proposed method performs well in a wealth of simulation scenarios by accurately recovering the true covariate-adjusted angular density. An empirical analysis is conducted and reveals relevant dynamics of the dependence between extreme air temperatures in two alpine resorts during the winter season.

2.1 Introduction

In this chapter, we address an extension of the standard approach for modelling non-stationary univariate extremes to the multivariate setting. In the univariate context, the limiting distribution for the maximum of a sequence of independent and identically distributed random variables, derived by Fisher and Tippett [1928], is given by a generalized extreme value distribution characterized by three parameters: μ (location), σ (scale), and ξ (shape). To take into account the effect of a vector of covariates $\mathbf{x}$, one can let these parameters depend on $\mathbf{x}$, and the resulting generalized extreme value distribution takes the form

$$G_{(\mu_{\mathbf{x}},\sigma_{\mathbf{x}},\xi_{\mathbf{x}})}(y) = \exp\left[-\left\{1+\xi_{\mathbf{x}}\left(\frac{y-\mu_{\mathbf{x}}}{\sigma_{\mathbf{x}}}\right)\right\}_{+}^{-1/\xi_{\mathbf{x}}}\right], \tag{2.1}$$

where $(a)_+ = \max\{0, a\}$; see Coles [2001, Chapter 6], Pauli and Coles [2001], Chavez-Demoulin and Davison [2005], Yee and Stephenson [2007], Wang and Tsai [2009], Eastoe

and Tawn [2009], and Chavez-Demoulin and Davison [2005] for related approaches.

In the multivariate context, consider $\mathbf{Y}^i = (Y_1^i, \ldots, Y_d^i)^\top$ independent and identically distributed random vectors with joint distribution F, and unit Fréchet marginal distribution functions $F_j(y) = \exp(-1/y)$, for $y > 0$. Pickands' representation theorem [Coles, 2001, Theorem 8.1] states that the law of the standardized componentwise maxima, $\mathbf{M}_n = n^{-1} \max\{\mathbf{Y}^1, \ldots, \mathbf{Y}^n\}$, converges in distribution to a multivariate extreme value distribution, $G_H(\mathbf{y}) = \exp\{-V_H(\mathbf{y})\}$, with

$$V_H(\mathbf{y}) = \int_{S_d} \max\left(\frac{w_1}{y_1}, \ldots, \frac{w_d}{y_d}\right) \mathrm{d}H(\mathbf{w}). \tag{2.2}$$

Here H is the so-called angular measure, that is, a positive finite measure on the unit simplex $S_d = \left\{(w_1, \ldots, w_d) \in [0, \infty)^d : w_1 + \cdots + w_d = 1\right\}$ that needs to obey

$$\int_{S_d} w_j \, \mathrm{d}H(\mathbf{w}) = 1, \quad j = 1, \ldots, d. \tag{2.3}$$

The function $V(\mathbf{y}) \equiv V_H(\mathbf{y})$, is the so-called exponent measure and is continuous, convex, and homogeneous of order -1, i.e., $V(t\mathbf{y}) = t^{-1}V(\mathbf{y})$ for all $t > 0$.

The class of limiting distributions of multivariate extreme values yields an infinite number of possible parametric representations [Coles, 2001, Chapter 8], as the validity of a multivariate extreme value distribution is conditional on its angular measure H satisfying the moment constraint (2.3). Therefore, most literature has focused on the estimation of the extremal dependence structures described by spectral measures or equivalently angular densities [Boldi and Davison, 2007, Einmahl et al., 2009, de Carvalho et al., 2013, Sabourin and Naveau, 2014, Hanson et al., 2017]. Related quantities, such as the Pickands' dependence function [Pickands, 1981] and the stable tail dependence function [Huang, 1992, Drees and Kaufmann, 1998], were investigated by many authors [Einmahl et al., 2006, Gudendorf and Segers, 2012, Wadsworth and Tawn, 2013, Marcon et al., 2016]. A wide variety of parametric models for the spectral density that allow flexible dependence structures were proposed [Kotz and Nadarajah, 2000, Section 3.4].

However, few papers were able to satisfactorily address the challenging but incredibly relevant setting of modelling nonstationarity at joint extreme levels. Some exceptions include de Carvalho and Davison [2014], who proposed a nonparametric approach, where a family of spectral densities is constructed using exponential tilting. Castro and de Carvalho [2017] developed an extension of this approach based on covariate-varying spectral densities. However, these approaches are limited to replicated one-way ANOVA types of settings. de Carvalho [2016] advocated the use of covariate-adjusted angular densities, and Escobar-Bach et al. [2016] discussed estimation—in the bivariate and covariate-dependent framework—of the Pickands' dependence function based on local estimation with a minimum density power divergence criterion. Finally, Mhalla et al. [2017a] constructed, in a nonparametric framework, smooth models for predictor-dependent Pickands' dependence functions based on generalized additive models.

Our approach is based on a non-linear model for covariate-varying extremal dependences. Specifically, we develop a vector generalized additive model that flexibly allows the extremal dependence to change with a set of covariates, but—keeping in mind that extreme values are scarce—it borrows strength from a parametric assumption. In other words, the goal is to develop a regression model for the extremal dependence through the parametric specification of an extremal dependence structure and then to model the parameters of that structure through a vector generalized additive model

(VGAM) [Yee and Wild, 1996, Yee, 2015]. One major advantage over existing methods is that our model may be used for handling an arbitrary number of dimensions and covariates of different types, and it is straightforward to implement, as illustrated in the R code [R Development Core Team, 2016] that can be found in the following repository https://github.com/lindamhalla/Regression_type_models_for_extremal_dependence .

The remainder of this chapter is organized as follows. In Section 2.2 we introduce the proposed model for covariate-adjusted extremal dependences. In Section 2.3 we develop our penalized likelihood approach and give details on the asymptotic properties of our estimator. In Section 2.4 we assess the performance of the proposed methods. An application to extreme temperatures in the Swiss Alps is given in Section 2.5. We close the chapter in Section 2.6 with a discussion.

2.2 Flexible covariate-adjusted angular densities

2.2.1 Statistics of multivariate extremes: preparations and background

The functions H and V in (2.2) can be used to describe the structure of dependence between the extremes, as in the case of independence between the extremes, where $V(\mathbf{y}) = \sum_{j=1}^{d} 1/y_j$, and in the case of perfect extremal dependence, where $V(\mathbf{y}) = \max\{1/y_1, \ldots, 1/y_d\}$. As a consequence, if H is differentiable with angular density denoted h, the more mass around the barycenter of S_d, $(d^{-1}, \ldots, d^{-1})$, the higher the level of extremal dependence. Further insight into these measures may be obtained by considering the point process $P_n = \{n^{-1}\mathbf{Y}^i : i = 1, \ldots, n\}$. Following de Haan and Resnick [1977] and Resnick [1987, Section 5.3], as $n \to \infty$, P_n converges to a non-homogeneous Poisson point process P defined on $[\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\}$ with a mean measure μ that verifies

$$\mu(A_{\mathbf{y}}) = V(\mathbf{y}),$$

where $A_{\mathbf{y}} = \mathbb{R}^d \setminus ([-\infty, y_1] \times \cdots \times [-\infty, y_d])$.

There are two representations of the intensity measure of the limiting Poisson point process P that will be handy for our purposes. First, it holds that

$$\mu(\mathrm{d}\mathbf{y}) = -V_{1:d}(\mathbf{y})\ \mathrm{d}\mathbf{y}, \tag{2.4}$$

with $V_{1:d}$ being the derivative of V with respect to all its arguments [Resnick, 1987, Section 5.4]. Second, another useful factorization of the intensity measure $\mu(\mathrm{d}\mathbf{y})$, called the spectral decomposition, can be obtained using the following decomposition of the random variable $\mathbf{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$ into radial and angular coordinates,

$$(R, \mathbf{W}) = \left(\|\mathbf{Y}\|, \frac{\mathbf{Y}}{\|\mathbf{Y}\|} \right), \tag{2.5}$$

where $\|\cdot\|$ denotes the L_1-norm. It can be shown that [Beirlant et al., 2004, Section 8.2.3] the limiting intensity measure factorizes across radial and angular components as follows:

$$\mu(\mathrm{d}\mathbf{y}) = \mu(\mathrm{d}r \times\ \mathrm{d}\mathbf{w}) = \frac{\mathrm{d}r}{r^2}\, \mathrm{d}H(\mathbf{w}).$$

The spectral decomposition (2.5) allows the separation of the marginal and the dependence parts in the multivariate extreme value distribution G_H, with the margins being unit Fréchet and the dependence structure being described by the angular measure H.

The inference approach that we build on in this work was developed by Coles and Tawn [1991] and is based on threshold excesses; see Huser et al. [2016] for a detailed review of likelihood estimators for multivariate extremes. The set of extreme events is defined as the set of observations with radial components exceeding a high fixed threshold, that is, the observations belonging to the extreme set,

$$E_{\mathbf{r}} = \left\{ (y_1, \ldots, y_d) \in (0, \infty)^d : \sum_{j=1}^{d} \frac{y_j}{r_j} > 1 \right\},$$

with $\mathbf{r} = (r_1, \ldots, r_d)$ being a large threshold vector. Since the points $n^{-1}\mathbf{Y}^i$ are mapped to the origin for non-extreme observations, the threshold $\mathbf{r}$ needs to be sufficiently large for the Poisson approximation to hold. Note that, $\mathbf{Y}^i \in E_{\mathbf{r}}$, if and only if,

$$R_i = \left\| \mathbf{Y}^i \right\| > \left(\sum_{j=1}^{d} \frac{\omega_{i,j}}{r_j} \right)^{-1}, \quad \text{where } \omega_{i,j} = \frac{Y_j^i}{R_i}.$$

Hence, the expected number of points of the Poisson process P located in the extreme region $E_{\mathbf{r}}$ is

$$\begin{aligned} \mu(E_{\mathbf{r}}) &= \int_{S_d} \int_{\left(\sum_{j=1}^{d} \frac{w_j}{r_j}\right)^{-1}}^{\infty} \frac{\mathrm{d}r}{r^2} \, \mathrm{d}H(\mathbf{w}) \\ &= \int_{S_d} \left(\sum_{j=1}^{d} \frac{w_j}{r_j} \right) \mathrm{d}H(\mathbf{w}) \\ &= \sum_{j=1}^{d} \frac{1}{r_j} \int_{S_d} w_j \, \mathrm{d}H(\mathbf{w}) = \sum_{j=1}^{d} \frac{1}{r_j}. \end{aligned} \tag{2.6}$$

Now, we can explicitly formulate the Poisson log-likelihood over the set E_r,

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) = -\mu(E_{\mathbf{r}}) + \sum_{i=1}^{n_{\mathbf{r}}} \log \left\{ \mu(\mathrm{d}R_i \times \mathrm{d}\mathbf{w}_i) \right\}, \tag{2.7}$$

where $\mathbf{w}_i = (\omega_{i,1}, \ldots, \omega_{i,d})$, $\boldsymbol{\theta}$ represents the p-vector of parameters of the measure μ and $n_{\mathbf{r}}$ represents the number of reindexed observations in the extreme set $E_{\mathbf{r}}$. Using (2.6), the first term in (2.7) can be omitted when maximizing the Poisson log-likelihood, which, using (2.4), boils down to

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n_{\mathbf{r}}} \log \left\{ -V_{1:d}(\mathbf{Y}^i; \boldsymbol{\theta}) \right\}. \tag{2.8}$$

Thanks to the differentiability of the exponent measure V and the support of the angular measure H in the unit simplex S_d, we can use the result of Coles and Tawn [1991, Theorem 1] that relates the angular density to the exponent measure via

$$V_{1:d}(\mathbf{y}; \boldsymbol{\theta}) = -\|\mathbf{y}\|^{-(d+1)} h\left(\frac{y_1}{\|\mathbf{y}\|}, \ldots, \frac{y_d}{\|\mathbf{y}\|}; \boldsymbol{\theta} \right)$$

and reformulate the log-likelihood (2.8) as follows

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) \equiv -(d+1) \sum_{i=1}^{n_{\mathbf{r}}} \log \left\| \mathbf{Y}^i \right\| + \sum_{i=1}^{n_{\mathbf{r}}} \log \left\{ h\left(\frac{Y_1^i}{\|\mathbf{Y}^i\|}, \ldots, \frac{Y_d^i}{\|\mathbf{Y}^i\|}; \boldsymbol{\theta} \right) \right\}. \tag{2.9}$$

2.2.2 Vector generalized additive models for covariate-adjusted angular densities

Our starting point for modelling is an extension of (2.1) to the multivariate setting. Whereas the model in (2.1) is based on indexing the parameters of the univariate extreme value distribution with a regressor, here we index the parameter (H) of a multivariate extreme value distribution (G_H) with a regressor $\mathbf{x} = (x_1, \ldots, x_q)^{\mathrm{T}} \in \mathcal{X} \subset \mathbb{R}^q$. Our target object of interest is thus given by a family of covariate-adjusted angular measures $H_{\mathbf{x}}$ obeying

$$\int_{S_d} w_j \mathrm{d}H_{\mathbf{x}}(\mathbf{w}) = 1, \quad j = 1, \ldots, d.$$

Of particular interest is the setting where $H_{\mathbf{x}}$ is differentiable, in which case the covariate-adjusted angular density can be defined as $h_{\mathbf{x}}(\mathbf{w}) = \mathrm{d}H_{\mathbf{x}}/\mathrm{d}\mathbf{w}$. This yields a corresponding family of covariate-indexed multivariate extreme value distributions

$$G_{\mathbf{x}}(\mathbf{y}) = \exp\left\{ -\int_{S_d} \max\left(\frac{w_1}{y_1}, \ldots, \frac{w_d}{y_d}\right) \mathrm{d}H_{\mathbf{x}}(\mathbf{w}) \right\}.$$

Other natural objects depending on $G_{\mathbf{x}}$ can be readily defined, such as the covariate-adjusted extremal coefficient, $\vartheta_{\mathbf{x}}$, which solves

$$G_{\mathbf{x}}(y\mathbf{1}_d) = \exp(-\vartheta_{\mathbf{x}}/y), \quad y > 0, \tag{2.10}$$

where $\mathbf{1}_d$ is a d-vector of ones. Here, $\vartheta_{\mathbf{x}}$ ranges from 1 to d, and the closer $\vartheta_{\mathbf{x}}$ is to one, the closer we get to the case of complete dependence at that value of the covariate.
Some parametric models [Hüsler and Reiss, 1989, Tawn, 1990, Coles and Tawn, 1991, Cooley et al., 2010] are used below to illustrate the concept of covariate-adjusted angular densities and of covariate-adjusted extremal coefficients, and we focus on the bivariate and trivariate settings for the sake of illustrating ideas. To develop insight and intuition on these models, see Figures 2.1 and 2.2.

Example 1 (Logistic angular surface)**.** Let

$$h_{\mathbf{x}}(w) = (1/\alpha_{\mathbf{x}} - 1)\left\{w(1-w)\right\}^{-1-1/\alpha_{\mathbf{x}}}\left\{w^{-1/\alpha_{\mathbf{x}}} + (1-w)^{-1/\alpha_{\mathbf{x}}}\right\}^{\alpha_{\mathbf{x}}-2}, \quad w \in (0,1),$$

with $\alpha : \mathcal{X} \subset \mathbb{R}^q \to (0,1]$. In Figure 2.1 (left) we represent the case $\alpha_x = \exp\{\eta(x)\}/[1 + \exp\{\eta(x)\}]$, with $\eta(x) = x^2 - 0.5x - 1$ and $x \in \mathcal{X} = [0.1, 2]$. This setup corresponds to be transitioning between a case of relatively high extremal dependence (lower values of x) to a case where we approach asymptotic independence (higher values of x).

Example 2 (Dirichlet angular surface)**.** Let

$$h_{\mathbf{x}}(w) = \frac{\alpha_{\mathbf{x}}\beta_{\mathbf{x}}\Gamma(\alpha_{\mathbf{x}} + \beta_{\mathbf{x}} + 1)(\alpha_{\mathbf{x}}w)^{\alpha_{\mathbf{x}}-1}\{\beta_{\mathbf{x}}(1-w)\}^{\beta_{\mathbf{x}}-1}}{\Gamma(\alpha_{\mathbf{x}})\Gamma(\beta_{\mathbf{x}})\{\alpha_{\mathbf{x}}w + \beta_{\mathbf{x}}(1-w)\}^{\alpha_{\mathbf{x}}+\beta_{\mathbf{x}}+1}}, \quad w \in (0,1),$$

with $\alpha : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$ and $\beta : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$. In Figure 2.1 (middle) we consider the case $\alpha_x = \exp(x)$ and $\beta_x = x^2$, with $x \in [0.9, 3]$. Note the different schemes of extremal dependence induced by the different values of the covariate x as well as the asymmetry of the angular surface underlying this model.

Example 3 (Hüsler–Reiss angular surface)**.** Let

$$h_{\mathbf{x}}(w) = \frac{\lambda_{\mathbf{x}}}{w(1-w)^2(2\pi)^{1/2}} \exp\left\{ -\frac{\left[2 + \lambda_{\mathbf{x}}^2 \log\left\{w/(1-w)\right\}\right]^2}{8\lambda_{\mathbf{x}}^2} \right\}, \quad w \in (0,1),$$

where $\lambda : \mathcal{X} \subset \mathbb{R}^q \to (0, \infty)$. In Figure 2.1 (right) we consider the case $\lambda_x = \exp(x)$, with $x \in [0.1, 2]$. Under this specification, lower values of x correspond to lower levels of extremal dependence, whereas higher values of x correspond to higher levels of extremal dependence.

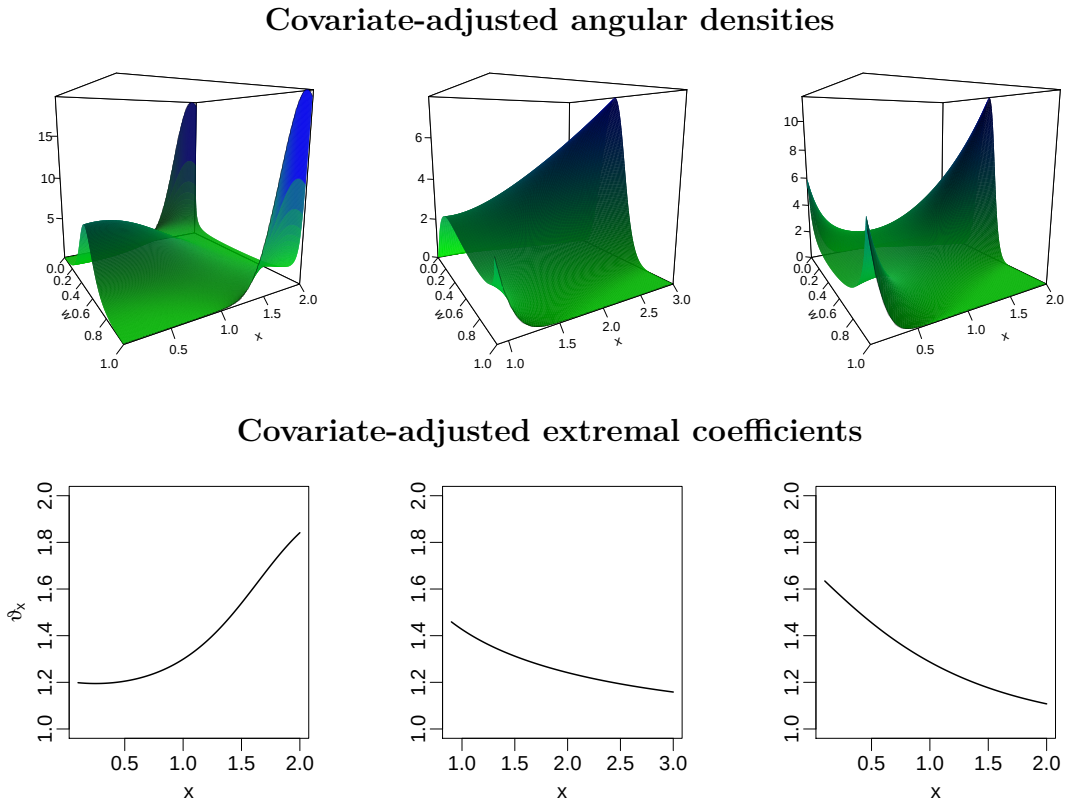

Figure 2.1: Covariate-adjusted angular densities and extremal coefficients of logistic (left panels), Dirichlet (middle panels), and Hüsler–Reiss (right panels) models, corresponding, respectively, to the specifications in Examples 1, 2, and 3.

Example 4 (Pairwise beta angular surface)**.** Let

$$h_{\mathbf{x}}(\mathbf{w}) = \frac{\Gamma(3\alpha_{\mathbf{x}} + 1)}{\Gamma(2\alpha_{\mathbf{x}} + 1)\Gamma(\alpha_{\mathbf{x}})} \sum_{1 \le i < j \le 3} h_{i,j_{\mathbf{x}}}(\mathbf{w}),$$

$$h_{i,j_{\mathbf{x}}}(\mathbf{w}) = (w_i + w_j)^{2\alpha_{\mathbf{x}} - 1} \left\{1 - (w_i + w_j)\right\}^{\alpha_{\mathbf{x}} - 1} \times \frac{\Gamma(2\beta_{i,j_{\mathbf{x}}})}{\Gamma^2(\beta_{i,j_{\mathbf{x}}})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j_{\mathbf{x}}} - 1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j_{\mathbf{x}}} - 1},$$

where $\mathbf{w} = (w_1, w_2, w_3) \in S_3$ and $\alpha, \beta_{i,j} : \mathcal{X} \subset \mathbb{R}^q \to (0, \infty)$ for $1 \le i < j \le 3$. In Figure 2.2, we consider the case $\alpha_{\mathbf{x}} = \exp\{\exp(x)\}$, $\beta_{1,2_{\mathbf{x}}} = \exp(x)$, $\beta_{1,3_{\mathbf{x}}} = x + 1$, and $\beta_{2,3_{\mathbf{x}}} = x + 2$, with $x \in [0.8, 3.3]$. For the different considered values of x, different strengths of global and pairwise dependences can be observed. The mass is concentrated mostly at the center of the simplex due to a large global dependence parameter $\alpha_{\mathbf{x}}$, compared to the pairwise dependence parameters.

The previous parametric models provide some examples of covariate-adjusted angular surfaces $h_{\mathbf{x}}$. But, how can we learn about $h_{\mathbf{x}}$ from the data? Suppose we observe the regression data $\{(\mathbf{x}^i, \mathbf{Y}^i)\}_{i=1}^n$, with $(\mathbf{x}^i, \mathbf{Y}^i) \in \mathcal{X} \times \mathbb{R}^d$, and where we assume that $\mathbf{Y}^i = (Y_1^i, \ldots, Y_d^i)^\top$ are independent random vectors with unit Fréchet marginal distributions

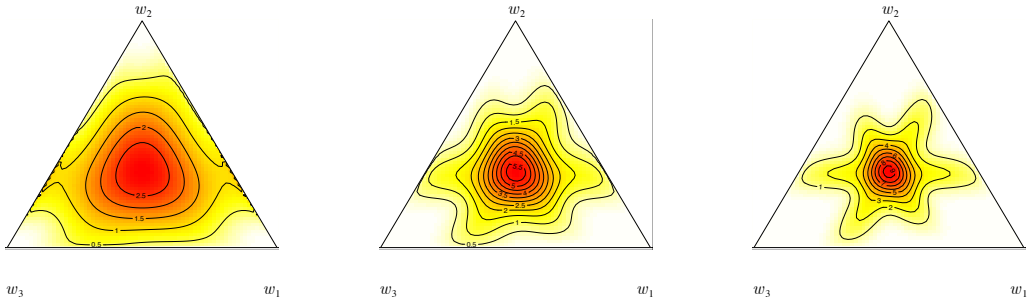

Figure 2.2: Trivariate covariate-adjusted angular density of the pairwise beta model corresponding to the specifications in Example 4 with $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right).

and a joint distribution in the maximum domain of attraction of a multivariate extreme value distribution with angular density $h_{\mathbf{x}_i}$. Using a similar approach as in Section 2.2.1, we convert the raw sample into a pseudo-sample of cardinality $n_{\mathbf{r}}$,

$$\{(\mathbf{x}^i, \mathbf{Y}^i) : \mathbf{Y}^i \in E_{\mathbf{r}}\},$$

and use the latter reindexed data to learn about $h_{\mathbf{x}}$ through the set of angular observations $\{\mathbf{w}_i\}_{i=1}^{n_{\mathbf{r}}}$, where $\mathbf{w}_i = \mathbf{Y}^i / \|\mathbf{Y}^{\mathbf{i}}\|$ for $\mathbf{Y}^i \in E_{\mathbf{r}}$.
Without loss of generality, we restrain ourselves to the bivariate extreme value framework ($d = 2$), so that

$$h_{\mathbf{x}}\left(\frac{Y_1^i}{\|\mathbf{Y}^{\mathbf{i}}\|}, \frac{Y_2^i}{\|\mathbf{Y}^{\mathbf{i}}\|}\right) = h_{\mathbf{x}}\left(w_i, 1 - w_i\right) \equiv h_{\mathbf{x}}(w_i), \quad \text{for } w_i \in [0,1], \quad i = 1, \ldots, n_{\mathbf{r}},$$

that is, the dimension of the angular observations w_i is $M = d - 1 = 1$. We model $h_{\mathbf{x}}(\cdot)$ using $h(\cdot; \boldsymbol{\theta}_{\mathbf{x}})$, where the parameter underlying the dependence structure

$$\begin{aligned} \boldsymbol{\theta}_{\mathbf{x}} &= (\theta_{1\mathbf{x}^1}, \ldots, \theta_{1\mathbf{x}^{n_{\mathbf{r}}}}, \ldots, \theta_{p\mathbf{x}^1}, \ldots \theta_{p\mathbf{x}^{n_{\mathbf{r}}}})^{\mathrm{T}} \in \mathbb{R}^{pn_{\mathbf{r}}}, \\ \mathbf{x} &= (\mathbf{x}^1, \ldots, \mathbf{x}^{n_{\mathbf{r}}})^{\mathrm{T}} \in \mathcal{X}^{n_{\mathbf{r}}} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_q)^{n_{\mathbf{r}}} \subseteq \mathbb{R}^{qn_{\mathbf{r}}}, \end{aligned}$$

is specified through a vector generalized additive model (VGAM) [Yee and Wild, 1996]. Specifically, we model $h_{\mathbf{x}}(w)$ using a fixed family of parametric extremal dependence structures $h(w; \boldsymbol{\theta}_{\mathbf{x}})$ with a covariate-dependent set of parameters $\boldsymbol{\theta}_{\mathbf{x}}$. To learn about $\boldsymbol{\theta}_{\mathbf{x}}$ from the pseudo-sample, we use a vector generalized additive model, which takes the form

$$\boldsymbol{\eta}(\mathbf{x}) \equiv \boldsymbol{\eta} = \mathbf{H}_0 \boldsymbol{\beta}_{[0]} + \sum_{k=1}^{q} \mathbf{H}_k \mathbf{f}_k(\mathbf{x}_k). \tag{2.11}$$

Here,

- $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta}_{\mathbf{x}}) = (g_1(\theta_{1x^1}), \ldots, g_1(\theta_{1x^{n_{\mathbf{r}}}}), \ldots, g_p(\theta_{px^1}), \ldots, g_p(\theta_{px^{n_{\mathbf{r}}}}))^{\top}$ is the vector of predictors and g_l is a link function that ensures that $\theta_{l\cdot}$ is well defined, for $l = 1, \ldots, p$,
- $\boldsymbol{\beta}_{[0]}$ is a $pn_{\mathbf{r}}$-vector of intercepts, with p distinct values each repeated $n_{\mathbf{r}}$ times,
- $\mathbf{x}_k = (x_k^1, \ldots, x_k^{n_{\mathbf{r}}})^{\top} \in \mathcal{X}_k^{n_{\mathbf{r}}}$, for $k = 1, \ldots, q$,

- $\mathbf{f}_k = (\mathbf{f}_{k,1}, \ldots, \mathbf{f}_{k,p})^\top$, where $\mathbf{f}_{k,l} = (f_{k,l}(x_k^1), \ldots, f_{k,l}(x_k^{n_\mathbf{r}}))^\top$, and $f_{k,l} : \mathcal{X}_k \to \mathbb{R}$ are smooth functions supported on $\mathcal{X}_k$, for $k = 1, \ldots, q$ and $l = 1, \ldots, p$, and
- $\mathbf{H}_k$ are $pn_\mathbf{r} \times pn_\mathbf{r}$ constraint matrices, for $k = 0, \ldots, q$.

The constraint matrices $\mathbf{H}_k$ are important quantities in the VGAM (2.11) that allow the tuning of the effects of the covariates on each of the $pn_\mathbf{r}$ components of $\boldsymbol{\eta}$. For example, in Example 4, one might want to impose the same smooth effect of a covariate on each of the $\binom{3}{2}$ pairwise dependence parameters and at the same time restrict the effect of this covariate to be zero on the global dependence parameter. To avoid clutter in the notation, we assume from now on that $\mathbf{H}_k = \mathbf{I}_{pn_\mathbf{r} \times pn_\mathbf{r}}$, for $k = 0, \ldots, q$.

The smooth functions $f_{k,l}$ are written as linear combinations of B-spline basis functions

$$f_{k,l}(x_k^i) = \sum_{s=1}^{d_k} \beta_{[kl]_s} B_{s,\tilde{q}}(x_k^i), \quad k = 1, \ldots, q, \quad l = 1, \ldots, p, \quad i = 1, \ldots, n_\mathbf{r},$$

where $B_{s,\tilde{q}}$ is the sth B-spline of order $\tilde{q}$ and $d_k = \tilde{q} + m_k$, with m_k the number of internal equidistant knots for $\mathbf{x}_k$ [Yee, 2015, Section 2.4.5]. To ease the notational burden, we suppose without loss of generality that $d_k \equiv \tilde{d}$, for $k = 1, \ldots, q$, and define

$$\boldsymbol{\beta}_{[k]} = \left(\beta_{[k1]_1}, \ldots, \beta_{[k1]_{\tilde{d}}}, \ldots, \beta_{[kp]_1}, \ldots, \beta_{[kp]_{\tilde{d}}}\right)^\top \in \mathbb{R}^{\tilde{d}p}.$$

Therefore, the VGAM (2.11), with identity constraint matrices $\mathbf{H}_k$, can be written as

$$\boldsymbol{\eta} = \boldsymbol{\beta}_{[0]} + \sum_{k=1}^{q} \mathbf{X}_{[k]} \boldsymbol{\beta}_{[k]} = \mathbf{X}_{\text{VAM}} \boldsymbol{\beta}, \tag{2.12}$$

where

$$\begin{cases} \boldsymbol{\beta} = \left(\boldsymbol{\beta}_{[0]} \quad \boldsymbol{\beta}_{[1]} \quad \cdots \quad \boldsymbol{\beta}_{[q]}\right)^\top \in \mathbf{B} \subset \mathbb{R}^{p(1+q\tilde{d})}, \\ \mathbf{X}_{\text{VAM}} = \left(\mathbf{1}_{pn_\mathbf{r} \times p} \quad \mathbf{X}_{[1]} \quad \cdots \quad \mathbf{X}_{[q]}\right) \in \mathbb{R}^{pn_\mathbf{r} \times \{p(1+q\tilde{d})\}}, \end{cases}$$

for some $pn_\mathbf{r} \times \tilde{d}p$ submatrices $\mathbf{X}_{[k]}$, $k = 1, \ldots, q$. The vector of parameters to be estimated in the VGAM (2.12) is $\boldsymbol{\beta}$.

The specification in (2.12) makes it possible to simultaneously fit ordinary generalized additive models [Wood, 2017] in each component of the vector of parameters $\boldsymbol{\theta}_\mathbf{x}$, hence avoiding any non orthogonality-related issues that could arise if the p components were to be treated separately [Chavez-Demoulin and Davison, 2005]. Finally, if the dimension M of the response vector of angular observations w_i is greater than one ($d > 2$), then the vector of predictors $\boldsymbol{\eta}$ will instead be a $Mpn_\mathbf{r}$-vector and the dimensions of the related quantities in (2.12) will change accordingly.

To give the unfamiliar reader insight on some of the quantities introduced above, we identify these quantities in the examples mentioned previously:

- In Examples 1 and 3, $d = 2$, $M = 1$, $p = 1$, $q = 1$, and $\mathcal{X} = [0.1, 2]$. The difference between the VGAMs modeled in these two examples resides in the form of dependence of η on x and the link function g. In Example 1, the parameter $\theta_x \in (0, 1]$, $\eta = x^2 - 0.5x - 1$, and the link function g is the logit function, whereas in Example 3 the parameter $\theta_x \in (0, \infty)$, $\eta = x$, and the link function g is the logarithm function.

- In Example 2, $d = 2$, $M = 1$, $p = 2$, $q = 1$, $\mathcal{X} = [0.9, 3]$, and $\boldsymbol{\eta} = (x, x)^\top$. The vector of parameters for the bivariate Dirichlet angular density $\boldsymbol{\theta}_x \in (0, \infty)^2$ and the link functions g_1 and g_2 are the logarithm and the square root functions, respectively.

- In Example 4, $d = 3$, $M = 2$, $p = 4$, $q = 1$, $\mathcal{X} = [0.8, 3.3]$, and $\boldsymbol{\eta} = (\exp(x), x, \log(x+1), \log(x+2))^\top$. The vector of parameters for the pairwise beta angular density $\boldsymbol{\theta}_x \in (0, \infty)^4$ and the link function g_l is the logarithm function, for $l = 1, \ldots, 4$.

2.3 Inference and asymptotic properties

The log-likelihood (2.9) with a covariate-dependent vector of parameters $\boldsymbol{\theta}_{\mathbf{x}}$ is now written as

$$\begin{aligned} \ell(\boldsymbol{\beta}) &:= \sum_{i=1}^{n_{\mathbf{r}}} c_i + \log\left\{h\left(\mathbf{w}_i; \boldsymbol{\beta}\right)\right\}, \\ &= \sum_{i=1}^{n_{\mathbf{r}}} c_i + \log\left(h\left[\mathbf{w}_i; \mathbf{g}^{-1}\{\boldsymbol{\eta}(\mathbf{x}^i)\}\right]\right), \end{aligned} \tag{2.13}$$

where c_i is a constant independent of $\boldsymbol{\beta}$ and $\mathbf{g}^{-1}$ is the componentwise inverse of $\mathbf{g}$.

Incorporating a covariate-dependence in the extremal dependence model through a non-linear smooth model adds considerable flexibility in the modelling of the dependence parameter $\boldsymbol{\theta}_{\mathbf{x}}$. The price to pay for this flexibility is reflected in the estimation procedure. The estimation of $\boldsymbol{\theta}_{\mathbf{x}}$, hence of $\boldsymbol{\beta}$, is performed by maximizing the penalized log-likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \ell(\boldsymbol{\beta}) - \frac{1}{2}\mathbf{J}(\boldsymbol{\gamma}), \tag{2.14}$$

where the penalty term can be written as

$$\mathbf{J}(\boldsymbol{\gamma}) = \sum_{k=1}^{q} \boldsymbol{\beta}_{[k]}^\top \left\{\mathbf{P}_k \otimes \operatorname{diag}(\gamma_{(1)k}, \ldots, \gamma_{(p)k})\right\} \boldsymbol{\beta}_{[k]} = \boldsymbol{\beta}^\top \mathbf{P}(\boldsymbol{\gamma}) \boldsymbol{\beta},$$

with $\mathbf{P}(\boldsymbol{\gamma})$ a $p(1 + q\tilde{d}) \times p(1 + q\tilde{d})$ block matrix with a first $p \times p$ block filled with zeros and q blocks, each formed by a $p\tilde{d} \times p\tilde{d}$ matrix $\mathbf{P}_k$ that depends only on the knots of the B-spline functions for the covariate $\mathbf{x}_k$. The matrix $\mathbf{P}(\boldsymbol{\gamma})$ can be written as $\mathbf{P}(\boldsymbol{\gamma}) = \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$ for some $p(1 + q\tilde{d}) \times p(1 + q\tilde{d})$ real matrix $\tilde{\mathbf{X}}$. The vectors $\boldsymbol{\beta}_{[k]}$ are defined in (2.12), and $\gamma_{(l)k}$ are termed the smoothing parameters.
The penalty term in (2.14) controls the wiggliness and the fidelity to the data of the component functions in (2.11) through the vector $\boldsymbol{\gamma}$ of the smoothing parameters $\gamma_{(l)k}$ for $l = 1, \ldots, p$ and $k = 1, \ldots, q$. Larger values of $\gamma_{(l)k}$ lead to smoother effects of the covariate $\mathbf{x}_k$ on the lth component of $\boldsymbol{\eta}$.

The maximization of the penalized log-likelihood (2.14) is based on a Newton–Raphson (N–R) algorithm. At each step of the N–R algorithm, a set of smoothing parameters is proposed by outer iteration [Wood, 2017], and a penalized iterative reweighted least squares (PIRLS) algorithm is performed, in an inner iteration, to update the model coefficients estimates. We detail the inner fitting procedure in the following section and the outer iteration in Section 2.3.2.

2.3.1 Fitting algorithm

We suppose that the penalized log-likelihood (2.14) depends only on the $p(1+q\tilde{d})$-vector $\boldsymbol{\beta}$ and that the vector of smoothing parameters $\boldsymbol{\gamma}$ is proposed (at each iteration of the N–R algorithm) by outer iteration and is therefore fixed in what follows.

The penalized maximum log-likelihood estimator (PMLE) $\hat{\boldsymbol{\beta}}$ satisfies the following score equation

$$\frac{\partial \ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} = \mathbf{X}_{\text{VAM}}^{\top} \mathbf{u}(\hat{\boldsymbol{\beta}}) - \mathbf{P}(\boldsymbol{\gamma})\hat{\boldsymbol{\beta}} = \mathbf{0},$$

where $\mathbf{u}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta})/\partial \boldsymbol{\eta} \in \mathbb{R}^{pn_{\mathbf{r}}}$ and $\mathbf{X}_{\text{VAM}}$ is as defined in (2.12). To obtain $\hat{\boldsymbol{\beta}}$, we update $\boldsymbol{\beta}^{(a-1)}$, the $(a-1)$th estimate of the true $\boldsymbol{\beta}_0$, by Newton–Raphson:

$$\boldsymbol{\beta}^{(a)} = \boldsymbol{\beta}^{(a-1)} + \mathbf{I}\left(\boldsymbol{\beta}^{(a-1)}\right)^{-1} \left\{ \mathbf{X}_{\text{VAM}}^{\top} \mathbf{u}(\boldsymbol{\beta}^{(a-1)}) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta}^{(a-1)} \right\}, \tag{2.15}$$

where

$$\begin{cases} \mathbf{I}\left(\boldsymbol{\beta}^{(a-1)}\right) = -\dfrac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\top}} = \mathbf{X}_{\text{VAM}}^{\top} \mathbf{W}(\boldsymbol{\beta}^{(a-1)}) \mathbf{X}_{\text{VAM}} + \mathbf{P}(\boldsymbol{\gamma}), \\ \mathbf{W}(\boldsymbol{\beta}^{(a-1)}) = -\dfrac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}} \in \mathbb{R}^{pn_{\mathbf{r}} \times pn_{\mathbf{r}}}. \end{cases}$$

The matrix $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$ is termed the working weight matrix. If the expectation $\mathrm{E}\{\partial^2 \ell(\boldsymbol{\beta})/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}\}$ is obtainable, a Fisher scoring algorithm is then preferred, as it ensures the positive definiteness of $\mathbf{W}(\boldsymbol{\beta})$ over a larger region of the parameter space $\mathbf{B}$ than in the N–R algorithm. When the working weight matrix is not positive definite, which might happen when the parameter $\boldsymbol{\beta}^{(a-1)}$ is far from the true $\boldsymbol{\beta}_0$, a Greenstadt [Greenstadt, 1967] modification is applied, and the negative eigenvalues of $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$ are replaced by their absolute values. With the different families of angular densities considered in Examples 1, 2, and 4, the expected information matrix is not obtainable and is hence replaced by the observed information matrix on which a Greenstadt modification is applied whenever needed. See Yee [2015, Section 9.2] for other remedies and techniques for deriving well-defined working weight matrices.

Let $\mathbf{z}^{(a-1)} := \mathbf{X}_{\text{VAM}} \boldsymbol{\beta}^{(a-1)} + \mathbf{W}(\boldsymbol{\beta}^{(a-1)})^{-1} \mathbf{u}(\boldsymbol{\beta}^{(a-1)})$ be the $pn_{\mathbf{r}}$-vector of working responses. Then, (2.15) can be rewritten in a PIRLS form as

$$\begin{aligned} \boldsymbol{\beta}^{(a)} &= \left\{ \mathbf{X}_{\text{VAM}}^{\top} \mathbf{W}(\boldsymbol{\beta}^{(a-1)}) \mathbf{X}_{\text{VAM}} + \mathbf{P}(\boldsymbol{\gamma}) \right\}^{-1} \mathbf{X}_{\text{VAM}}^{\top} \mathbf{W}(\boldsymbol{\beta}^{(a-1)}) \mathbf{z}^{(a-1)} \\ &= \left\{ \mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)} \mathbf{X}_{\text{PVAM}} \right\}^{-1} \mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)} \mathbf{y}^{(a-1)}, \end{aligned}$$

where $\mathbf{X}_{\text{PVAM}}$, $\mathbf{y}^{(a-1)}$, and $\tilde{\mathbf{W}}^{(a-1)}$ are augmented versions of $\mathbf{X}_{\text{VAM}}$, $\mathbf{z}^{(a-1)}$ and $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$, respectively, and are defined as

$$\begin{cases} \mathbf{X}_{\text{PVAM}} = \begin{pmatrix} \mathbf{X}_{\text{VAM}}^{\top} & \tilde{\mathbf{X}} \end{pmatrix}^{\top} \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d}) \times p(1+q\tilde{d})}, \\ \mathbf{y}^{(a-1)} = \begin{pmatrix} \mathbf{z}^{(a-1)} & \mathbf{0}_{p(1+q\tilde{d})} \end{pmatrix}^{\top} \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d})}, \\ \tilde{\mathbf{W}}^{(a-1)} = \operatorname{diag}\left(\mathbf{W}(\boldsymbol{\beta}^{(a-1)}), \mathbf{I}_{p(1+q\tilde{d}) \times p(1+q\tilde{d})} \right) \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d}) \times p(1+n_{\mathbf{r}}+q\tilde{d})}. \end{cases}$$

The algorithm stops when the change in the coefficients $\boldsymbol{\beta}$ between two successive iterations is sufficiently small. Convergence of the N–R algorithm is not guaranteed and might not occur if the quadratic approximation of $\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ around $\hat{\boldsymbol{\beta}}$ is poor. See Yee [2015, 2016] for more details.

The plug-in penalized maximum log-likelihood estimator of the covariate-dependent angular density is defined as

$$\widehat{h}_{\mathbf{x}}(\mathbf{w}) \equiv h\{\mathbf{w}; \mathbf{g}^{-1}(\mathbf{X}_{\text{VAM}}\hat{\boldsymbol{\beta}})\}. \tag{2.16}$$

In the following section, we give details about the selection of the smoothing parameters $\boldsymbol{\gamma}$, which is outer to the PIRLS algorithm.

2.3.2 Selection of the smoothing parameters

To implement the PIRLS algorithm performed at each iteration of the N–R algorithm, a smoothing parameter selection procedure is conducted by minimizing a prediction error estimate given by the generalized cross validation (GCV) score.

Let $\mathbf{A}^{(a-1)}(\boldsymbol{\gamma})$ be the influence matrix of the fitting problem at the ath iteration, defined as

$$\mathbf{A}^{(a-1)}(\boldsymbol{\gamma}) = \mathbf{X}_{\text{PVAM}} \left\{\mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)} \mathbf{X}_{\text{PVAM}}\right\}^{-1} \mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)}.$$

Then, by minimizing the GCV score

$$\text{GCV}^{(a-1)} = \frac{n_{\mathbf{r}} \left\{\mathbf{y}^{(a-1)} - \mathbf{A}^{(a-1)}(\boldsymbol{\gamma})\mathbf{y}^{(a-1)}\right\}^{\top} \tilde{\mathbf{W}}^{(a-1)} \left\{\mathbf{y}^{(a-1)} - \mathbf{A}^{(a-1)}(\boldsymbol{\gamma})\mathbf{y}^{(a-1)}\right\}}{\left[n_{\mathbf{r}} - \text{trace}\left\{\mathbf{A}^{(a-1)}(\boldsymbol{\gamma})\right\}\right]^2},$$

we aim at balancing between goodness of fit and complexity of the model, which is measured by the trace of the influence matrix and termed the effective degrees of freedom (EDF). The EDF of the fitted VGAM (2.12) are defined as the EDF obtained at convergence, that is, trace $\left\{\mathbf{A}^{(c-1)}(\boldsymbol{\gamma})\right\}$, where c is the iteration at which convergence occurs.

Both the fitting algorithm of Section 2.3.1 and the smoothing parameter selection are implemented in the `R` package `VGAM` [Yee, 2017], with the latter being required from the `R` package `mgcv` [Wood, 2017].

Model selection between different, not necessarily nested, fitted VGAMs is performed based on the Akaike information criterion (AIC), where the number of parameters of the model is replaced by its EDF to account for penalization. More details on the (conditional) AIC for models with smoothers along with a corrected version of this criterion, which takes into account the smoothing parameter uncertainty, can be found in Wood [2017, Section 6.11].

2.3.3 Large sample properties

We now derive the consistency and asymptotic normality of the PMLE $\hat{\boldsymbol{\beta}}$ defined in Section 2.3.1. Throughout this section, we assume that the set of angular observations $\{\mathbf{w}_i\}_{i=1}^{n_{\mathbf{r}}}$ defined in the log-likelihood (2.13), stems from a parametric family of angular densities of multivariate extreme value distributions $\{h(\cdot, \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathbf{B}\}$, with $\mathbf{B}$ a compact subset of $\mathbb{R}^{p(1+q\tilde{d})}$. The following large sample properties are derived for an increasing size $n_{\mathbf{r}}$ and under the assumption that the angular density $h(\cdot; \boldsymbol{\beta})$ has all its density in the interior of the unit simplex S_d. Additionally, we assume that the dependence parameter of the angular density is a linear combination of known spline functions where the number and the location of the internal knots are fixed, that is, the dimension of the parameter $\boldsymbol{\beta}$ is fixed. The bias resulting from this assumption is negligible compared to the bias due

to the smoothing parameter uncertainty [Ruppert, 2002].
Based on the penalized log-likelihood (2.14), $\hat{\boldsymbol{\beta}}$ satisfies the following score equation

$$\mathbf{m}(\boldsymbol{\beta}) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta} = \mathbf{0}_{p(1+q\tilde{d})}, \tag{2.17}$$

where $\mathbf{m}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$.

Let $\mathbf{B}_0$ be an open neighbourhood around the true parameter $\boldsymbol{\beta}_0$. Moreover, we define $\mathbf{m}(\mathbf{w}, \boldsymbol{\beta}) = \partial \log\{h(\mathbf{w}; \boldsymbol{\beta})\}/\partial \boldsymbol{\beta}$.
Our asymptotic results hold under the following customary assumptions on the smoothing parameters $\boldsymbol{\gamma}$ and the angular density $h(\cdot; \boldsymbol{\beta})$:

(A1) $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_{(1)1} & \cdots & \gamma_{(p)1} & \cdots & \gamma_{(1)q} & \cdots & \gamma_{(p)q} \end{pmatrix}^\top = o(n_{\mathbf{r}}^{1/2})\mathbf{1}_{pq}$.

(A2) Regularity conditions:

- The angular density support $\{\mathbf{w} \in S_d : h(\mathbf{w}; \boldsymbol{\beta}) > 0\}$ does not depend on $\boldsymbol{\beta} \in \mathbf{B}_0$.
- If $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, then $h(\mathbf{w}; \boldsymbol{\beta}) \neq h(\mathbf{w}; \boldsymbol{\beta}_0)$, with $\boldsymbol{\beta} \in \mathbf{B}_0$. Moreover, $\mathrm{E}[\sup_{\boldsymbol{\beta} \in \mathbf{B}} |\log\{h(\mathbf{w}; \boldsymbol{\beta})\}|] < \infty$.
- For $\mathbf{w} \in S_d$, $h(\mathbf{w}; \boldsymbol{\beta}) \in C^3(\mathbf{B})$ and $h(\mathbf{w}; \boldsymbol{\beta}) > 0$ on $\mathbf{B}_0$.
- $\int \sup_{\boldsymbol{\beta} \in \mathbf{B}} \|\mathbf{m}(\mathbf{w}, \boldsymbol{\beta})\| \, \mathrm{d}\mathbf{w} < \infty$ and $\int \sup_{\boldsymbol{\beta} \in \mathbf{B}} \|\partial \mathbf{m}(\mathbf{w}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^\top\| \mathrm{d}\mathbf{w} < \infty$.
- For $\boldsymbol{\beta} \in \mathbf{B}_0$, $\mathbf{i}(\boldsymbol{\beta}) := \mathrm{cov}\{\mathbf{m}(\mathbf{W}, \boldsymbol{\beta})\} = \mathbf{X}_{\mathrm{VAM}}^\top \mathbf{W}(\boldsymbol{\beta}) \mathbf{X}_{\mathrm{VAM}}$ exists and is positive-definite.
- For each triplet $1 \leq q, r, s \leq p(1 + q\tilde{d})$, there exists a function $M_{qrs} : S_d \to \mathbb{R}$ such that, for $\mathbf{w} \in S_d$ and $\boldsymbol{\beta} \in \mathbf{B}_0$, $|\partial^3 \log\{h(\mathbf{w}; \boldsymbol{\beta})\}/\partial \boldsymbol{\beta}_{qrs}| \leq M_{qrs}(\mathbf{w})$, and $\mathrm{E}\{M_{qrs}(\mathbf{W})\} < \infty$.

Assumption (A1) is needed to control the influence of the smoothing parameters as $n_{\mathbf{r}} \to \infty$ such that the asymptotic unbiasedness of $\hat{\boldsymbol{\beta}}$ can be established. This assumption is rather weak as it allows the smoothing parameters to grow as the size of the threshold exceedances grows, at a rate smaller than $n_{\mathbf{r}}^{1/2}$, implying therefore that heavy oversmoothing is avoided. As discussed by Marra and Wood [2012], heavy oversmoothing is likely to happen under two situations. The first one corresponds to situations where two covariates are highly correlated but the effect of one is very smooth and the effect of the other is very wiggly, leading therefore to a possible inversion of the degrees of smoothness and to one of the covariates being highly oversmoothed. The second situation corresponds to the setting where the true effect of one of the covariates is close to a function in the null space of the penalty associated to this covariate, i.e., a straight line, and might therefore be estimated exactly as this function. Assumption (A2) consists of standard regularity assumptions under which the classical asymptotic properties of MLEs hold. These assumptions are expressed in terms of the angular density as the first term in the log-likelihood (2.9) does not influence the maximization. Although difficult to verify, these assumptions are similar to the ones considered in Bienvenüe and Robert [2017] and Padoan et al. [2010] where in the latter, the assumptions are imposed on the components of the composite likelihood along with additional assumptions on the composite score equation. Note that Dombry et al. [2016] discussed the asymptotic properties of the maximum likelihood estimator of max-stable distributions under the notion of differentiability in quadratic mean. Their

regularity conditions on the exponent function and angular measure were proved to hold for several popular parametric models.
The next theorem characterizes the large sample behaviour of our estimator $\hat{\boldsymbol{\beta}}$.

Theorem 2.3.1. *Let* $\{h(\cdot, \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathbf{B}\}$ *be a parametric family of angular densities of multivariate extreme value distributions, where* $\mathbf{B}$ *is a compact subset of* $\mathbb{R}^{p(1+q\tilde{d})}$. *Let* $\boldsymbol{\beta}_0$ *be an interior point of* $\mathbf{B}$ *and* $\mathbf{B}_0$ *an open neighbourhood around* $\boldsymbol{\beta}_0$. *Under (A1) and (A2), and for independent observations* $\mathbf{w}_1, \ldots, \mathbf{w}_{n_\mathbf{r}}$ *with distribution* $h(\cdot, \boldsymbol{\beta}_0)$, *it follows that, as* $n_\mathbf{r} \to \infty$, *the estimator* $\hat{\boldsymbol{\beta}}$ *maximizing the penalized log-likelihood* (2.14) *verifies:*

1. $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n_\mathbf{r}^{-1/2})$.
2. $n_\mathbf{r}^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{d}} N(\mathbf{0}, \mathbf{i}(\boldsymbol{\beta}_0)^{-1})$.

These results are derived from a second-order Taylor expansion of the score equation (2.17) around the true parameter $\boldsymbol{\beta}_0$ along the same lines as in Vatter and Chavez-Demoulin [2015] and Davison [2003, p. 147]. The proof of Theorem 2.3.1 is deferred to the Supplementary Materials of Section 2.7. Similar results on the large sample behaviour of the corresponding plug-in estimator (2.16) can be derived using the multivariate delta method. These results are useful to derive and construct approximate confidence intervals for conditional angular densities and to compare nested models based on likelihood ratio tests. Our proviso is similar to that of de Carvalho and Davison [2014] in the sense that asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}$ are derived under the assumption of known margins and we sample from the limiting object $h_\mathbf{x}$, whereas in practice only a sample of (estimated) pseudo-angles, $\{\widehat{\mathbf{w}}_i\}_{i=1}^{n_\mathbf{r}}$, would be available and the uncertainty arising from the marginal fitting would not be accounted for. Asymptotic properties under misspecification of the parametric model set for $h_\mathbf{x}$ could in principle be derived under additional assumptions on $\boldsymbol{\beta}$ and $\mathbf{m}$, along the same lines as in standard likelihood theory [Knight, 2000]. The resulting theory is outside the scope of this work and is deliberately not studied here.

2.4 Simulation study

2.4.1 Data generating processes and preliminary experiments

We assess the performance of our methods using the bivariate extremal dependence structures presented in Section 2.2.2—and displayed in Figure 2.1—as well as the trivariate pairwise beta dependence model from Example 4—depicted in Figure 2.2. Monte Carlo evidence will be reported in Section 2.4.2. For now, we concentrate on illustrating the methods over a single-run experiment on these scenarios. For each dependence model from Examples 1–3, we draw a sample $\{(w_{i,1}, w_{i,2})\}_{i=1}^{n_\mathbf{r}}$ from the corresponding angular density h_x with sample size $n_\mathbf{r} = 300$ and where each angular observation $(w_{i,1}, w_{i,2})$ is drawn from the chosen dependence model conditional on a fixed value x^i of the covariate x. To gain insight into the bias and variance of our covariate-adjusted spectral density estimator, we compute its 95% asymptotic confidence bands based on Theorem 2.3.1 and at different values of w in $(0, 1)$. The only source of bias in our estimation procedure is due to the penalization of the model likelihood causing a smoothing bias [Wood, 2017] if the smoothing parameters do not vanish at a certain rate (see Section 2.3.3). The uncertainty due to the choice of the parametric model is deliberately not taken into account, that is, the simulations are performed in a well-specified framework.

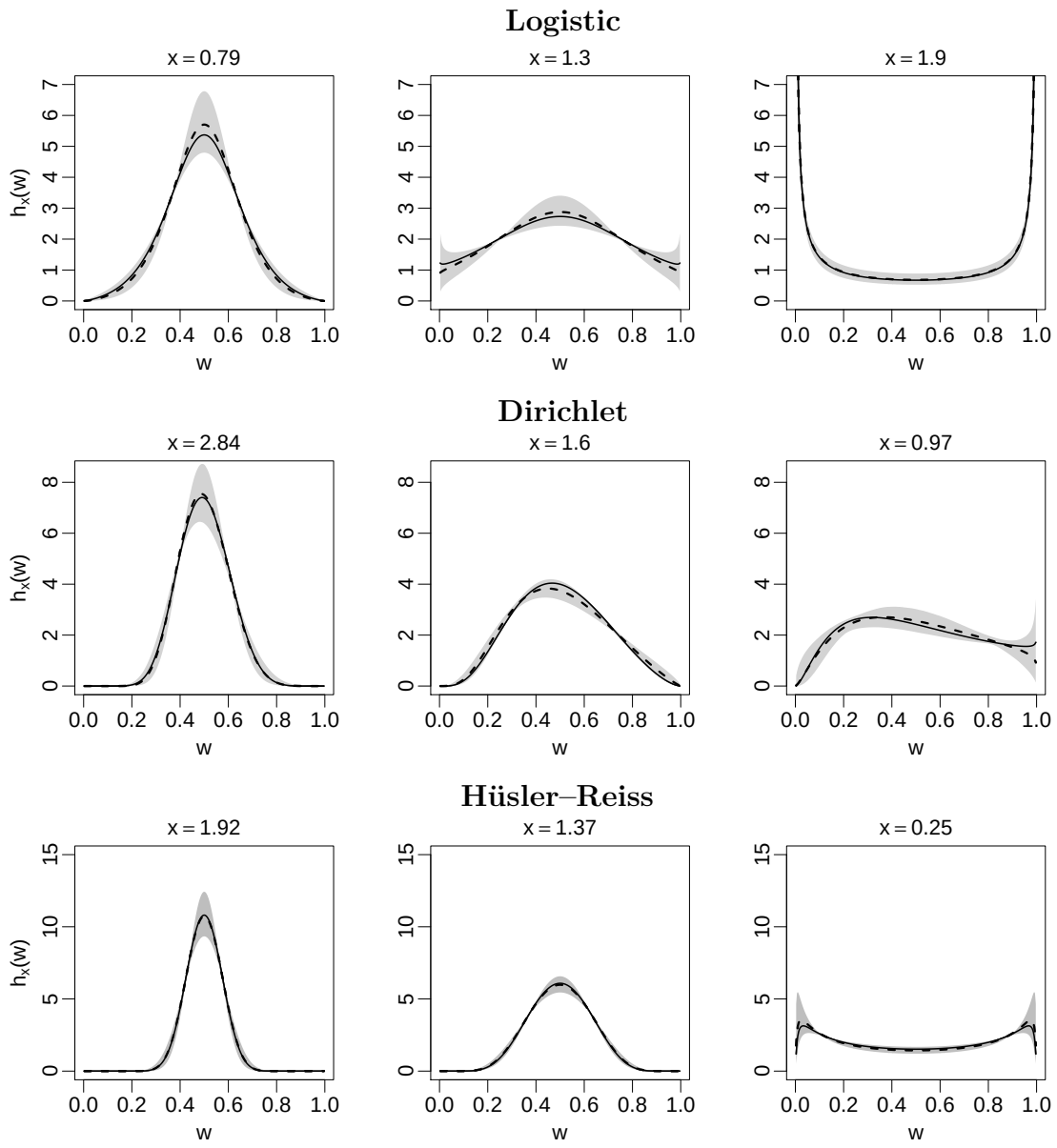

Figure 2.3: Estimates of the covariate-adjusted spectral densities in Examples 1, 2, and 3 conditional on different values of the covariate x (dashed lines) along with their 95% (pointwise) asymptotic confidence bands (grey area). The true spectral densities are displayed in solid lines.

Figure 2.3 displays the estimates of the covariate-adjusted spectral densities from Examples 1, 2, and 3 for various fixed values of the covariate x that induce different extremal dependence strengths. All panels show that for the different extremal dependence schemes (strength and asymmetry), the covariate-adjusted spectral densities are accurately estimated and the true curves fall well within the 95% confidence bands. The estimates in the Dirichlet case seem to be a bit more biased, and this might be explained by the fact that both of the two non-orthogonal parameters of the model depend smoothly on the covariate x.

We now consider the case of the trivariate pairwise beta dependence model from Example 4. We draw a sample $\{(w_{i,1}, w_{i,2}, w_{i,3})\}_{i=1}^{n_{\mathbf{r}}}$ with sample size $n_{\mathbf{r}} = 300$ where each observation $(w_{i,1}, w_{i,2}, w_{i,3})$ is drawn from the pairwise beta model conditional on a fixed

value x_i of the covariate x, as illustrated in Figure 2.2. Figure 2.4 displays the contour plots of the estimates of the covariate-adjusted spectral density from Example 4 at three fixed values of x. All panels in Figure 2.4 show that, for the different extremal dependence

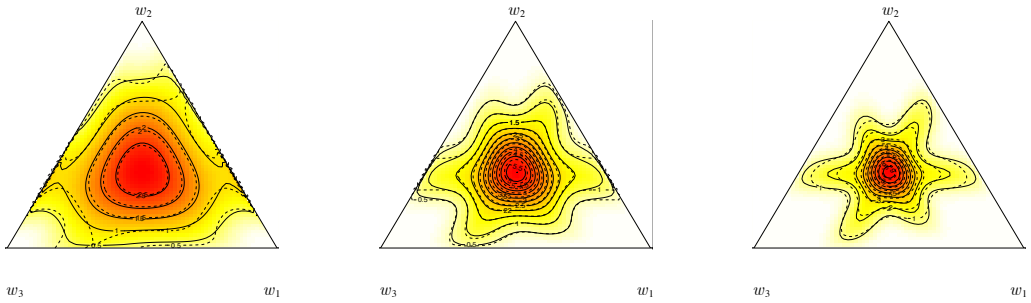

Figure 2.4: Contour plots of the covariate-adjusted pairwise beta spectral density estimate (dashed lines) at $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right). The contour plots of the true spectral density are displayed in solid lines.

schemes, i.e., for the different considered values of x, the contour plots of the estimates are remarkably close to the actual contour plots. The estimates are slightly more biased near the edges of the simplex than in the center, reflecting a better estimation of the global dependence parameter compared to the pairwise dependence parameters.

2.4.2 Monte Carlo evidence

A Monte Carlo study was conducted by simulating 500 independent samples of sizes $n_{\mathbf{r}} = 300$ and $n_{\mathbf{r}} = 500$ angular observations, respectively. In what follows we focus on documenting how the level of accuracy increases when the number of observations increases by assessing the mean integrated absolute error (MIAE)—which for the bivariate case can be written as

$$\mathrm{MIAE} = \mathrm{E}\left\{\int_{\mathcal{X}}\int_0^1 |\widehat{h}_x(w) - h_x(w)|\,\mathrm{d}w\,\mathrm{d}x\right\}.$$

The results are reported in Table 2.1.

Table 2.1: Mean integrated absolute error (MIAE) estimates computed from 500 samples for the covariate-adjusted spectral densities in Examples 1–3; $n_{\mathbf{r}}$ denotes the number of angular observations

$n_{\mathbf{r}}$	Covariate-adjusted angular density	MIAE
300	Logistic	0.2185
	Dirichlet	0.3372
	Hüsler–Reiss	0.2205
500	Logistic	0.1781
	Dirichlet	0.2459
	Hüsler–Reiss	0.1648

As expected, an increase in the number of angular observations leads to a reduction of MIAE. To give a more granular level of detail than that of Table 2.1 on the behaviour

of the estimator over specific values of the covariate and of the unit simplex, Figure 2.5 displays the Monte Carlo confidence intervals of the covariate-adjusted spectral densities from Examples 1–3 for various fixed values of the covariate x, along with the Monte Carlo means and Figure 2.6 displays the contour plots of the Monte Carlo mean of the covariate-adjusted spectral density from Example 4 at three fixed values of x.

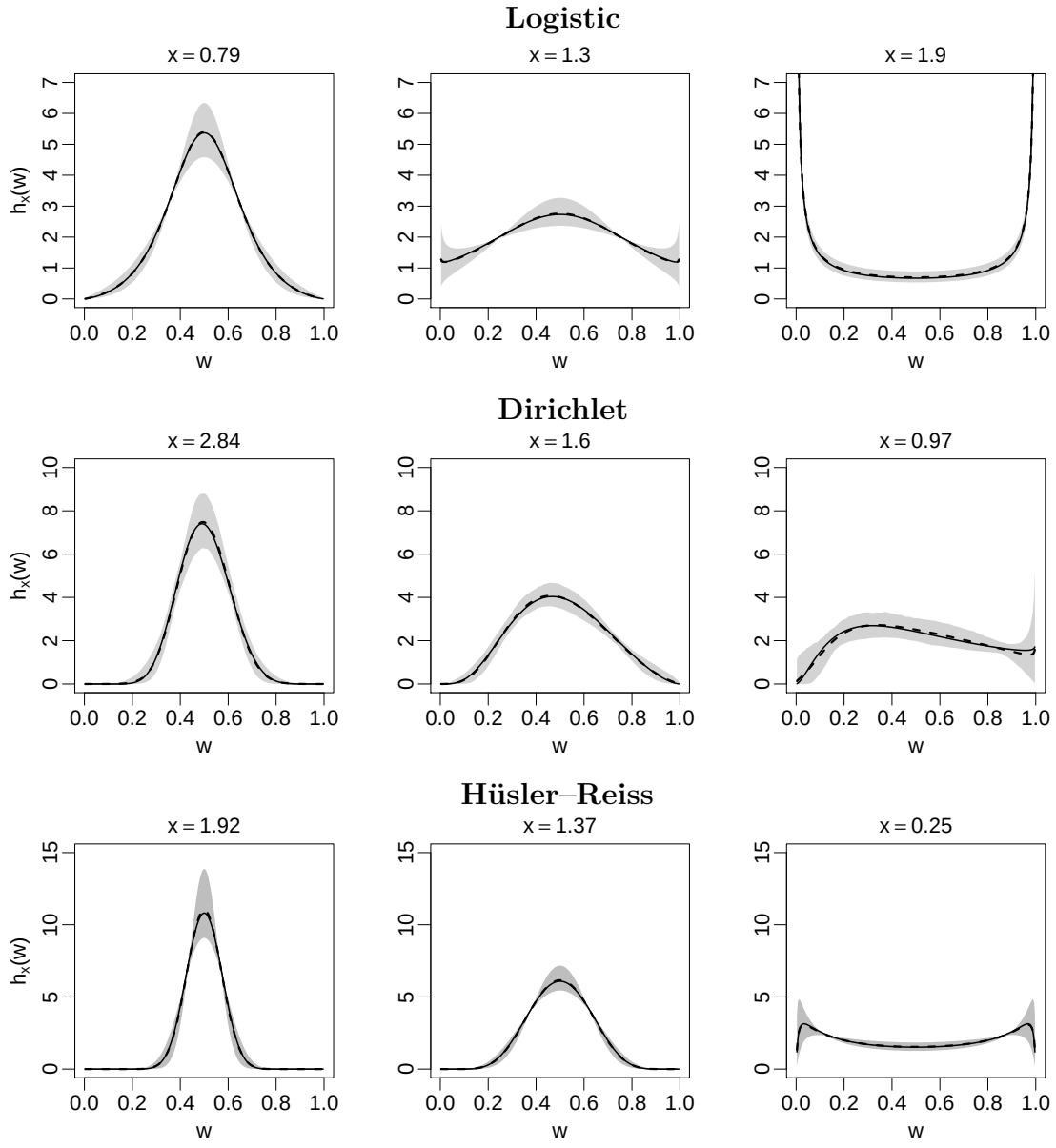

Figure 2.5: The Monte Carlo 95% confidence intervals of the spectral densities in Examples 1, 2, and 3 conditional on different values of the covariate x (grey area) along with their Monte Carlo means (dashed lines). The true spectral densities are displayed in solid lines.

As can be seen from Figures 2.5 and 2.6, our method successfully recovers the corresponding target covariate-adjusted angular densities with a high level of precision over the simulation study. Additionally, the variability in the Monte Carlo study in Figure 2.5 is comparable to the asymptotic variability displayed in Figure 2.3.

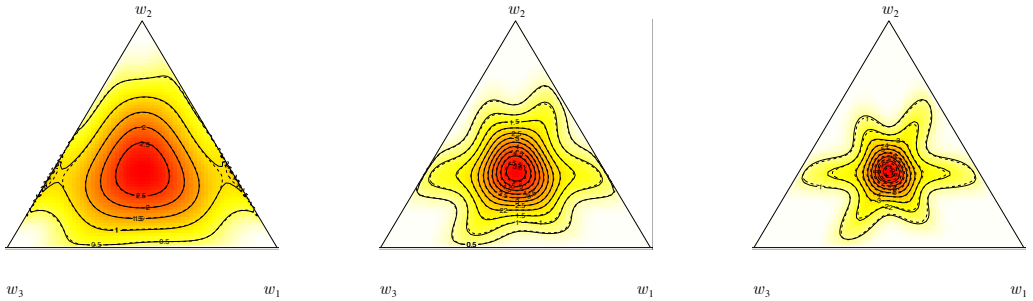

Figure 2.6: Contour plots of the Monte Carlo mean estimate of the covariate-adjusted pairwise beta spectral density (dashed lines) at $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right). The contour plots of the true spectral density are displayed in solid lines.

2.5 Extreme temperature analysis

2.5.1 Data description, motivation for the analysis, and preprocessing

In this section, we describe an application to modelling the dependence between extreme air winter (December–January–February) temperatures at two sites in the Swiss Alps: Montana—at an elevation of 1427m—and Zermatt—at an elevation of 1638m. The sites are approximatively 37km apart.
In the Alpine regions of Switzerland, there is an obvious motivation to focus on extreme climatic events, as their impact on the local population and infrastructure can be very costly. As stated by Beniston [2007], warm winter spells, that is, periods with strong positive temperature exceedances in winter, can exert significant impacts on the natural ecosystems, agriculture, and water supply:

> "Temperatures persistently above 0°C will result in early snow-melt and a shorter seasonal snow cover, early water runoff into river basins, an early start of the vegetation cycle, reduced income for alpine ski resorts and changes in hydro-power supply because of seasonal shifts in the filling of dams [Beniston, 2004]."

In this analysis, we are interested in the dynamics of the dependence between extreme air temperatures in Montana and Zermatt during the winter season. The dynamics of both extreme high and extreme low winter temperatures in these two sites will be assessed and linked to the following explanatory factors: time (in years) (t), day within season (d), and the NAO (North Atlantic Oscillation) index (z); the latter is a normalized pressure difference between Iceland and the Azores that is known to have a major direct influence on the alpine region temperatures, especially during winter [Beniston, 2005]. The choice of the studied sites is of great importance in this analysis. Beniston and Rebetez [1996] showed that both cold and warm winters exhibit temperature anomalies that are altitude-dependent, with high-elevation resorts being more representative of free atmospheric conditions and less likely to be contaminated by urban effects. Therefore, to study the "pure" effect of the above-mentioned explanatory covariates on the winter temperature extremal dependence, we choose the two high elevation sites Montana and Zermatt.

The data consist of daily winter temperature minima and maxima measured at 2m above ground surface and were obtained from the MeteoSwiss website

www.meteoswiss.admin.ch

The data were available from 1981 to 2016, giving a total of 3190 winter observations per site. Daily NAO index measurements were obtained from the NOAA (National Centers for Environmental Information), at

www.ngdc.noaa.gov/ftp.html

We first transform the minimum temperature data by multiplication by -1 and then fit at each site—and to both daily minimum and maximum temperatures—a generalized Pareto Distribution (GPD) [Coles, 2001, ch. 4]

$$\mathrm{G}_{\sigma,\xi}(y) = 1 - \left(1 + \xi\frac{y}{\sigma}\right)_+^{-1/\xi}, \tag{2.18}$$

to model events above the 95% quantile u_{95} for each of the four temperature time series. In (2.18), $\sigma > 0$ is the scale parameter that depends on u_{95}, and $-\infty < \xi < \infty$ is the shape parameter. As is common with temperature data analysis, we test the effect of time t on the behaviour of the threshold exceedances by allowing the scale parameter of the GPD (2.18) to smoothly vary with t [Chavez-Demoulin and Davison, 2005]. Based on the likelihood ratio tests, a model with a non-stationary scale parameter is preferred only in Zermatt for the threshold exceedances of the daily minimum temperatures (p-value ≈ 0.022). Graphical goodness-of-fit tests for the four GPD models are conducted by comparing the distribution of a test statistic S with the unit exponential distribution (if $Y \sim G_{\sigma,\xi}$, then $S = -\ln\{1 - G_{\sigma,\xi}(Y)\}$ is unit exponentially distributed). Figure 2.7 displays the resulting qq-plots and confirms the validity of these models. The fitted models are then used to transform the data to a common unit Fréchet scale by probability integral transform and where the empirical distribution is used below u_{95}. This results in two datasets of bivariate observations (in Montana and Zermatt) with unit Fréchet margins: one for the daily maximum temperatures and the other one for the daily minimum temperatures.

Following the theory developed in Section 2.2.1, we transform each of the two datasets into pseudo-datasets of radial and angular components. By retaining the angular observations corresponding to a radial component exceeding its 95% quantile in each pseudo-dataset, we end up with two pseudo-samples of 160 extreme bivariate (angular) observations in each pseudo-dataset.

2.5.2 Covariate-adjusted dependence of extreme temperatures

In the following analyses of the dynamics of the dependence between extreme temperatures in Montana and Zermatt—and in line with findings from previous analyses of extreme temperatures in Switzerland [Davison and Gholamrezaee, 2011, Davison et al., 2013, Dombry et al., 2013]—we assume asymptotic dependence in both extremely high and extremely low winter temperatures.

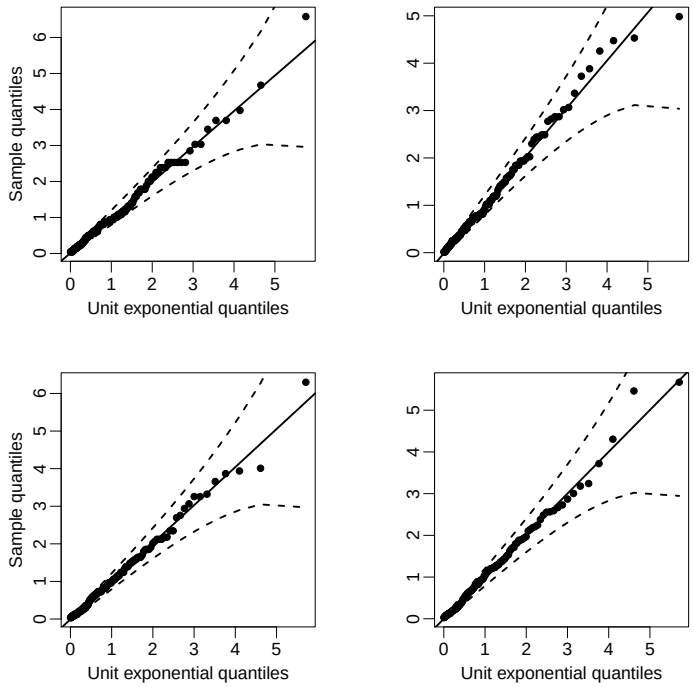

Figure 2.7: Diagnostic plots of the GPD modelling of the threshold exceedances of the daily maximum winter temperatures (left) and the daily minimum winter temperatures (right) in Montana (top) and Zermatt (bottom).

Dependence of extreme high winter temperatures

The covariate-adjusted bivariate angular densities presented in Section 2.2.2 are now fitted to the pseudo-sample of extreme high temperatures. The effects of the explanatory covariates t, z, and d are tested in each of the three angular densities: the logistic model (Example 1) with parameter $\alpha(t, z, d)$, the Dirichlet model (Example 2) with parameters $\alpha(t, z, d)$ and $\beta(t, z, d)$, and the Hüsler–Reiss model (Example 3) with parameter $\lambda(t, z, d)$. Within each family of covariate-adjusted angular densities, likelihood ratio tests (LRT) are performed to select the most adequate VGAM for the dependence parameters. Table 2.2 shows the best models in each of the three families of angular densities. All the considered covariates have a significant effect on the strength of dependence between extreme high temperatures in Montana and Zermatt. For the covariate-dependent Dirichlet model, the covariates affect the dependence parameters α and β differently. However, these parameters lack interpretability, and Coles and Tawn [1994] mention the quantities $(\alpha + \beta)/2$ and $(\alpha - \beta)/2$ that can be interpreted as the strength and asymmetry of the extremal dependence, respectively. In this case, the best Dirichlet dependence model found in Table 2.2 is such that both the intensity and the asymmetry of the dependence are affected by time, NAO, and day in season.

The best models in the studied angular density families are then compared by means of the AIC (see Section 2.3.2) displayed in Table 2.2. The Dirichlet model with $\alpha(z)$ and $\beta(t, d)$ parameters has the lowest AIC and is hence selected. This suggests the presence of asymmetry in the dependence of extreme high temperatures between Montana and Zermatt. Figure 2.8 shows the fitted smooth effects of the covariates on the extremal coefficient—constructed via the covariate-adjusted extremal coefficient as in (2.10)—that lies between 1 for perfect extremal dependence and 2 for perfect extremal independence.

Table 2.2: Selected models in each family of angular densities along with their AICs. The link functions g are the logit function for the logistic model and the logarithm function for the Dirichlet and the Hüsler–Reiss models. The functions $\hat{f}$ with subscripts t, z, and d are fitted smooth functions of time, NAO, and day in season, respectively

Covariate-adjusted angular density	VGAM	AIC
Logistic	$\hat{\alpha}(t,z,d) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_t(t) + \hat{f}_z(z) + \hat{f}_d(d)\}$	-280.15
Dirichlet	$\hat{\alpha}(z) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_z(z)\}$ $\hat{\beta}(t,d) = g^{-1}\{\hat{\beta}_0 + \hat{f}_t(t) + \hat{f}_d(d)\}$	-290.05
Hüsler–Reiss	$\hat{\lambda}(t,z,d) = g^{-1}\{\hat{\lambda}_0 + \hat{f}_t(t) + \hat{f}_z(z) + \hat{f}_d(d)\}$	-275.64

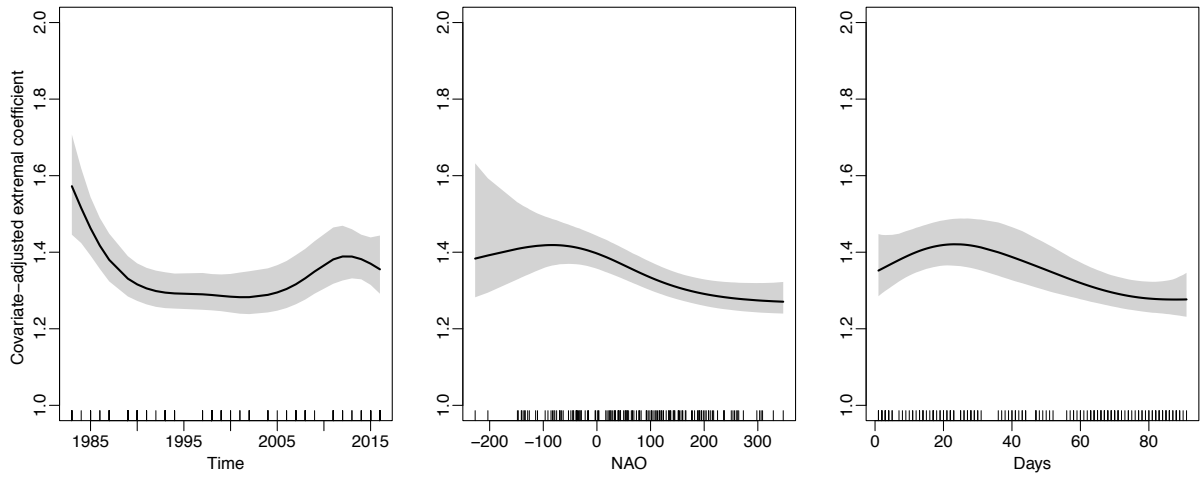

Figure 2.8: Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 2.2 along with their associated 95% (pointwise) asymptotic confidence bands.

A decrease in the extremal coefficient, or equivalently an increase in the extremal dependence between high winter temperatures in Montana and Zermatt, is observed from 1988 until 2006. This change might be explained first by a warm phase of very pronounced and persistent warm anomalies during the winter season, which occured countrywide from 1988 to 1999 [Jungo and Beniston, 2001], and then by an exceptionally warm 2006/2007 winter that took place in Europe [Luterbacher et al., 2007]. Regarding the NAO effect, as expected, we observe an increase in the extremal dependence during the positive phase of NAO that has a geographically global influence on the Alps and results in warmer and milder winters, as depicted by Beniston [1997]. In terms of the very negative NAO values (less than -100), there is an important uncertainty due to the corresponding small amount of joint extreme high temperatures (8%). The right panel of Figure 2.8 suggests an increase in the extremal dependence around mid-December. This evidence also seems compatible with the countrywide findings by Beniston [1997], who claims that

> "The anomalously warm winters have resulted from the presence of very persistent high pressure episodes which have occurred essentially during periods from late Fall to early Spring."

The sensitivity of the dependence modelling to the choice of the radial threshold is assessed by fitting the Dirichlet model of Table 2.2 to the angular observations corresponding to a radial component exceeding its 90%, 93%, and 97% quantile. The fitted smooth effects

Table 2.3: Selected models in each family of angular densities along with their AICs. The link functions g are the logit function for the logistic model and the logarithm function for the Dirichlet and the Hüsler–Reiss models. The functions $\hat{f}$ with subscripts t and d are fitted smooth functions of time and day in season, respectively

Covariate-adjusted angular density	VGAM	AIC
Logistic	$\hat{\alpha}(d) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_d(d)\}$	-402.76
Dirichlet	$\hat{\alpha} \equiv g^{-1}(\hat{\alpha}_0)$ $\hat{\beta}(t,d) = g^{-1}\{\hat{\beta}_0 + \hat{f}_t(t) + \hat{f}_d(d)\}$	-404.95
Hüsler–Reiss	$\hat{\lambda}(t,d) = g^{-1}\{\hat{\lambda}_0 + \hat{f}_t(t) + \hat{f}_d(d)\}$	-402.98

of the covariates on the extremal coefficient are reported in Figure 2.11 of Section 2.7.2 and are found to be essentially unaffected by the choice of the radial threshold.

Dependence of extreme low winter temperatures

The effects of the covariates time, NAO, and day in season on the dependence between extreme cold winters in Montana and Zermatt are now tested by fitting the bivariate angular densities of Section 2.2.2. Within each of the logistic, Dirichlet, and Hüsler–Reiss families, LRTs are performed, and the selected models are displayed in Table 2.3.

The explanatory covariates have different effects on the extremal dependence, depending on the family of angular densities. The AICs for the fitted models are quite close, and the asymmetric Dirichlet model has the lowest AIC and is hence the retained model. As opposed to the extremal dependence between warm winters in the two mountain sites, the NAO has a non-significant effect on the extremal dependence between cold winters. This might be explained by the fact that high values of the NAO index will affect the frequency of extreme low winter temperatures (less extremes) and hence the marginal behaviour of the extremes at both sites, but not necessarily the dependence of the extremes between these sites [Beniston, 2004, sec. 7.3.2].

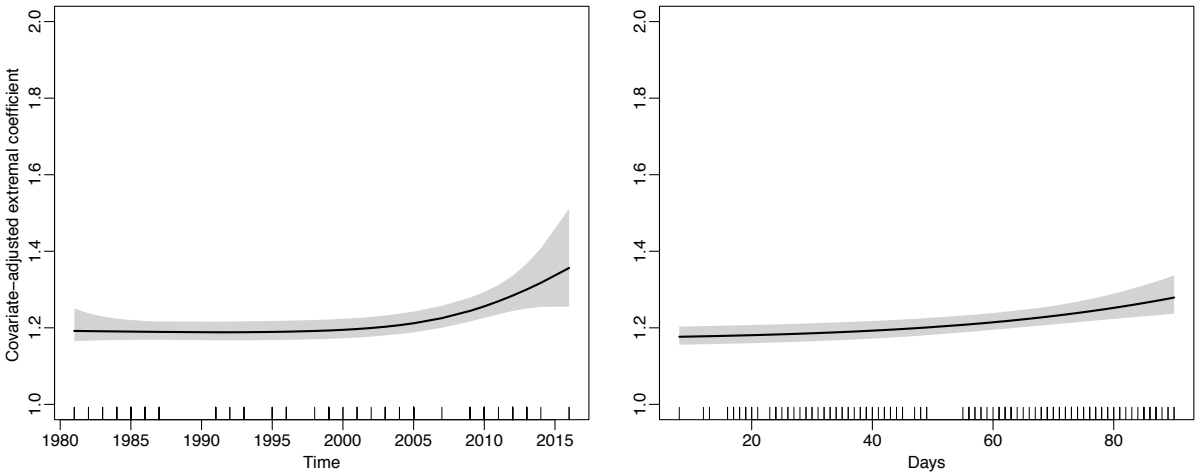

Figure 2.9: Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 2.3 along with their associated 95% (pointwise) asymptotic confidence bands.

Figure 2.9 shows the fitted smooth effects of time and day in season. A sensitivity

analysis of these effects to the radial threshold choice is conducted in Section 2.7.2 and Figure 2.12 shows that the effects of the considered covariates on the extremal dependence are unaltered by changing the radial threshold. The extremal dependence between low winter temperatures in Montana and Zermatt is high, regardless of the values taken by the covariates t and d. The range of values of the extremal coefficient observed in Figure 2.9 is in line with the findings of Davison et al. [2013], where the value of the extremal coefficient for the dependence between extreme low winter temperatures (in Switzerland) is around 1.3 for pairs of resorts separated by up to 100km. Overall, the extremal coefficient is lower in the extreme low winter temperatures than in the extreme high winter temperatures. This could be explained by the fact that minimum winter temperatures are usually observed overnight when the atmosphere is purer and not affected by local sunshine effects and hence is more favourable to the propagation over space of cold winter spells.

A decrease in the extremal dependence is observed from around 2007 and results in values of the extremal coefficient that are comparable to those obtained under the warm winter spells scenario (see Figure 2.8). This can be explained by a decrease in the intensity of the joint extreme low temperatures, that is, milder joint extreme low temperatures, occurring during the last years of the analysis, as can be observed in Figure 2.10. The right panel of Figure 2.9 highlights a decrease in the extremal dependence when approaching spring. This effect can be explained by the fact that mountains often produce their own local winds, as can be seen for instance from

www.morznet.com/morzine/climate/local-climate-in-the-alps

These warm dry winds are mostly noticeable in spring and are called Foehn in the Alps. Local effects obviously lead to a decrease of extremal dependence between the two resorts.

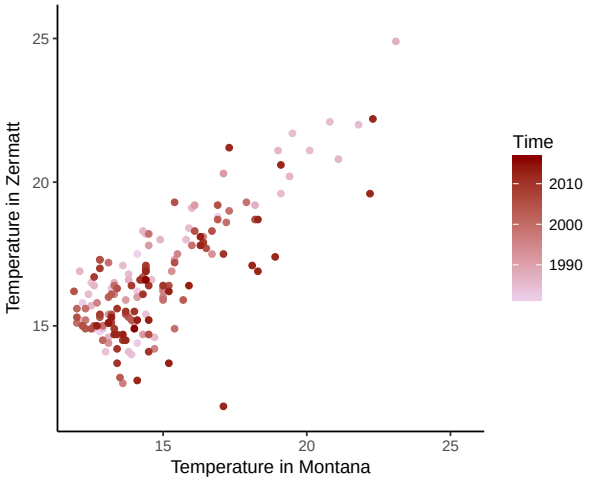

Figure 2.10: Scatterplot of (minus) extreme low winter temperatures (in °C) in Montana and Zermatt.

2.6 Final remarks

In this work, we have introduced a sturdy and general approach to model the influence of covariates on the extremal dependence structure. Keeping in mind that extreme values are scarce, our methodology borrows strength from a parametric assumption and benefits directly from the flexibility of VGAMs. Our non-linear approach for covariate-varying extremal dependences can be regarded as a model for conditional extreme value copulas—or equivalently as a model for nonstationary multivariate extremes. An important advantage over existing methods is that our model profits from the VGAM framework, allowing the incorporation of a large number of covariates of different types (continuous, factor, etc) as well as the possibility for the smooth functions to accommodate different shapes. The fitting procedure is an iterative ridge regression, the implementation of which is based on an ordinary N–R type algorithm that is available in many statistical software. An illustration is provided in the R code that can be found in the following repository https://github.com/lindamhalla/Regression_type_models_for_extremal_dependence.

The method paves the way for novel applications, as it is naturally tailored for assessing how covariates affect dependence between extreme values—and thus it offers a natural approach for modelling conditional risk. Conceptually, the proposed approach is valid in high dimensions. Yet, as for the classical setting without covariates, the number of parameters would increase quickly with the dimension and additional complications would arise. Relying on composite likelihoods [Padoan et al., 2010] instead of the full likelihood seems to represent a promising path for future extensions of the proposed methodology in a high-dimensional context.

2.7 Supplementary Materials

2.7.1 Auxiliary lemmas and proofs

Recall the following notations

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n_{\mathbf{r}}} c_i + \log\{h(\mathbf{w}_i;\boldsymbol{\beta})\}, \\ \mathbf{m}(\boldsymbol{\beta}) &= \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \\ \mathbf{m}(\mathbf{w},\boldsymbol{\beta}) &= \frac{\partial \log\{h(\mathbf{w};\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}}, \end{aligned}$$

where c_i is a constant independent of $\boldsymbol{\beta}$, for $i = 1, \ldots, n_{\mathbf{r}}$.
As mentioned in Section 2.3.3, the penalized log-likelihood estimator (PMLE) $\hat{\boldsymbol{\beta}}$ satisfies the following score equation

$$\mathbf{m}(\boldsymbol{\beta}) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta} = \mathbf{0}_{p(1+q\tilde{d})}. \tag{2.19}$$

We now define

$$\boldsymbol{\phi}(\boldsymbol{\beta}) = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}, \quad \boldsymbol{\phi}(\mathbf{w},\boldsymbol{\beta}) = \frac{\partial^2 \log\{h(\mathbf{w};\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}.$$

Based on Assumption (A2), we prove the following two lemmas that will streamline the proof of the first part of Theorem 2.3.1, i.e., the weak consistency of $\hat{\boldsymbol{\beta}}$.

Lemma 1. *Let $h(\mathbf{w};\boldsymbol{\beta})$ be continuously differentiable a.e. for $\boldsymbol{\beta} \in \mathbf{B}$. If $\int \sup_{\boldsymbol{\beta}\in\mathbf{B}} \|\partial h(\mathbf{w};\boldsymbol{\beta})/\partial\boldsymbol{\beta}\| \, \mathrm{d}\mathbf{w} < \infty$, then for $\boldsymbol{\beta} \in \mathbf{B}$:*

1. *$\int h(\mathbf{w};\boldsymbol{\beta}) \, \mathrm{d}\mathbf{w}$ is continuously differentiable.*
2. *$\int \partial h(\mathbf{w};\boldsymbol{\beta})/\partial\boldsymbol{\beta} \, \mathrm{d}\mathbf{w} = \partial \int h(\mathbf{w};\boldsymbol{\beta}) \, \mathrm{d}\mathbf{w}/\partial\boldsymbol{\beta}$.*

Proof. See Newey and McFadden [1994, Lemma 3.6]. □

Lemma 2. *If (A2) holds, then*

$$\mathrm{E}\left\{\mathbf{m}(\mathbf{W},\boldsymbol{\beta}_0)\right\} = \mathbf{0}, \quad -\mathrm{E}\left\{\boldsymbol{\phi}(\mathbf{W},\boldsymbol{\beta}_0)\right\} = \mathbf{i}(\boldsymbol{\beta}_0).$$

Proof. By Lemma 1, it follows that

$$\mathrm{E}\left\{\mathbf{m}(\mathbf{W},\boldsymbol{\beta}_0)\right\} = \int \mathbf{m}(\mathbf{w},\boldsymbol{\beta}_0) h(\mathbf{w};\boldsymbol{\beta}_0) \, \mathrm{d}\mathbf{w} = \int \left.\frac{\partial h(\mathbf{w};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \mathrm{d}\mathbf{w} = \left.\frac{\partial \int h(\mathbf{w};\boldsymbol{\beta}) \, \mathrm{d}\mathbf{w}}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \mathbf{0}.$$

Now, using (A2) (parts 3 and 4) we have that

$$\begin{aligned}
\mathrm{E}\left\{\boldsymbol{\phi}(\mathbf{W},\boldsymbol{\beta}_0)\right\} &= \int \boldsymbol{\phi}(\mathbf{w},\boldsymbol{\beta}_0) h(\mathbf{w};\boldsymbol{\beta}_0) \, \mathrm{d}\mathbf{w} \\
&= \int \left.\frac{\partial \mathbf{m}(\mathbf{w},\boldsymbol{\beta}) h(\mathbf{w};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} h(\mathbf{w};\boldsymbol{\beta}_0) \, \mathrm{d}\mathbf{w} - \int \mathbf{m}(\mathbf{w},\boldsymbol{\beta}_0)\mathbf{m}(\mathbf{w},\boldsymbol{\beta}_0)^\top h(\mathbf{w};\boldsymbol{\beta}_0) \, \mathrm{d}\mathbf{w} \\
&= -\int \mathbf{m}(\mathbf{w},\boldsymbol{\beta}_0)\mathbf{m}(\mathbf{w},\boldsymbol{\beta}_0)^\top h(\mathbf{w};\boldsymbol{\beta}_0) \, \mathrm{d}\mathbf{w}.
\end{aligned}$$

□

We now prove the first part of Theorem 2.3.1 by proving the following lemma.

Lemma 3. *Let $Q_{\mathbf{a}} = \{\boldsymbol{\beta} \in \mathbf{B} : \boldsymbol{\beta} = \boldsymbol{\beta}_0 + n_{\mathbf{r}}^{-1/2}\mathbf{a}\}$ be the surface of the sphere around $\boldsymbol{\beta}_0$ with radius $n_{\mathbf{r}}^{-1/2}\|\mathbf{a}\|$. Then, for every $\varepsilon > 0$, there exists $\mathbf{a}$ such that*

$$\Pr\left\{\sup_{\boldsymbol{\beta}\in Q_{\mathbf{a}}} \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) < \ell(\boldsymbol{\beta}_0,\boldsymbol{\gamma})\right\} \geq 1-\varepsilon,$$

for $n_{\mathbf{r}}$ large enough.

Proof. Let $\boldsymbol{\beta} \in Q_a$, i.e., there exists $\mathbf{a}$ such that $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + n_{\mathbf{r}}^{-1/2}\mathbf{a}$. Applying a second-order Taylor expansion around $\boldsymbol{\beta}_0$ of the penalized log-likelihood $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$, we have

$$\begin{aligned}
\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) - \ell(\boldsymbol{\beta}_0,\boldsymbol{\gamma}) &= \ell(\boldsymbol{\beta}) - \ell(\boldsymbol{\beta}_0) - \frac{1}{2}\left\{\boldsymbol{\beta}^\top\mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta} - \boldsymbol{\beta}_0^\top\mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta}_0\right\} \\
&= n_{\mathbf{r}}^{-1/2}\mathbf{m}(\boldsymbol{\beta}_0)^\top\mathbf{a} + \frac{n_{\mathbf{r}}^{-1}}{2}\mathbf{a}^\top\boldsymbol{\phi}(\boldsymbol{\beta}_0)\mathbf{a} - \frac{n_{\mathbf{r}}^{-1}}{2}\mathbf{a}^\top\mathbf{P}(\boldsymbol{\gamma})\mathbf{a} - n_{\mathbf{r}}^{-1/2}\boldsymbol{\beta}_0^\top\mathbf{P}(\boldsymbol{\gamma})\mathbf{a} \\
&\quad + \frac{n_{\mathbf{r}}^{-3/2}}{2}\sum_q\sum_r\sum_s a_q a_r a_s \left.\frac{\partial^3\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{qrs}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}, \qquad (2.20)
\end{aligned}$$

with $\boldsymbol{\beta}^*$ in the interior of $Q_{\mathbf{a}}$. The terms involving the penalty matrix $\mathbf{P}(\boldsymbol{\gamma})$ converge in probability to $\mathbf{0}$ due to the vanishing penalty from (A1). By the central limit theorem (based on a Lindeberg-type condition), Lemma 2, and (A2), we have that $n_{\mathbf{r}}^{-1/2}\mathbf{m}(\boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{i}(\boldsymbol{\beta}_0))$ implying that $|n_{\mathbf{r}}^{-1/2}\mathbf{m}(\boldsymbol{\beta}_0)^\top\mathbf{a}| = O_p(1)\|\mathbf{a}\|$. By the law of large numbers and

Lemma 2, we have that $n_{\mathbf{r}}^{-1}\boldsymbol{\phi}(\boldsymbol{\beta}_0) \xrightarrow{p} -\mathbf{i}(\boldsymbol{\beta}_0)$. Hence, applying the continuous mapping theorem, we end up with $\mathbf{a}^\top n_{\mathbf{r}}^{-1}\boldsymbol{\phi}(\boldsymbol{\beta}_0)\mathbf{a}/2 \xrightarrow{p} -\mathbf{a}^\top\mathbf{i}(\boldsymbol{\beta}_0)\mathbf{a}/2 \leq \|\mathbf{a}\|^2\lambda_{\min}/2$, where $\lambda_{\min} > 0$ is the smallest eigenvalue of $\mathbf{i}(\boldsymbol{\beta}_0)$. Finally, Assumption (A2) (last part) implies that the terms $\partial^3\ell(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{qrs}|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} < \infty$ and that, by Cauchy–Schwartz inequality, the remainder term in (2.20) vanishes in probability (is $O_p(n_{\mathbf{r}}^{-1/2})$). Leaving out the terms vanishing in probability, (2.20) yields

$$\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) - \ell(\boldsymbol{\beta}_0,\boldsymbol{\gamma}) \leq O_p(1)\|\mathbf{a}\| - \|\mathbf{a}\|^2\lambda_{\min}/2 = T, \quad \boldsymbol{\beta} \in Q_{\mathbf{a}},$$

for $n_{\mathbf{r}}$ large enough. Thus,

$$\Pr\left\{\sup_{\boldsymbol{\beta}\in Q_{\mathbf{a}}} \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) < \ell(\boldsymbol{\beta}_0,\boldsymbol{\gamma})\right\} \geq \Pr(T < 0),$$

implying that for every $\varepsilon > 0$, there exists an $\mathbf{a}$ such that $\Pr\{\sup_{\boldsymbol{\beta}\in Q_{\mathbf{a}}}\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) < \ell(\boldsymbol{\beta}_0,\boldsymbol{\gamma})\} \geq 1-\varepsilon$. Hence, with probability tending to 1, the penalized log-likelihood $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$ has a local maximum $\hat{\boldsymbol{\beta}}$ in the interior of a sphere around $\boldsymbol{\beta}_0$. □

Lemma 3 yields the first part of Theorem 2.3.1. Now we move to the second part of Theorem 2.3.1. The asymptotic normality of the PMLE $\hat{\boldsymbol{\beta}}$ is derived from a second-order Taylor expansion of the score equation (2.19) around the true parameter $\boldsymbol{\beta}_0$. The Taylor expansion of (2.19) yields

$$\mathbf{m}(\boldsymbol{\beta}_0) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta}_0 + \{\boldsymbol{\phi}(\boldsymbol{\beta}_0) - \mathbf{P}(\boldsymbol{\gamma})\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \mathbf{r} = \mathbf{0}_{p(1+q\tilde{d})},$$

where

$$\mathbf{r} = \frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \left.\frac{\partial^2\mathbf{m}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{2.21}$$

and $\boldsymbol{\beta}^*$ is such that $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \leq \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|$. Dividing (2.21) by $n_{\mathbf{r}}^{1/2}$, we obtain

$$\frac{1}{n_{\mathbf{r}}}\{\boldsymbol{\phi}(\boldsymbol{\beta}_0) - \mathbf{P}(\boldsymbol{\gamma}) + \tilde{\mathbf{r}}\}n_{\mathbf{r}}^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n_{\mathbf{r}}^{-1/2}\{\mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta}_0 - \mathbf{m}(\boldsymbol{\beta}_0)\}, \tag{2.22}$$

where

$$\tilde{\mathbf{r}} = \frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \left.\frac{\partial^2\mathbf{m}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}.$$

The consistency of $\hat{\boldsymbol{\beta}}$ and the assumption on the third order derivative of $\log\{h(\mathbf{w};\boldsymbol{\beta})\}$ in (A2) implies that $\tilde{\mathbf{r}} \xrightarrow{p} 0$. Assumption (A1) implies that the terms involving $\mathbf{P}(\boldsymbol{\gamma})$ vanish in probability. Since $n_{\mathbf{r}}^{-1/2}\mathbf{m}(\boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{i}(\boldsymbol{\beta}_0))$ and $n_{\mathbf{r}}^{-1}\boldsymbol{\phi}(\boldsymbol{\beta}_0) \xrightarrow{p} -\mathbf{i}(\boldsymbol{\beta}_0)$ (see above), Slutsky's theorem implies that $n_{\mathbf{r}}^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{i}(\boldsymbol{\beta}_0)^{-1})$ and proves hence the second part of Theorem 2.3.1.

2.7.2 Extreme temperature analysis

This section supplements Section 2.5.2.

Dependence of extreme high winter temperatures

The Dirichlet model of Table 2.2 is fitted to the pseudo-sample of extreme high temperatures where the angular observations corresponding to a radial component exceeding its 90%, 93%, and 97% quantiles, are considered. Figure 2.11 displays the fitted smooth effects of time, NAO, and day in season on the extremal coefficient, along with their associated 95% confidence intervals.

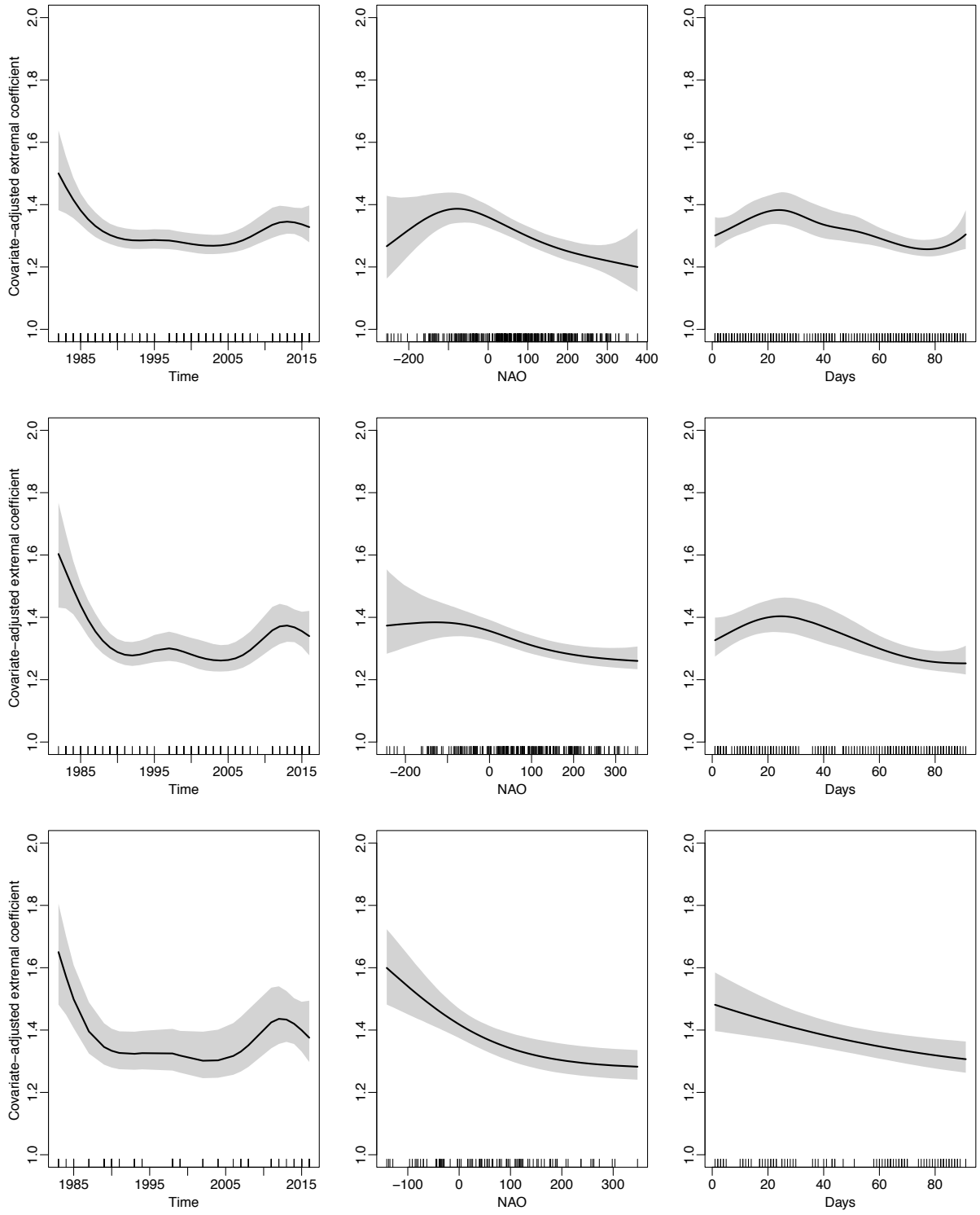

Figure 2.11: Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 2.2 along with their associated 95% (pointwise) asymptotic confidence bands. Different radial thresholds are considered: the 90% quantile (top), the 93% quantile (middle), and the 97% quantile (bottom).

Dependence of extreme low winter temperatures

The Dirichlet model of Table 2.3 is fitted to the pseudo-sample of extreme low temperatures where the angular observations corresponding to a radial component exceeding its

90%, 93%, and 97% quantiles, are considered. Figure 2.12 displays the fitted smooth effects of time and day in season on the extremal coefficient, along with their associated 95% confidence intervals.

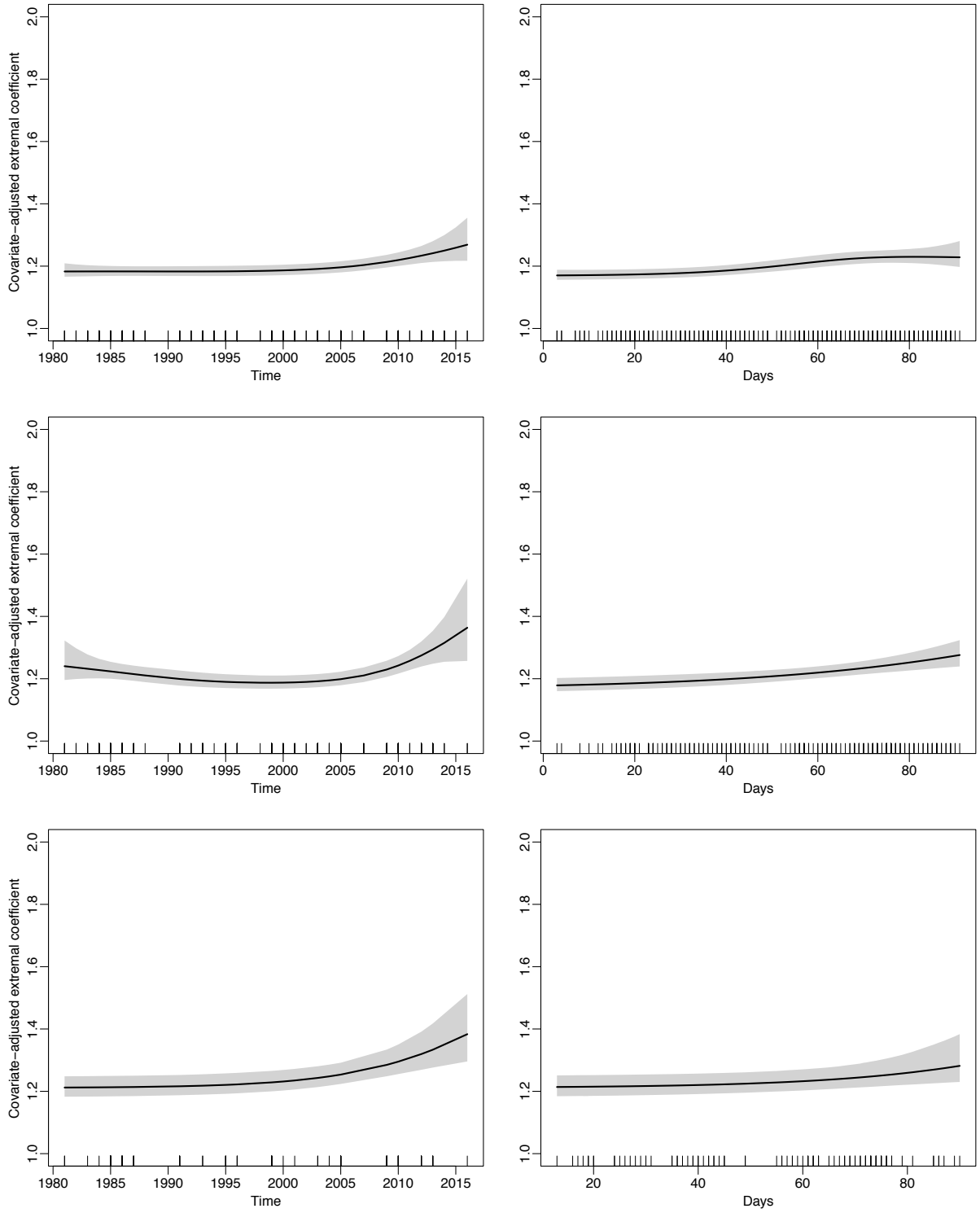

Figure 2.12: Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 2.3 along with their associated 95% (pointwise) asymptotic confidence bands. Different radial thresholds are considered: the 90% quantile (top), the 93% quantile (middle), and the 97% quantile (bottom).

Chapter 3

Non-linear models for extremal dependence

As seen in Section 1.1.2.4, the dependence structure of max-stable random vectors can be characterized by their Pickands' dependence function. In the constant perspective of borrowing information from covariates to model the extremal dependence, we develop, in the work presented in this chapter, a flexible, semi-parametric method for the estimation of non-stationary multivariate Pickands' dependence functions. The term "semi-parametric" refers here to the covariate model as we might have a part with a parametric dependence on the covariates and a part with a smooth dependence on the covariates. The estimation of the Pickands' function is however performed in a non-parametric way as any finite-parametric model for the extremal dependence cannot fully cover the class of Pickands' dependence functions, and further adjustments are conducted to ensure that the estimator is a genuine Pickands' function. The correction of (non-parametric) Pickands' estimators $\hat{A}(\omega)$ is a standard procedure and we show in Figure 3.1 an example of this correction in the bivariate case and based on the greatest convex minorant corrections $\{\hat{A}(\omega) \vee (1-\omega) \vee \omega\} \wedge 1$ [Deheuvels, 1991].

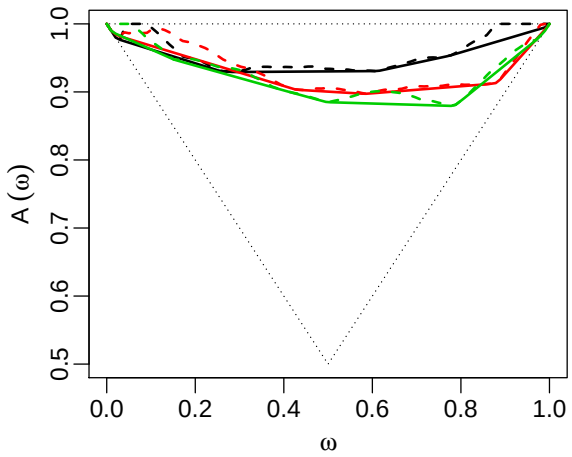

Figure 3.1: Non-corrected Pickands estimators [Pickands, 1981] (dashed lines) and their convex minorant corrections (solid lines).

We propose, in this chapter, new non-parametric estimators of covariate-dependent Pickands' dependence functions based on an accurate max-projection allowing to pass

from the multivariate to the univariate setting and to rely on the generalized additive modelling framework when incorporating covariate information. In the bivariate case, the resulting estimator of the Pickands' dependence function is regularized using constrained median smoothing B-splines, and bootstrap variability bands are constructed. In higher dimensions, we tailor our approach to the estimation of the extremal coefficient. An extended simulation study suggests that our estimator performs well and is competitive with the standard estimators in the absence of covariates. We apply the new methodology to a temperature dataset in the U.S. where the extremal dependence is linked to time and altitude.

3.1 Introduction

Extreme events are of major importance in different fields. Such events, defined as rare and severe, may cause considerable material damage, deaths, and economic losses. A proper statistical framework should then provide the tools for measuring extreme events through the quantification of their frequencies, as well as intensities, which can be formulated in terms of quantiles. Often, in practice, there is much more to gain by understanding simultaneous extreme events of several quantities. For example, in finance, people are interested in the values of several assets that constitute the portfolio and the risk that the assets collapse together. In environmental sciences, it is vital to measure the risk of simultaneous flooding at different points of a river [Asadi et al., 2015] or at nearby rivers. In both cases, the quantity of interest is measured by the extremal dependence between different variables. There is a well-established literature on modelling multivariate extremes and thus, extremal dependence structures; some authors addressed this matter using Gaussian copulas [Renard and Lang, 2007] while others analysed multivariate extreme value copulas [Gudendorf and Segers, 2012]. When the interest is in modelling multivariate extremes, the notion of max-stable processes is very important. Max-stable processes arise as the limiting distribution of suitably normalized componentwise maxima of independent replications of some continuous stochastic process [de Haan, 1984, de Haan and Ferreira, 2006, Schlather, 2002]. The dependence structure of a multivariate max-stable random vector can be characterized by its Pickands' dependence function A (see Section 3.2.1) defined on the unit simplex $S_{d-1} = \left\{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in \mathbb{R}_+^d : \omega_1 + \cdots + \omega_d = 1\right\}$ through the Pickands' representation theorem [Pickands, 1981].

In this work, our main interest is modelling possible changes in the dependence structure among multivariate block maxima according to a given set of covariates $\mathbf{x}$. Moreover, the effect of some covariates does not need to have a linear form but may vary smoothly. In classical regression analysis, a popular and flexible set of models is the class of generalized additive models [Green and Silverman, 1994, Hastie and Tibshirani, 1990], which link the mean behaviour of a random variable Y with a set of covariates $\mathbf{X} \in \mathbb{R}^q$ through

$$\mathrm{E}\left(Y \mid \mathbf{X} = \mathbf{x}\right) = g\left\{\mathbf{u}^\top \boldsymbol{\beta} + \sum_{k=1}^{K} h_k(t_k)\right\}, \tag{3.1}$$

where

- g is a link function,
- $(u_1, \ldots, u_s)$ and $(t_1, \ldots, t_K)$ are subsets of $\{x_1, \ldots, x_q\}$,

- $\boldsymbol{\beta} \in \mathbb{R}^s$ is a vector of parameters, and
- $h_k : \mathbb{T}_k \to \mathbb{R}$ are smooth functions supported on closed $\mathbb{T}_k \subset \mathbb{R}$, for all k.

Chavez-Demoulin and Davison [2005] proposed a first step to bridge extreme value theory and generalized additive models (GAM). They modelled the marginal behaviour of extremes with a GAM, but they did not investigate the effect of covariates on the dependence structure. In this work, we fill this gap in the context of block maxima analysis by developing a methodology for estimating conditional Pickands' dependence functions $A(\cdot|\mathbf{x})$. The method relies on the max-projection of a d-dimensional max-stable random vector $\mathbf{Z}(\mathbf{x})$, where $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_d(\mathbf{x}))^\top$ given a set of covariates $\mathbf{x}$. We show that, conditionally on $\mathbf{x}$ and at any fixed value in the unit simplex, the univariate max-projection follows a classical two-parameter Beta distribution. The second parameter of the Beta distribution is equal to 1, and the first parameter is exactly the Pickands' function evaluated at the specific value in the unit simplex. Under the resulting likelihood-based framework, we benefit from the well-established framework of the GAM to fit non-stationary Pickands' functions evaluated at the specific value in the unit simplex. Repeating the max-projection procedure along with the GAM at several values in the unit simplex generates estimates of semi-parametric models for the Pickands' function conditional on $\mathbf{x}$. In the bivariate case, for a fixed value of $\mathbf{x}$, the estimated Pickands' function is regularized by constrained median smoothing B-splines. When $d \geq 3$, the focus is on the regularization of the non-stationary extremal coefficient estimate (see Section 3.3.4), as it allows for a simple interpretation of the extremal dependence in high dimensions.

To the best of our knowledge, only Escobar-Bach et al. [2016] addressed the dependence of the Pickands' function on covariates. They considered a robust estimator, in the bivariate case, using local estimation with the minimum density power divergence criterion. Few works addressed specific non-stationarity in the multivariate and spatial contexts: Huser and Genton [2016] developed non-stationary max-stable dependence structures in which covariates are incorporated, and their inference is based on pairwise likelihoods. de Carvalho and Davison [2014] constructed a model for a family of spectral measures from several populations, to each of which a set of predictors is considered. They introduced the spectral density ratio model that they fitted using empirical likelihood methods and showed how to relate the resulting tilting parameters to covariates. Jonathan et al. [2014b] proposed a spline-based methodology to incorporate multiple covariates. Their approach used a Poisson process model for the rate of occurrence of the threshold excesses and a generalized Pareto model for the size of the threshold excesses. Our method provides a semi-parametric, flexible framework based on the classical setup of generalized additive models.

The chapter is organized as follows: In Section 3.2, we summarize the theory of max-stable random vectors and then introduce the notion of max-projection from multivariate to univariate settings. In Section 3.3, we develop the generalized additive model for the Pickands' dependence function in the multivariate context. We then propose a way to regularize the Pickands' function in the bivariate case and the extremal coefficient in higher dimensions. We implement a bootstrapping procedure to compute the variability bands in Section 3.3.3. Finally, in the last two sections, we perform simulations and address the estimation of a conditional model for the Pickands' dependence function of the monthly maxima of the maximum and minimum temperatures in the U.S.

3.2 Max-stable random vectors and max-projection

In this section, we first briefly recall some basics of max-stable random vectors and then propose a max-projection procedure that synthesizes the extremal dependence of multivariate max-stable random vectors within a univariate Beta random variable.

3.2.1 Max-stable random vectors

The usual strategy for modelling multivariate block maxima is to take advantage of the solid theoretical basis for modelling extreme events by completely characterizing the tail behaviour of the underlying generating process. Multivariate extreme value theory (see [Resnick, 1987, Chapter 5] and [de Haan and Ferreira, 2006, Chapter 6]) addresses the limiting behaviour of suitably normalized componentwise maxima $\mathbf{M}_n = (M_{n,1}, \ldots, M_{n,d})^\top$ where $M_{n,j}$ represents the maximum, over a block of size n, of the jth component of a vector of dimension d. The asymptotic theory of multivariate extremes is concerned with finding a non-degenerate multivariate distribution function G such that as $n \to \infty$,

$$\Pr\left\{(M_{n,1} - b_{n,1})/a_{n,1} \leq z_1, \ldots, (M_{n,d} - b_{n,d})/a_{n,d} \leq z_d\right\} \to G(z_1, \ldots, z_d), \tag{3.2}$$

with sequences $a_{n,j} > 0$, $b_{n,j} \in \mathbb{R}$, $j \in \{1, \ldots, d\}$, and G a distribution function with non-degenerate margins. The marginal distributions of G are univariate max-stable or equivalently generalized extreme value (GEV) distributions (see, e.g., Coles [2001] and Section 3.5 for our treatment of the marginal behaviours). Without loss of generality (see [Resnick, 1987, Section 5.4]), we assume that G has common unit Fréchet margins; that is, $G_j(z_j) = \exp(-1/z_j)$ for $z_j > 0$ and $j \in \{1, \ldots, d\}$.

If the limit (3.2) exists, and G is non-degenerate, then

$$G(z_1, \ldots, z_d) = \exp\left\{-\int_{S_{d-1}} \max_{i=1,\ldots,d}\left(\frac{\omega_i}{z_i}\right) dH(\omega_1, \ldots, \omega_d)\right\} = \exp\left\{-V(\mathbf{z})\right\},$$

where $\mathbf{z} = (z_1, \ldots, z_d)^\top$ and H is an arbitrary finite measure that satisfies

$$\int_{S_{d-1}} \omega_j dH(\omega_1, \ldots, \omega_d) = 1, \quad j \in \{1, \ldots, d\}. \tag{3.3}$$

Condition (3.3) is necessary for the margins to be unit Fréchet. The positive function $V(\mathbf{z})$ is called the exponent function and is continuous, convex, and homogeneous of order -1, i.e., $V(t\mathbf{z}) = t^{-1}V(\mathbf{z})$, for all $t > 0$. The homogeneity property of $V(\cdot)$ implies that $V(\mathbf{z}) = V(\mathbf{z}/|\mathbf{z}|)/|\mathbf{z}|$ where $|\mathbf{z}| = z_1 + \cdots + z_d$. An alternative expression for $V(\cdot)$ is derived by writing it as a departure from the independence case $V_\perp(\mathbf{z}) = 1/z_1 + \cdots + 1/z_d$ as follows

$$V(\mathbf{z}) = \left(\frac{1}{z_1} + \cdots + \frac{1}{z_d}\right) A(\boldsymbol{\omega}),$$

where the non-negative scalar vector $\boldsymbol{\omega} = (z_1/|\mathbf{z}|, \ldots, z_d/|\mathbf{z}|)^\top$ belongs to the unit simplex S_{d-1}, and $A(\boldsymbol{\omega})$ is the Pickands' dependence function mentioned in Section 3.1. The exponent function can then be thought of as the extension of the Pickands' function to $[0, \infty)^d$. If $A(\boldsymbol{\omega}) = 1$ for all $\boldsymbol{\omega} \in S_{d-1}$, we retrieve $V_\perp(\mathbf{z})$. To be a valid Pickands' function, A has to satisfy a few conditions. The classical necessary conditions are the following

C1) A is continuous and convex, i.e., for $a \in [0, 1]$,

$$A\left\{a\boldsymbol{\omega}_1 + (1-a)\boldsymbol{\omega}_2\right\} \leq aA(\boldsymbol{\omega}_1) + (1-a)A(\boldsymbol{\omega}_2), \quad \boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \in S_{d-1},$$

C2) For any $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)^\top \in S_{d-1}$, $A(\boldsymbol{\omega})$ has lower and upper limits

$$\max(\omega_1, \ldots, \omega_d) \leq A(\boldsymbol{\omega}) \leq 1,$$

C3) $A(0, \ldots, 0) = 1$ and $A(e_i) = 1$, for the boundary points with $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ for $i \in \{1, \ldots, d\}$.

In condition C2, the lower and upper bounds represent the complete dependence and complete independence cases, respectively. Conditions C1–C2–C3 are sufficient in the bivariate case and necessary but not sufficient when $d \geq 3$. In fact, Beirlant et al. [2004, p. 257] discussed the example of a function that satisfies C1–C2–C3 but is not a valid Pickands' dependence function. Complete, sufficient, and necessary conditions for the validity of the stable tail dependence function [Drees and Kaufmann, 1998, Huang, 1992], and therefore the Pickands' function, can be found in Ressel [2013], see also Charpentier et al. [2014] for further details of the framework of d-variate stable tail dependence functions, with $d \geq 3$.

Two well-known non-parametric estimators of the Pickands' dependence function are those proposed by Pickands [1981] and Capéraà et al. [1997]. However, these estimators are not intrinsic, i.e., they do not yield valid Pickands' functions, and many new estimators or modified versions of these estimators have been studied. For example, Deheuvels [1991] investigated a truncation and convexification technique that allows any estimator to verify conditions C1–C2. Hall and Tajvidi [2000a] corrected the estimator defined in Pickands [1981] by using the greatest convex minorant and constrained splines while Genest and Segers [2009] derived a rank-based modified version of the Pickands [1981] and Capéraà et al. [1997] estimators with a correction on the end-points, and Cormier et al. [2014] used a quadratic constrained B-spline smoothing technique. The above mentioned papers contributed to the estimation of the Pickands' dependence function in the bivariate case. In the multivariate case, Fils-Villetard et al. [2008] projected the Pickands' estimator into a convex and closed set $\mathcal{A}$ of valid Pickands' functions by minimizing the L^2 distance between this estimate and the members of $\mathcal{A}$. Similar ideas are investigated in Gudendorf and Segers [2011, 2012]. A more recent projection method was derived in Marcon et al. [2014] and consisted of projecting an initial estimator into $\mathcal{A}$ through a Bernstein polynomials-based technique. In our context of non-stationary Pickands' functions, we propose in Section 3.3 a smoothing regularization technique based on constrained splines.

3.2.2 Max-projection

We assume a d-dimensional max-stable random vector $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_d(\mathbf{x}))^\top$ with unit Fréchet marginal distributions F and a predictor-dependent extremal dependence $A(\boldsymbol{\omega}|\mathbf{x})$, $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^q$. As our emphasis is on non-stationary extremal dependence structures, one may also consider the margins of $\mathbf{Z}(\mathbf{x})$ to be predictor-dependent, i.e., $F_{j,\mathbf{x}}(z) = \Pr_{\mathbf{x}}\{Z_j(\mathbf{x}) \leq z\}$, $j \in \{1, \ldots, d\}$. In that case, a regression model applied on the marginal GEV distributions' parameters (see, e.g., Chavez-Demoulin and Davison [2005]) is required to transform the margins to a unit Fréchet scale.

We now introduce a new random variable

$$Y(\boldsymbol{\omega})|\mathbf{x} = \max\left[F^{\lambda_1}\{Z_1(\mathbf{x})\}, \ldots, F^{\lambda_d}\{Z_d(\mathbf{x})\}\right], \tag{3.4}$$

with

$$\lambda_j = \omega_j \left(\frac{1}{\omega_1} + \cdots + \frac{1}{\omega_d}\right),$$

for $j \in \{1, \ldots, d\}$, and where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)^\top$ is a fixed, non-negative scalar vector on the unit simplex with non-zero entries.

Here and hereafter, (3.4) will be referred to as the max-projection of the max-stable random vector $\mathbf{Z}(\mathbf{x})$. The following proposition establishes the distribution and mean of $Y(\boldsymbol{\omega})|\mathbf{x}$.

PROPOSITION 3.2.1. *The max-projection of a d-dimensional max-stable random vector* $\mathbf{Z}(\mathbf{x})$ *with unit Fréchet margins results in a Beta distributed random variable with parameters* $A_{\boldsymbol{\omega}}(\mathbf{x})$ *and* 1*, i.e.,*

$$\Pr\{Y(\boldsymbol{\omega}) \leq y|\mathbf{x}\} = y^{A_{\boldsymbol{\omega}}(\mathbf{x})}, \text{ for } 0 \leq y \leq 1,$$

where $Y(\boldsymbol{\omega})|\mathbf{x}$ *is defined in* (3.4)*. In particular,* $\mathrm{E}\{Y(\boldsymbol{\omega})|\mathbf{x}\} = g^{-1}\{A_{\boldsymbol{\omega}}(\mathbf{x})\}$*, where the concave link function equals* $g(u) = u/(1-u)$*, with* $0 < u < 1$*.*

The quantity $A_{\boldsymbol{\omega}}(\mathbf{x})$ defines the conditional Pickands' function $A(\boldsymbol{\omega}|\mathbf{x})$ evaluated at a specific value of $\boldsymbol{\omega}$. Proposition 3.2.1 allows us, given any fixed $\boldsymbol{\omega}$, to take advantage of the well-established theory of GAM or GLM (generalized linear models) frameworks to estimate the predictor-dependent Pickands' function $A_{\boldsymbol{\omega}}(\mathbf{x})$. We develop this idea in Section 3.3.

3.3 Smoothed models for max-stable dependence structures

In this section, we first propose a GAM for $A_{\boldsymbol{\omega}}(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$ in the multivariate context (Section 3.3.1). The resulting estimates of the conditional Pickands' functions at any value of $\boldsymbol{\omega}$ and covariate $\mathbf{x}$ are denoted by $\tilde{A}(\boldsymbol{\omega}|\mathbf{x})$. We then derive a regularized estimate of the conditional Pickands' function denoted $\hat{A}(\boldsymbol{\omega}|\mathbf{x})$ in the bivariate context (Section 3.3.2), as well as a regularized estimate of the extremal coefficient in higher dimensions (Section 3.3.4).

3.3.1 Semi-parametric estimation of the Pickands' dependence function in the multivariate setting

For $\mathbf{x} \in \mathcal{X}$, a generalized additive model for the transformed conditional mean of $Y(\boldsymbol{\omega})|\mathbf{x}$ can be written as

$$A_{\boldsymbol{\omega}}(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{u}^\top \boldsymbol{\beta} + \sum_{k=1}^{K} h_k(t_k), \tag{3.5}$$

where $\mathbf{u}$, $(t_1, \ldots, t_K)$, $\boldsymbol{\beta}$, and h_k are as in the classical GAM (3.1), and $\boldsymbol{\theta} \in \Theta$ is a vector of parameters containing $\boldsymbol{\beta}$ and h_k for all k. We assume that each smooth function $h_k \in \mathcal{C}^2(\mathbb{T}_k)$, the class of twice continuously differentiable functions on $\mathbb{T}_k$, admits a finite m_k-dimensional basis parametrized by $\mathbf{h}_k = (h_{k,1}, \ldots, h_{k,m_k})^\top \in \mathbb{R}^{m_k}$ and a quadratic penalty representation $\int_{\mathbb{T}_k} h_k''(t)^2\, dt = \mathbf{h}_k^\top S_k \mathbf{h}_k$, where S_k is a uniquely determined symmetric matrix [Green and Silverman, 1994, Hastie and Tibshirani, 1990, Wood, 2006b]. The class of twice continuously differentiable smoothers with finite quadratic penalty representation is broad and encompasses, among many flexible smoothers, the natural cubic splines, the tensor product splines, and the cyclic cubic splines which are all included in the `R` package `mgcv` [Wood, 2006b].

Such models are estimated by maximizing a penalized log-likelihood where the penalty term controls the roughness of the smoothers through a vector of smoothing parameters

$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ with higher values yielding smoother curves. The related effective degrees of freedom of each smooth function h_k are then defined as trace $(I + \gamma_k S_k)^{-1}$. Considering a sample of n observations $\{y_i(\boldsymbol{\omega}), \mathbf{x}_i\}_{i=1}^n$ built sequentially from $\{\mathbf{Z}(\mathbf{x}_1), \ldots, \mathbf{Z}(\mathbf{x}_n)\}$ by fixing the value of $\boldsymbol{\omega} \in \mathcal{S}_{d-1} \setminus \{\boldsymbol{\omega} \in \mathcal{S}_{d-1} | \exists_i, \omega_i = 0\}$ and applying the max-projection (3.4), we can write the penalized log-likelihood as

$$\ell_p(\boldsymbol{\theta}, \boldsymbol{\gamma}; \boldsymbol{\omega}) = \ell_{\boldsymbol{\omega}}(\boldsymbol{\theta}) - \frac{1}{2} \sum_{k=1}^{K} \gamma_k \int_{\mathbb{T}_k} h_k''(t_k)^2 dt_k = \ell_{\boldsymbol{\omega}}(\boldsymbol{\theta}) - \frac{1}{2} \sum_{k=1}^{K} \gamma_k \mathbf{h}_k^\top S_k \mathbf{h}_k, \tag{3.6}$$

where the first term $\ell_{\boldsymbol{\omega}}(\boldsymbol{\theta})$ is the log-likelihood of the Beta distributed random variables $\{y_i(\boldsymbol{\omega}), \mathbf{x}_i\}_{i=1}^n$, i.e.,

$$\ell_{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\{A_{\boldsymbol{\omega}}(\mathbf{x}_i; \boldsymbol{\theta}) - 1\} \ln y_i(\boldsymbol{\omega}) + \ln A_{\boldsymbol{\omega}}(\mathbf{x}_i; \boldsymbol{\theta}) \right]. \tag{3.7}$$

The smoothing parameters are chosen based on the Akaike information criterion (AIC), and we refer the interested reader to Section 4.5 in Wood [2006b] for more details regarding the selection of the smoothing parameters. Given $\boldsymbol{\gamma}$, the penalized log-likelihood (3.6) is maximized using an iterative weighted least squares procedure based on a Newton–Raphson algorithm (see Appendix A.1 in Chavez-Demoulin et al. [2015] for more details). The penalized maximum likelihood estimator is defined as

$$\tilde{\boldsymbol{\theta}}^{\boldsymbol{\omega}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}}\ \ell_p(\boldsymbol{\theta}, \boldsymbol{\gamma}; \boldsymbol{\omega}),$$

and the resulting estimator of the non-stationary Pickands' dependence function is $A_{\boldsymbol{\omega}}(\mathbf{x}; \tilde{\boldsymbol{\theta}}^{\boldsymbol{\omega}})$. Applying the model (3.5) at different values $\{\boldsymbol{\omega}^v\}_{v=1}^{\tilde{n}}$ with $\boldsymbol{\omega}^v \in \mathcal{S}_{d-1} \setminus \{\boldsymbol{\omega} \in \mathcal{S}_{d-1} | \exists_i, \omega_i = 0\}$ yields a set of estimates of the conditional Pickands' function

$$\left\{ \tilde{A}(\boldsymbol{\omega}^v | \mathbf{x}) \right\}_{v=1}^{\tilde{n}} = \left\{ A_{\boldsymbol{\omega}^v} \left(\mathbf{x}; \tilde{\boldsymbol{\theta}}^{\boldsymbol{\omega}^v} \right) \right\}_{v=1}^{\tilde{n}}. \tag{3.8}$$

Note that for $\boldsymbol{\omega} \in \{S_{d-1} | \exists_i \omega_i = 1\}$, the values of $\tilde{A}(\boldsymbol{\omega}|\mathbf{x})$ are fixed at 1 so that condition C3 is satisfied. In practice, the GAM (3.5) is applied $\tilde{n}$ times, with $\tilde{n}$ large enough to get accurate estimates of $A(\boldsymbol{\omega}|\mathbf{x})$ in (3.8). From a computational aspect, the time to convergence of the Newton–Raphson algorithm at each $v \in \{1, \ldots, \tilde{n}\}$ depends on the size of the data and the covariate set size q, as for every standard GAM. The second parameter of the Beta distribution being equal to 1, we have only one parameter to estimate, and the algorithm typically converges quickly. For each model, or equivalently for each value of $\boldsymbol{\omega}$, the selection of the relevant covariates among $\mathbf{x}$ is performed using variable selection methods for GAM (see Marra and Wood [2011] for an extensive study and comparison of the available techniques). Theoretically, the selected models can differ from one value of $\boldsymbol{\omega}$ to another, but practically, we impose the same set of covariates for all values of $\boldsymbol{\omega}$. The imposed set of covariates is likely to be the same for any value of $\boldsymbol{\omega} \in \mathcal{S}_{d-1}$, except near the borders, and we set it equal to the set of covariates selected at $\boldsymbol{\omega} = (1/d, \ldots, 1/d)$ as most of the information about the extremal dependence is summarized at this point. This constraint ensures a coherence of the covariates, independently of $\boldsymbol{\omega}$, and simplifies our algorithm.

3.3.2 Corrected Pickands' function estimator in the bivariate setting

The estimator $\tilde{A}(\boldsymbol{\omega}|\mathbf{x})$ in (3.8) can be implemented for any finite d-dimensional max-stable random vector with an extremal dependence structure that is either stationary or varying according to a set of predictors. However, for a fixed value of the covariates $\mathbf{x}$, this estimator is not a valid Pickands' dependence function since it does not verify conditions C1–C2–C3.

In this section, we focus on the correction of the estimator $\tilde{A}(\boldsymbol{\omega}|\mathbf{x})$ in the bivariate case and for a fixed value of the covariates $\mathbf{x} = \mathbf{x}_0$. Our method consists of smoothing $\tilde{A}\{(\omega, 1-\omega)|\mathbf{x} = \mathbf{x}_0\}$ for $\omega \in [0,1]$ so that we obtain a regularized estimator of the Pickands' dependence function that verifies C1–C2–C3. This is performed using the median smoothing technique described in Koenker et al. [1994] combined with the constrained optimization algorithm proposed by He and Ng [1999] and improved by Ng and Maechler [2007] to be more efficient and faster. This technique is used in Cormier et al. [2014], where the authors estimated the Pickands' dependence function based on the A-plot, which is a graphical tool for detecting extremal dependence. The constrained quantile smoothing algorithm is implemented in the `R` package `COBS` [Ng and Maechler, 2015]. For notational convenience, we set $\tilde{A}(\omega|\mathbf{x}) = \tilde{A}\{(\omega, 1-\omega)|\mathbf{x}\}$. We discretize the interval $[0,1]$ by fixing $\tilde{n}$ regularly spaced values $0 = \omega_1 < \omega_2 < \cdots < \omega_{\tilde{n}-1} < \omega_{\tilde{n}} = 1$ and apply the constrained smoothing technique to $\tilde{A}(\omega_1|\mathbf{x} = \mathbf{x}_0), \ldots, \tilde{A}(\omega_{\tilde{n}}|\mathbf{x} = \mathbf{x}_0)$.

We define the median L_∞ smoothing spline as

$$\hat{g}_{\gamma,L_\infty}(.|\mathbf{x} = \mathbf{x}_0) = \min_g \sum_{i=1}^{\tilde{n}} |\tilde{A}(\omega_i|\mathbf{x} = \mathbf{x}_0) - g(\omega_i)| + \gamma \max_\omega |g''(\omega)|, \tag{3.9}$$

where the smoothing parameter γ is chosen based on the Schwarz information criterion [Schwarz, 1978]. In (3.9), the L_1 norm adds robustness to the estimator, and the penalty term ensures, as in (3.6), its smoothness. Koenker et al. [1994] showed that $\hat{g}_{\gamma,L_\infty}(.|\mathbf{x})$ can be approximated by a quadratic smoothing spline $g(\cdot|\mathbf{x})$ which has an equivalent B-spline representation

$$g(\omega|\mathbf{x}) = \sum_{j=1}^{N+3} a_j(\mathbf{x}) B_j(\omega),$$

where N is the number of interior knots, i.e., the number of fixed equidistant points in the support of the spline $g(\cdot|\mathbf{x})$ [Green and Silverman, 1994, Section 2.1.1].
The optimization problem (3.9) is equivalent to

$$\begin{aligned} &\min_g \sum_{i=1}^{\tilde{n}} |\tilde{A}(\omega_i|\mathbf{x} = \mathbf{x}_0) - g(\omega_i)| + \gamma\sigma \\ &\text{s.t.} \quad -\sigma \le g''(\omega_i) \le \sigma, \quad i \in \{1, \ldots, \tilde{n}\}, \end{aligned}$$

or equivalently,

$$\begin{aligned} &\min_{\boldsymbol{\theta} = (a_1(\mathbf{x}_0), \ldots, a_{N+3}(\mathbf{x}_0), \sigma)^\top \in \mathbb{R}^{N+4}} \sum_{i=1}^{\tilde{n}+1} |\tilde{\tilde{A}}_i(\mathbf{x}_0) - \mathbf{c}_i^\top \boldsymbol{\theta}| \\ &\text{s.t.} \quad \begin{pmatrix} D & 1 \\ -D & 1 \end{pmatrix} \boldsymbol{\theta} \ge 0, \end{aligned} \tag{3.10}$$

where

$$\tilde{\tilde{A}}(\mathbf{x}_0) = \begin{pmatrix} \tilde{A}(\omega_1|\mathbf{x}_0) & \cdots & \tilde{A}(\omega_{\tilde{n}}|\mathbf{x}_0) & 0 \end{pmatrix}^\top \text{ is the } (\tilde{n}+1)\times 1 \text{ pseudo response vector,}$$

$$\begin{pmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_{\tilde{n}+1} \end{pmatrix}^\top = \begin{pmatrix} B_1(\omega_1) & \dots & B_{N+3}(\omega_1) & 0 \\ \vdots & \cdots & \vdots & \vdots \\ B_1(\omega_{\tilde{n}}) & \dots & B_{N+3}(\omega_{\tilde{n}}) & 0 \\ 0 & \cdots & 0 & \gamma \end{pmatrix} \text{ is the } (\tilde{n}+1)\times(N+4) \text{ pseudo design matrix,}$$

$$\text{and } D = \begin{pmatrix} B_1''(\omega_2) & \dots & B_{N+3}''(\omega_2) \\ \vdots & \cdots & \vdots \\ B_1''(\omega_{\tilde{n}-1}) & \dots & B_{N+3}''(\omega_{\tilde{n}-1}) \end{pmatrix},$$

(see Ng and Maechler [2007] for more details).

We can now easily incorporate conditions C1–C2–C3 as a set of additional constraints in the minimization problem (3.10):

Convexity constraint:

We add a set of $\tilde{n}-2$ inequality constraints $\begin{pmatrix} D & \mathbf{0} \end{pmatrix} \boldsymbol{\theta} \geq 0$.

Lower bound constraint:

$$\begin{pmatrix} B_1(\omega_i) & \cdots & B_{N+3}(\omega_i) & 0 \end{pmatrix} \boldsymbol{\theta} \geq \max(\omega_i, 1-\omega_i), \quad \text{for } i \in \{1,\dots,\tilde{n}\}.$$

Border constraints:

$$\begin{pmatrix} B_1(\omega_1) & \cdots & B_{N+3}(\omega_1) & 0 \end{pmatrix} \boldsymbol{\theta} = 1, \quad \omega_1 = 0,$$
$$\begin{pmatrix} B_1(\omega_{\tilde{n}}) & \cdots & B_{N+3}(\omega_{\tilde{n}}) & 0 \end{pmatrix} \boldsymbol{\theta} = 1, \quad \omega_{\tilde{n}} = 1.$$

Finally, we define our regularized estimator of the Pickands' dependence function for a fixed value of the covariates $\mathbf{x}_0$ as $\hat{A}(\omega|\mathbf{x}_0) = \hat{g}_{\gamma,L_\infty}(\omega|\mathbf{x}_0)$, with

$$\hat{g}_{\gamma,L_\infty}(\omega|\mathbf{x}_0) = \sum_{j=1}^{N+3} \hat{a}_j(\mathbf{x}_0) B_j(\omega)$$

being the constrained quadratic median smoothing spline defined for any value $\omega \in [0,1]$ and satisfying conditions C1–C2–C3. To sum up, starting from the sequence (3.8) in the bivariate case, we sequentially apply this regularization procedure at any value $\mathbf{x}_0$ of $\mathbf{x}$ and therefore obtain the following smoothed conditional valid Pickands' function estimators

$$\left\{\hat{A}(\omega_v|\mathbf{x}_0)\right\}_{v=1}^{\tilde{n}}. \tag{3.11}$$

3.3.3 Variability bands

We construct variability bands [Akritas and Politis, 2003, Ruppert et al., 2003] for (3.11) using bootstrap techniques and, more specifically, the resampling cases method [Davison and Hinkley, 1997, Section 6.2.4]. This bootstrapping method is safe in that it does not assume any functional form of the regression, and it assumes only that the observations at hand are independent.

For a fixed value $\omega \in \{\omega_v\}_{v=1}^{\tilde{n}}$, consider a sample of n observations $\{y_i(\omega), \mathbf{x}_i\}_{i=1}^n$ built sequentially from the bivariate observations $\{(Z_1(\mathbf{x}_i), Z_2(\mathbf{x}_i))\}_{i=1}^n$ by max-projection. Then, the resampling cases bootstrap is performed as follows:

Resampling cases bootstrap algorithm:

For each $r \in \{1, \ldots, R\}$,

1. sample with replacement from $\{1, \ldots, n\}$ and get $i_1^*, \ldots, i_n^*$;
2. for $j \in \{1, \ldots, n\}$, set $\{y_j^*(\omega), \mathbf{x}_j^*\} = \{y_{i_j^*}(\omega), \mathbf{x}_{i_j^*}\}$; then,
3. apply our estimation procedure to $\{y_1^*(\omega), \mathbf{x}_1^*\}, \ldots, \{y_n^*(\omega), \mathbf{x}_n^*\}$, giving estimates $\hat{A}_r^*(\omega|\mathbf{x})$.

For a fixed value of the covariates $\mathbf{x} = \mathbf{x}_0$, we compute pointwise $(1-\alpha)$-percentile bootstrap variability bands of the conditional Pickands' dependence function. This is performed by setting the interval lower and upper bounds at each angular observation ω_v, equal to the $\lfloor R(\alpha/2) \rfloor$ and $\lfloor R(1-\alpha/2) \rfloor$ ordered statistics of $\hat{A}_1^*(\omega_v|\mathbf{x}_0), \ldots, \hat{A}_R^*(\omega_v|\mathbf{x}_0)$, respectively.

Finally, we mention that the variability bands resulting from this approach are not guaranteed to be convex. However, we notice in simulations that this condition is violated only for very weak extremal dependences.

3.3.4 Extension to $d \geq 3$

The notion of extremal dependence for high-dimensional max-stable random vectors is often less intuitive than in the bivariate case. For instance, the Pickands' dependence function for $d = 5$ is defined on the four-dimensional unit simplex which makes its interpretation difficult. In such instances, scalar quantities are needed to help practitioners.

A useful way of summarizing the extremal dependence in high dimensions d is via the extremal coefficient. This quantity denoted η reflects the degree of tail dependence in a multivariate extreme value distribution and is related to the Pickands' dependence function A via the following relation

$$\eta = dA(1/d, \ldots, 1/d). \tag{3.12}$$

The extremal coefficient lies in the interval $[1, d]$, with 1 corresponding to perfect dependence and d to independence. Schlather and Tawn [2002, 2003] examine the consistency properties that must be verified by the set of extremal coefficients yielding a valid multivariate extreme value distribution.

Recall that when $d \geq 3$, conditions C1–C2–C3 are necessary but not sufficient for the Pickands' dependence function to be valid. Therefore, we focus on the estimation of extremal coefficients depending on the set of covariates $\mathbf{x} \in \mathcal{X}$. Using the same notations as before, the covariate-dependent extremal coefficient is estimated using the equation

$$\tilde{\eta}(\mathbf{x}) = d\tilde{A}(\boldsymbol{\omega}_0|\mathbf{x}), \tag{3.13}$$

where $\boldsymbol{\omega}_0 = (1/d, \ldots, 1/d)^\top$ is a d-dimensional vector belonging to the unit-simplex S_{d-1}, and $\tilde{A}(\cdot|\mathbf{x})$ is the first stage estimator defined in (3.8).

To force the extremal coefficient to belong to $[1, d]$ for all values of $\mathbf{x}$, we regularize $\tilde{\eta}(\mathbf{x})$ using constrained median B-splines. Recall that the first-stage estimator $\tilde{A}(\boldsymbol{\omega}_0|\mathbf{x})$ results from the generalized additive modelling of the max-projected Beta random variables $Y(\boldsymbol{\omega}_0)|\mathbf{x}$. Due to the penalized likelihood maximization step in the GAM fitting procedure, $\tilde{A}(\boldsymbol{\omega}_0|\mathbf{x})$ is a smooth function of the covariates $\mathbf{x}$. Thus, the regularization

step is performed by dropping the penalty term in the COBS minimization problem (3.9). Using the same notations as in Section 3.3.2, the regularization problem is now defined as

$$\begin{aligned} &\min_{\boldsymbol{\theta}=(a_1,\ldots,a_{N+3})^\top \in \mathbb{R}^{N+3}} \sum_{i=1}^{m} \left| \tilde{\eta}\left(\mathbf{x}_i\right) - \sum_{j=1}^{N+3} \boldsymbol{\theta}_j B_j(\mathbf{x}_i) \right| \\ &\text{s.t.} \quad \begin{pmatrix} B_1(\mathbf{x}_i) & \cdots & B_{N+3}(\mathbf{x}_i) \end{pmatrix} \boldsymbol{\theta} \geq 1, \\ &\qquad \begin{pmatrix} B_1(\mathbf{x}_i) & \cdots & B_{N+3}(\mathbf{x}_i) \end{pmatrix} \boldsymbol{\theta} \leq d, \quad \text{for } i \in \{1,\ldots,m\}, \end{aligned} \tag{3.14}$$

with m the number of distinct values of the covariates $\mathbf{x}$. Notice that the solution to the optimization problem (3.14) is now a function of $\mathbf{x}$, i.e.,

$$\hat{\eta}\left(\mathbf{x}\right) = \sum_{j=1}^{N+3} \hat{a}_j B_j(\mathbf{x}), \tag{3.15}$$

as opposed to the corrected Pickands' function estimator (3.11), which is a function defined on the unit simplex S_1 for a fixed value of $\mathbf{x}$.

Variability bands for the extremal coefficient estimates (3.15) are constructed straightforwardly based on (3.13) and bootstrapping techniques.

3.4 Simulation study

In this section, we assess the performance of our methodology by first studying the estimates obtained from the max-projection combined with the median smoothing regularization procedure (no covariate) and then including the generalized additive modelling in the presence of covariates. The parameters of our simulation study are chosen to mimic the range of values observed in the application in Section 3.5. Note that the uncertainty assessment in the subsequent simulation studies does not take into account the uncertainty resulting from the separate marginal fitting, or the use of a limiting model at a sub-asymptotic level.

3.4.1 No covariate dependence

Since we are in the case where the Pickands' dependence function does not depend on covariates, the estimation procedure of A_ω, for ω in $[0,1]$ boils down to maximizing the log-likelihood (3.7) of the Beta distributed random variables $\{y_i(\omega)\}_{i=1}^n$, where n is the number of observations at hand. Thus, the estimate of A_ω is given by

$$\tilde{A}_{\omega,\text{mle}} = -n \left[\sum_{i=1}^{n} \ln\left\{y_i(\omega)\right\} \right]^{-1}.$$

We now need to smooth the corresponding estimates $\tilde{A}_{\omega,\text{mle}}$ using the constrained median smoothing splines technique described in Section 3.3.2 to obtain a valid Pickands' function estimator $\hat{A}_{\omega,\text{mle}}$.

We perform a simulation study in which we consider two sample sizes $n = 200, 1000$ from a bivariate extreme value distribution with unit Fréchet margins and the Dirichlet

dependence model [Coles and Tawn, 1991]

$$V(z_1, z_2; \alpha, \beta) = \frac{1}{z_1}\left\{1 - \mathcal{B}\left(\alpha + 1, \beta; \frac{\alpha z_1}{\alpha z_1 + \beta z_2}\right)\right\} + \frac{1}{z_2}\left\{1 - \mathcal{B}\left(\alpha, \beta + 1; \frac{\alpha z_1}{\alpha z_1 + \beta z_2}\right)\right\}, \tag{3.16}$$

where α, $\beta > 0$ and

$$\mathcal{B}(\alpha, \beta; u) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^u w^{\alpha-1}(1 - w)^{\beta-1} dw,$$

a normalized incomplete beta function.

We fix $\tilde{n} = 800$ regularly spaced values of ω in $[0, 1]$ and compute $\hat{A}_{\omega,\text{mle}}$ for each value. Note that the estimation of the Pickands' function is sensitive to the choice of $\tilde{n}$, i.e., the size of the discretization of the unit simplex, particularly when dealing with asymmetric extremal dependence. Small discretization intervals, i.e., large values of $\tilde{n}$, allow us to capture more features of the Pickands' function. To have insight into the performance of the Pickands' function estimator for different levels of extremal dependence, we fix $\beta = 3$ and set different values for α that illustrate different extremal dependence schemes: weak, mild, and strong. In Figure 3.2, in both settings, our estimator is compared with the Pickands [1981] and Capéraà et al. [1997] corrected estimators. In the small sample size setting, our estimator outperforms the other estimators. When a larger sample size is considered, our estimator is competitive with, if not slightly better than, the other estimators.

3.4.2 Covariate-dependent Pickands' function

We now consider two cases in which the extremal dependence depends on a set of covariates. In the first setup, we focus on the estimation of the bivariate logistic dependence model [Tawn, 1990] with the exponent measure

$$V(z_1, z_2; \rho) = \left(z_1^{-1/\rho} + z_2^{-1/\rho}\right)^{\rho},$$

and where the parameter $\rho \in (0, 1)$ depends on a set of covariates through

$$\rho(t, x) = \tau + \mathbf{1}_{\{x=\text{“Level 1”}\}} a_0 \sin(a_1 t) + \mathbf{1}_{\{x=\text{“Level 2”}\}} \left\{b_0 \sin(b_1 t^2) + b_2\right\}, \tag{3.17}$$

with $\tau = 0.1$; $(a_0, a_1) = (0.8, 0.8)$; $(b_0, b_1, b_2) = (0.5, 0.21, 0.35)$; t is a continuous covariate (time, say) over the range 0.3 to 3.8 discretized into 12 equally spaced values, and x is a factor covariate with two levels "Level 1" and "Level 2". The specific form of covariate-dependence (3.17) is illustrated in Figure 3.3 and is ecologically motivated. If t represents time in months, then we can observe that for each level, the extremal dependence in December and January is the same. Moreover, (3.17) allows us to assess the performance of our estimator in the different cases of weak, mild, and strong extremal dependence. We consider two sample sizes ($n_c = 30, 400$) for each distinct pair of covariates from the bivariate logistic extreme value distribution with the corresponding dependence parameter $\rho(t, x)$, i.e., two samples of size $n = n_c \times 2 \times 12$ ($= 720, 9600$) in total. We suppose that we have no prior knowledge of the form of dependence of the Pickands' function on the available covariates and fit the following model to the Pickands' function

$$A_\omega(t, x; \boldsymbol{\theta}) = \mathbf{1}_{\{x=\text{“Level 1”}\}}\{L_1 + h_{L_1}(t)\} + \mathbf{1}_{\{x=\text{“Level 2”}\}}\{L_2 + h_{L_2}(t)\}, \quad \omega \in [0, 1]. \tag{3.18}$$

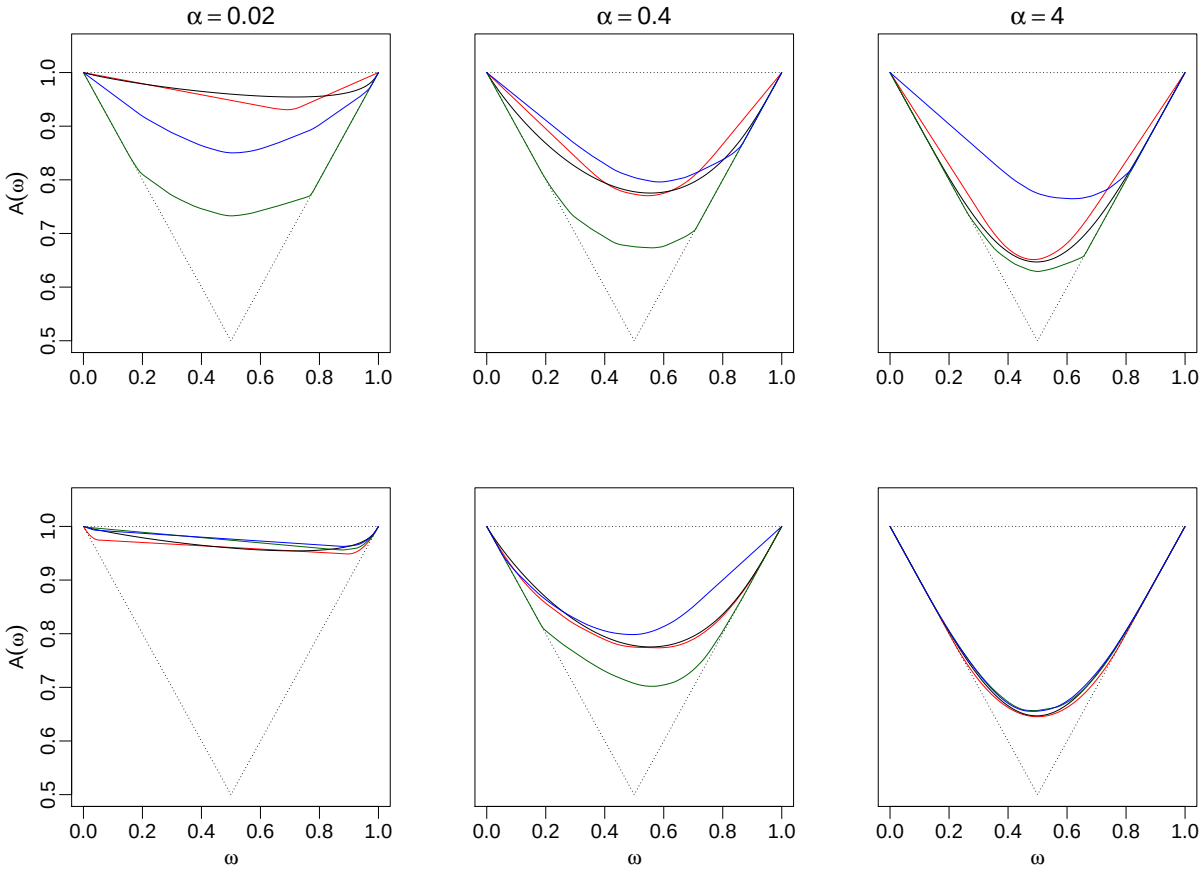

Figure 3.2: Inference of the Pickands' dependence function for the Dirichlet model with $\beta = 3$ and different values of α with three different estimators: our proposed estimator (in red), the Pickands [1981] corrected estimator (in green), and the Capéraà et al. [1997] corrected estimator (in blue). The true value of the Pickands' function is shown in solid black lines. Each row represents a different sample size: $n = 200$ (top panels) and $n = 1000$ (bottom panels).

In the second setup, we study the asymmetric Dirichlet extremal dependence (3.16) where we fix $\beta = 3$ and let the parameter α depend linearly on a continuous covariate $\tilde{t}$ as follows

$$\alpha(\tilde{t}) = 0.02 \times \tilde{t}, \quad \tilde{t} \in [1, 50] \cup [100, 350].$$

We discretize the intervals $[1, 50]$ and $[100, 350]$ into 40 and 15 equally spaced values, respectively. This choice of value range for $\tilde{t}$ allows the pairs of parameters $(\alpha(\tilde{t}), \beta)$ to produce different scenarios of extremal dependence. For each of the 55 selected values of $\tilde{t}$, we take two samples of size n_c ($n_c = 12, 175$), i.e., we consider two sample sizes of $n = n_c \times 55$ ($= 660, 9625$) in total. Again, with no prior knowledge of the form of dependence on the covariate $\tilde{t}$, we fit the following smooth model to the Pickands' function

$$A_\omega(\tilde{t}; \boldsymbol{\theta}) = a + h(\tilde{t}), \quad \omega \in [0, 1]. \tag{3.19}$$

Table 4.1 summarizes the fitted models (3.18) and (3.19) at the value $\omega = 0.5$.
The models are fitted reasonably well to the data. In both GAM fittings, we do not fix the basis dimension of the smooth terms but instead estimate it from the data. Therefore,

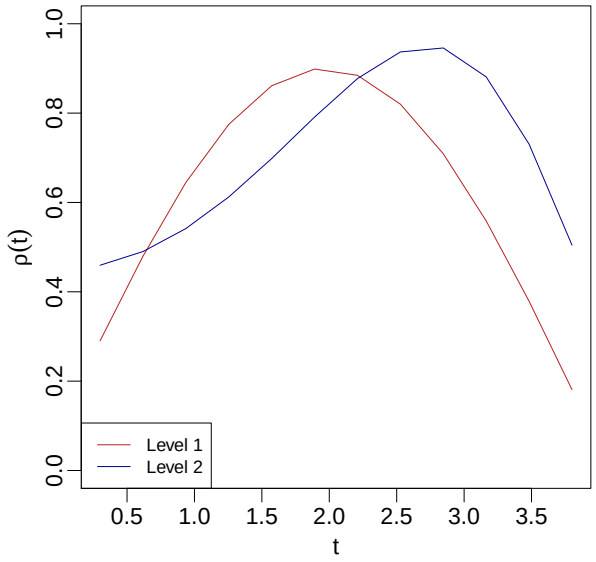

Figure 3.3: Covariate-dependence of the logistic dependence parameter $\rho(t, x)$ (3.17).

			L_1	L_2	h_{L_1}	h_{L_2}	a	h
Logistic	$n = 720$	estimate/edf	0.81 (0.04)	0.83 (0.04)	3.04	2.14		
		p-value	$< 10^{-16}$	$< 10^{-16}$	0.004	0.003		
	$n = 9600$	estimate/edf	0.76 (0.01)	0.81 (0.01)	3.12	3.53		
		p-value	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	2.59×10^{-12}		
Dirichlet	$n = 660$	estimate/edf					0.75 (0.03)	1
		p-value					$< 10^{-16}$	4.16×10^{-6}
	$n = 9625$	estimate/edf					0.75 (0.007)	5.24
		p-value					$< 10^{-16}$	$< 10^{-16}$

Table 3.1: Estimates (se) of the intercepts and the degrees of freedom (edf) of the smooth functions in the models (3.18) and (3.19) with $\omega = 0.5$.

if the dependence on a covariate is linear, for example, then the degree of freedom of the smooth term will be estimated at 1.

In both setups, we take $R = 1000$ bootstrap samples and fix $\tilde{n} = 800$ regularly spaced values in the unit simplex. For illustration purposes, we focus on the estimation of the Pickands' dependence function for selected values of the covariate t (logistic dependence) and the covariate $\tilde{t}$ (Dirichlet dependence) that produce different strengths of extremal dependence. Moreover, we predict in the Dirichlet dependence case the Pickands' dependence function for the unobserved value $\tilde{t} = 55$.

Figures 3.4 and 3.5 display, for the different values of t and both levels, the Pickands' function estimate and its 95% bootstrap variability bands in the logistic case and for the two sample sizes $n = 720$ and $n = 9600$, respectively.

For the different sample sizes, we observe that regardless of the strength of the extremal dependence, the variability bands always contain the true Pickands' function. A comparison of Figure 3.4 and Figure 3.5 shows that the noticeable bias and large variance observed when $n = 720$ are reduced when a larger sample size is considered ($n = 9600$). The bias reduction is more striking when approaching extremal independence whereas the

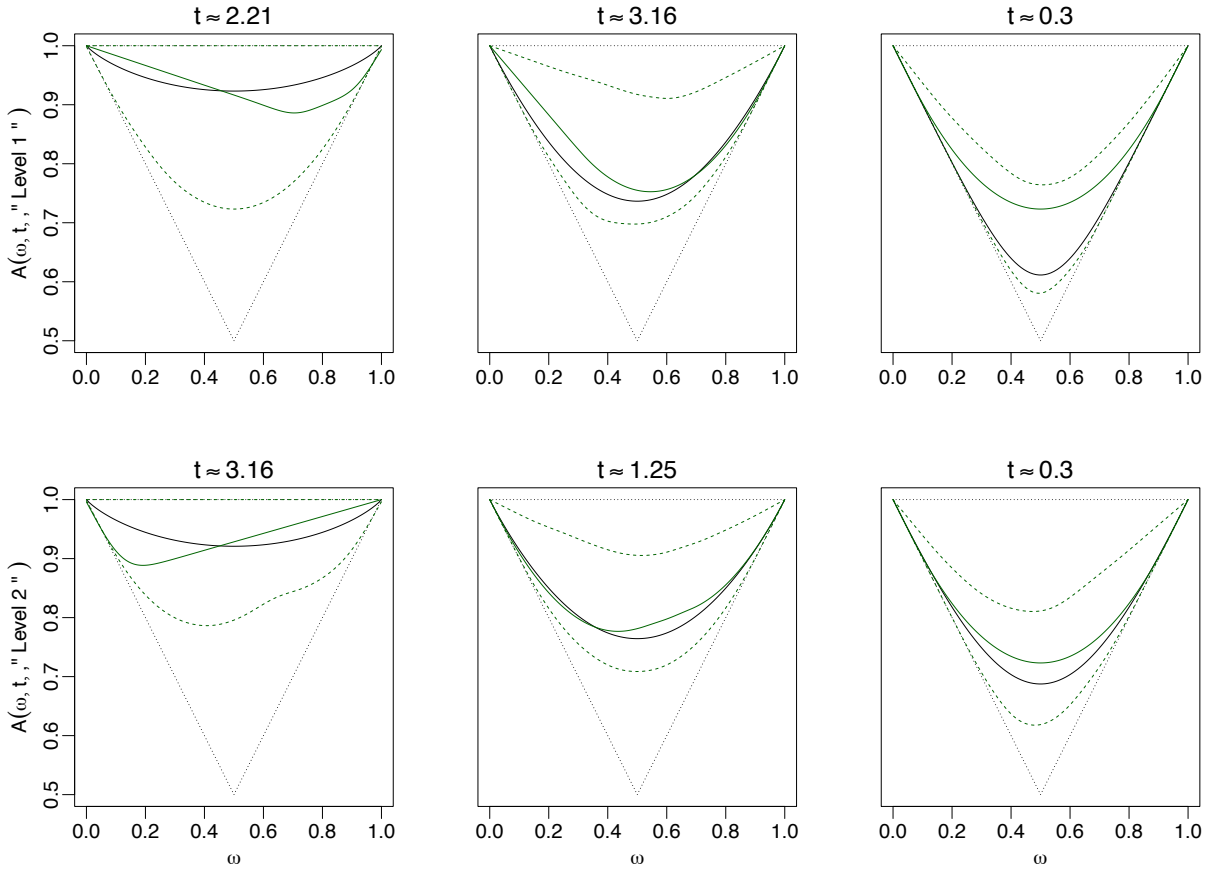

Figure 3.4: Estimates (solid green lines) of the Pickands' function along with the 95% bootstrap variability bands (dashed green lines) for the logistic model with weak, mild, and strong dependence (from left to right) for "Level 1" (top panels) and "Level 2" (bottom panels). The true value of the Pickands' function is represented by solid black lines. The simulation is performed with the sample size $n = 720$.

reduction in variability is notable at the different levels of extremal dependence.
As explained in Section 3.3.3, the bootstrap variability bands are not guaranteed to be convex. However, Figure 3.4 and Figure 3.5 show that the convexity of the variability bands is almost always verified, except for very weak extremal dependences.

Figure 3.6 displays for the different values of $\tilde{t}$, the Pickands' function estimate and the 95% bootstrap variability bands in the Dirichlet case with sample sizes $n = 660, 9625$.

As for the logistic dependence case, we observe that the variability bands contain the true Pickands' functions for the different levels of extremal dependence but are relatively wide when $n = 660$. A noticeable improvement in the bias and the variance is worth mentioning when a larger sample size is considered. The estimation procedure seems to work well at predicting the Pickands' function at unobserved values. For instance, for $\tilde{t} = 55$, we observe a small bias with both sample sizes and a striking variance reduction when $n = 9625$.

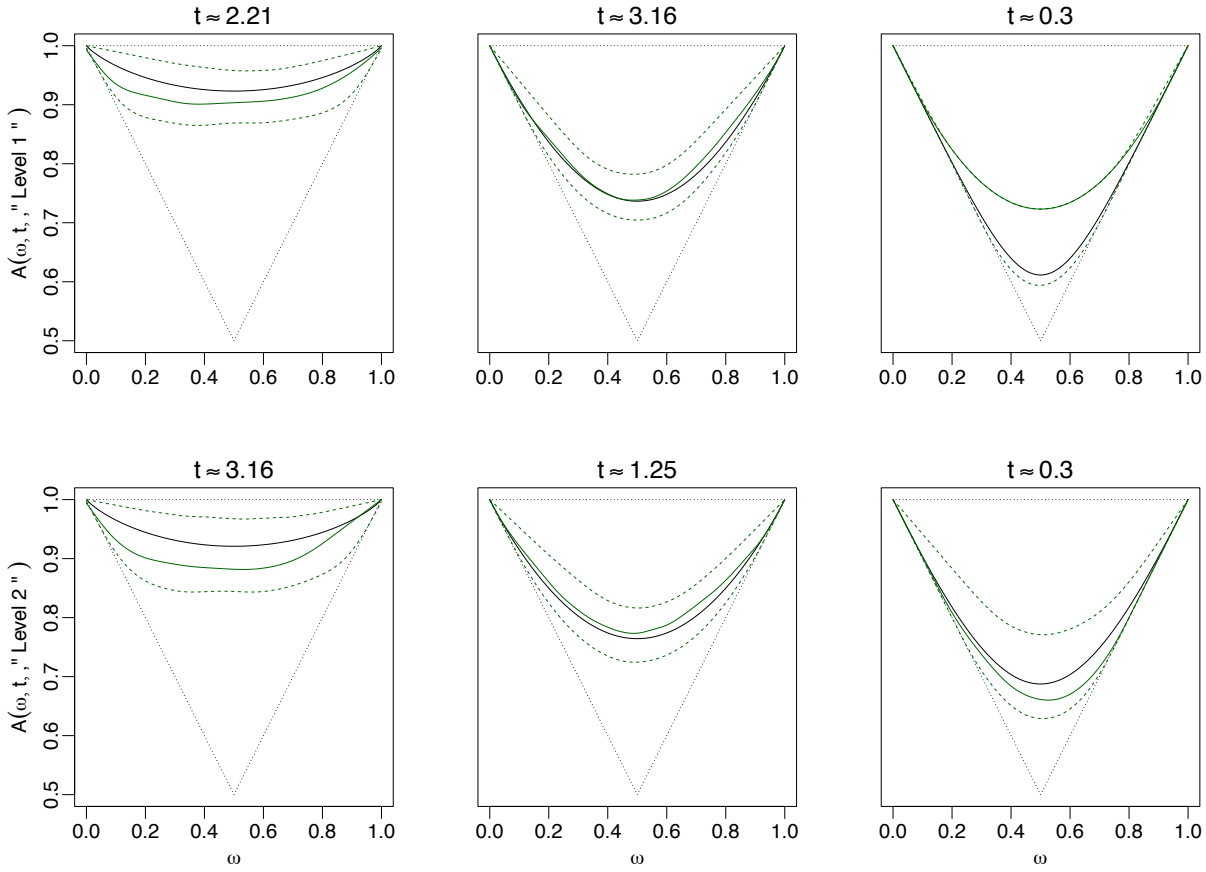

Figure 3.5: Same as Figure 3.4 but with $n = 9600$.

3.4.3 Multivariate covariate-dependent extremal coefficient

We illustrate the proposed methodology for estimating non-stationary extremal coefficients in the case of the symmetric logistic model [Tawn, 1990] with the exponent measure

$$V(z_1, \ldots, z_d; \rho) = \left(\sum_{i=1}^{d} z_i^{-1/\rho} \right)^{\rho}, \quad \rho \in [0, 1].$$

It can easily be proven that the extremal coefficient can be written as $\eta(\mathbf{x}) = d^{\rho(\mathbf{x})}$. The simulation study is performed under the dependence dynamic $\rho(t) = \exp(t)/\{1 + \exp(t)\}$, where $t \in [-1, 2]$, allowing for different strengths of extremal dependence. We consider $m = 20$ distinct values of t, two sample sizes $n = 150 \times m$ and $n = 1000 \times m$, and $R = 500$ bootstrap samples. In Figure 3.7, the theoretical extremal coefficients are compared to the estimated ones along with their 95% bootstrap variability bands, with different values of the dimension d. The proposed estimation methodology provides very reasonable estimates of the extremal coefficient with narrow variability bands containing the true values, even in the smaller sample size scenario.

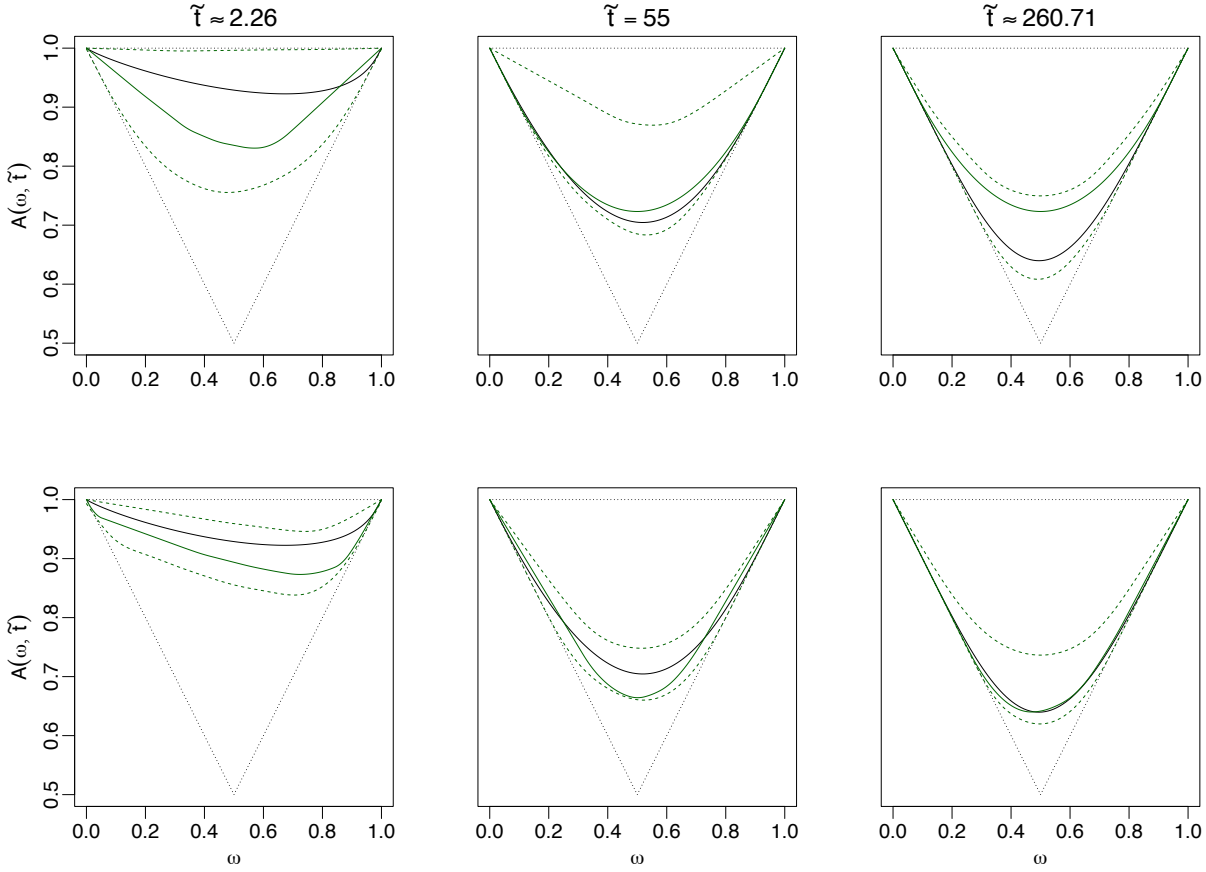

Figure 3.6: Estimates (solid green lines) of the Pickands' function along with the 95% bootstrap variability bands (dashed green lines) for the Dirichlet model with parameters $(\alpha(\tilde{t}), 3)$ allowing for weak, mild, and strong dependence (from left to right). The true value of the Pickands' function is represented by solid black lines. Each row represents a different sample size: $n = 660$ (top panels) and $n = 9625$ (bottom panels).

3.5 Application to U.S. temperature data

In this section, we investigate the dependence between hot days and warm nights in a non-stationary way, i.e., the dynamics of extremal dependence between maximum and minimum temperatures. Daily temperature maxima and minima measured at more than 5000 weather stations across the U.S. during the period 1950–2004 were extracted from the U.S. Cooperative Observer Program (COOP) and are shown in Figure 3.8. The dataset can be freely downloaded from NCAR [2016].

The first part of the analysis focuses on modelling a time-varying extremal dependence between the daily maxima and daily minima temperatures in Colorado. We consider the monthly maxima of daily maxima $W_{\max}$ and daily minima $W_{\min}$ taken over the 151 stations in Colorado. Figure 3.9 shows the time series of these monthly maxima. As expected, both time series exhibit seasonality. Thus, we fit marginally non-stationary generalized extreme value distributions where regression models on the location, scale, and shape parameters are considered in order to take into account the seasonality resulting from the choice of the block size, as well as the possible long-term trend [Chavez-Demoulin and Davison, 2012, Coles, 2001, Davison and Huser, 2015, Davison and Smith, 1990, Katz et al.,

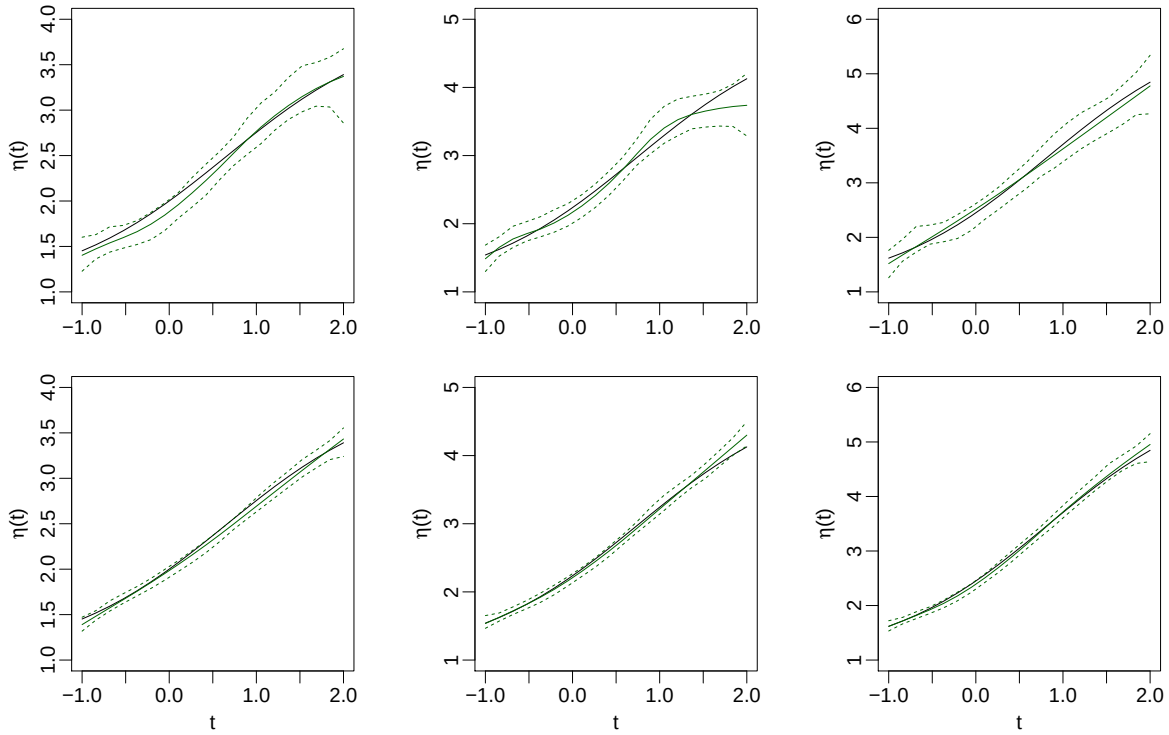

Figure 3.7: Estimates (solid green lines) of the extremal coefficient for the symmetric logistic model with parameter $\rho(t)$ and $d = 4$ (left panels), 5 (middle panels), and 6 (right panels), along with the 95% bootstrap variability bands (dashed green lines). The true extremal coefficient is represented in solid black lines. Each row represents a different sample size: $n = 3000$ (top panels) and $n = 20000$ (bottom panels).

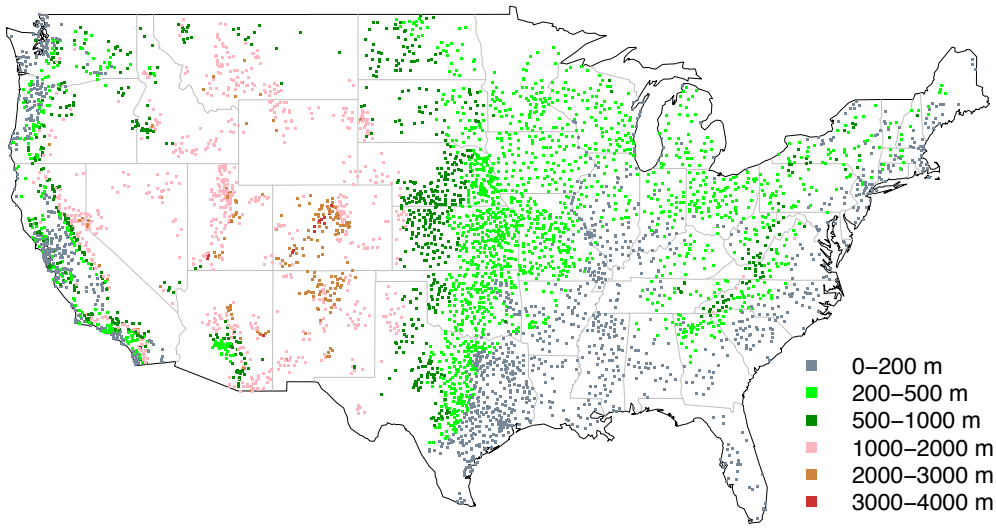

Figure 3.8: Monitoring stations available from the U.S. COOP. The colour scheme is chosen according to the altitude levels in Table 3.3.

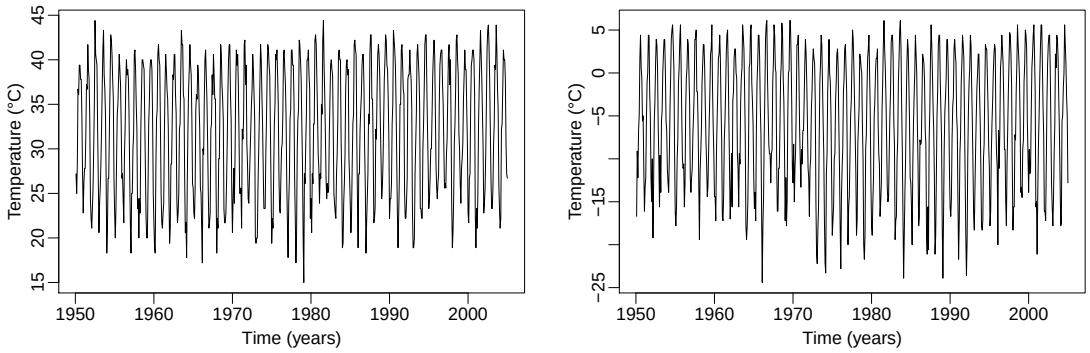

Figure 3.9: Monthly maxima of daily maximum (left) and daily minimum (right) temperatures in Colorado.

2002]. The location, scale, and shape parameters of the non-stationary GEVs fitted to the monthly maxima of the daily maxima and daily minima are denoted $(\mu_{\max}, \sigma_{\max}, \xi_{\max})$ and $(\mu_{\min}, \sigma_{\min}, \xi_{\min})$, respectively. Different non-stationary models with constant shape parameters are considered. A forward model selection is performed marginally based on likelihood ratio tests. In the first step, the model for the location parameter is chosen while the scale parameter is constant. Then, the model for the scale parameter is chosen while the location parameter is fixed at the selected model. The fitted model is the following

$$\mu_{\max}(t,m) = \mu_0 t + \mu_1 \mathbf{1}_{\{m=1\}} + \sum_{\ell=2}^{12} (\mu_1 + \mu_\ell)\mathbf{1}_{\{m=\ell\}}, \qquad \mu_{\min}(t,m) = \tilde{\mu}_1 \mathbf{1}_{\{m=1\}} + \sum_{\ell=2}^{12} (\tilde{\mu}_1 + \tilde{\mu}_\ell)\mathbf{1}_{\{m=\ell\}},$$

$$\sigma_{\max}(t,m) = \sigma_1 \mathbf{1}_{\{m=1\}} + \sum_{\ell=2}^{12} (\sigma_1 + \sigma_\ell)\mathbf{1}_{\{m=\ell\}}, \qquad \sigma_{\min}(t,m) = \tilde{\sigma}_1 \mathbf{1}_{\{m=1\}} + \sum_{\ell=2}^{12} (\tilde{\sigma}_1 + \tilde{\sigma}_\ell)\mathbf{1}_{\{m=\ell\}},$$

$$\xi_{\max}(t,m) = \xi_{\max}, \qquad \xi_{\min}(t,m) = \xi_{\min},$$

where t represents the years (standardized between 0 and 1) and m the months. Parameter estimates along with their standard errors are reported in Table 3.2.

m	0	1	2	3	4	5	6	7	8	9	10	11	12
μ_m	20.56	0.02	1.66	5.97	10.55	14.03	18.40	19.64	18.08	15.96	11.41	4.29	0.30
(se)	(0.37)	(0.004)	(0.51)	(0.47)	(0.45)	(0.44)	(0.44)	(0.39)	(0.39)	(0.40)	(0.42)	(0.46)	(0.48)
σ_m	–	2.55	0.12	−0.24	−0.44	−0.59	−0.64	−1.34	−1.25	−1.19	−0.85	−0.34	−0.14
(se)	–	(0.22)	(0.32)	(0.30)	(0.28)	(0.28)	(0.27)	(0.24)	(0.25)	(0.25)	(0.27)	(0.30)	(0.30)
$\xi_{\max}$							$-0.30_{(0.03)}$						
$\tilde{\mu}_m$	–	−18.20	2.19	6.42	11.34	15.98	19.26	21.86	21.35	18.20	13.54	7.96	2.23
(se)	–	(0.57)	(0.73)	(0.69)	(0.61)	(0.60)	(0.59)	(0.59)	(0.60)	(0.60)	(0.62)	(0.65)	(0.70)
$\tilde{\sigma}_m$	–	4.15	−0.75	−1.31	−2.45	−2.63	−2.92	−2.98	−2.73	−2.74	−2.30	−1.83	−1.15
(se)	–	(0.37)	(0.45)	(0.45)	(0.40)	(0.37)	(0.38)	(0.38)	(0.39)	(0.38)	(0.38)	(0.42)	(0.45)
$\xi_{\min}$	–						$-0.30_{(0.03)}$						

Table 3.2: Parameter estimates and their standard errors (se) for the monthly maxima of the daily maximum and daily minimum temperatures in Colorado.

Diagnostic plots are produced by transforming the data $w_{\max}(t,m)$ and $w_{\min}(t,m)$ to

a standardized Gumbel distribution using the following transformations

$$g_{\max}(t,m) = \frac{1}{\hat{\xi}_{\max}(t,m)} \ln\left[1 + \hat{\xi}_{\max}(t,m)\left\{\frac{w_{\max}(t,m) - \hat{\mu}_{\max}(t,m)}{\hat{\sigma}_{\max}(t,m)}\right\}\right],$$
$$g_{\min}(t,m) = \frac{1}{\hat{\xi}_{\min}(t,m)} \ln\left[1 + \hat{\xi}_{\min}(t,m)\left\{\frac{w_{\min}(t,m) - \hat{\mu}_{\min}(t,m)}{\hat{\sigma}_{\min}(t,m)}\right\}\right].$$

The resulting QQ plots are displayed in Figure 3.10, and the fitted models appear to capture the non-stationarity reasonably well in the margins. Thus, once the data $w_{\max}(t,m)$

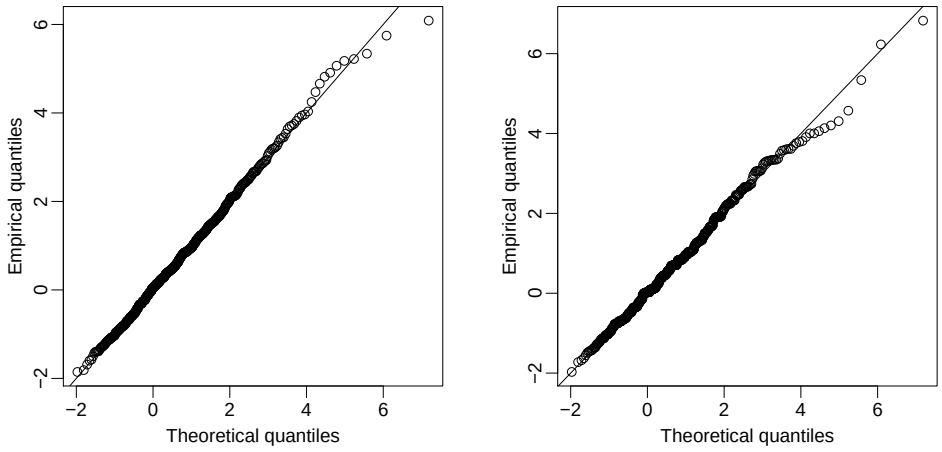

Figure 3.10: Gumbel marginal QQ plots for the monthly maxima of the daily maximum (left) and daily minimum (right) temperatures in Colorado.

and $w_{\min}(t,m)$ are transformed to unit Fréchet, we can proceed by assuming that the non-stationarity left in our data is at the extremal dependence level. We model a time-varying dependence between the monthly maxima of the daily maximum and daily minimum temperatures in Colorado. We fix the covariate to be the time in years and set $\tilde{n} = 300$ points in the unit simplex. Figure 3.11 displays the estimate of the non-stationary Pickands' dependence function as a function of time (in years) and the angular observations $\omega \in [0,1]$. Different levels of extremal dependence can be observed through time. In particular, the observations start to exhibit extremal dependence from around 1965 until 1985. These observations are supported by the findings in Figure 3.12 where cross-sections of $\hat{A}(\omega|t)$ at different years are displayed along with their 95% bootstrap variability bands. Throughout the data analysis, variability bands are constructed based on the block bootstrap using seasonal blocks, so that the dependence within the blocks is preserved.

Remark. Issues related to the convergence in distribution of the monthly-maxima could arise as the length of the considered block is small. Here, a two-monthly block maxima analysis, not shown for sake of brevity, confirms the stability of the results and the non-stationarity of the monthly-maxima appears not to be a major issue for the estimation of the extremal dependence structure.

To see whether the observed behaviour of the extremal dependence is region-specific, we perform the same data analysis in different states and compare the resulting estimates of the non-stationary extremal coefficient. Recall that the extremal coefficient (3.12) summarizes the strength of the extremal dependence with a value of 1 corresponding to

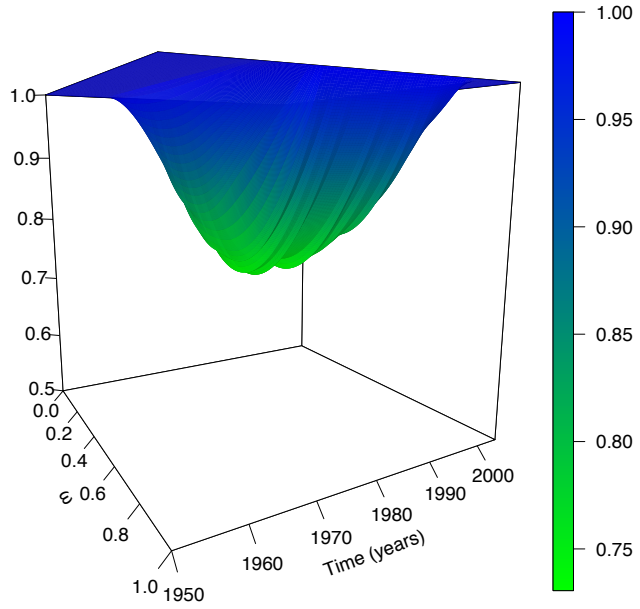

Figure 3.11: Estimate of the time-dependent Pickands' dependence function $\hat{A}(\omega|t)$ for temperatures in Colorado. For sake of visibility, the variability bands are not shown.

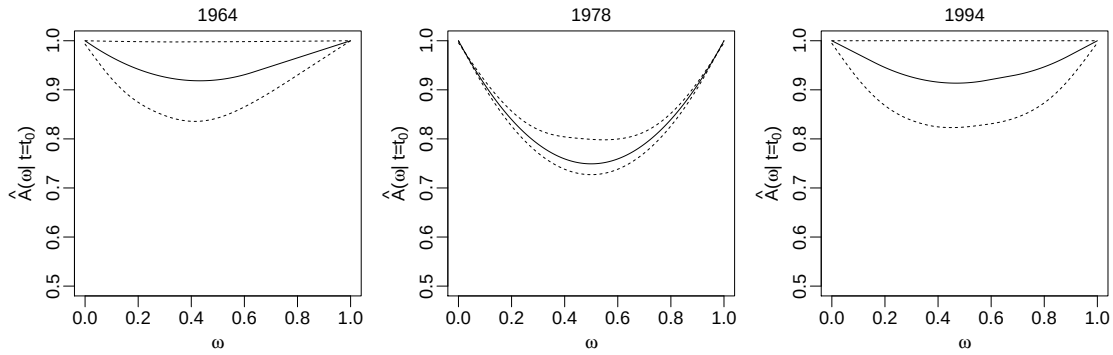

Figure 3.12: Estimates of $A(\omega|t = t_0)$, with $t_0 = 1964, 1978$, and 1994 in Colorado (solid lines). The 95% bootstrap variability bands (dotted lines) are based on 300 bootstrap samples.

perfect dependence and a value of 2 to independence. The estimates of the non-stationary extremal coefficients are shown in Figure 3.13.

The graphs confirm that the global behaviour of the extremal dependence between 1965 and 1985, observed in the state of Colorado, is common to many states. However, the shape of the extremal dependence during the entire period of study is quite different from one state to another. One can argue that the shape of the estimated extremal coefficients might be caused by boundary effects in the GAM fitting. Therefore, we estimate the extremal dependence during three different non-overlapping time periods. Figure 3.14 displays the estimates of the extremal coefficient in Colorado during the time periods 1950–1965, 1966–1985, and 1986–2004. The behaviour of the extremal coefficient during these time periods is consistent with the findings shown in Figure 3.13, and therefore, we eliminate the possibility of boundary effects, as far as the GAM fitting is concerned.

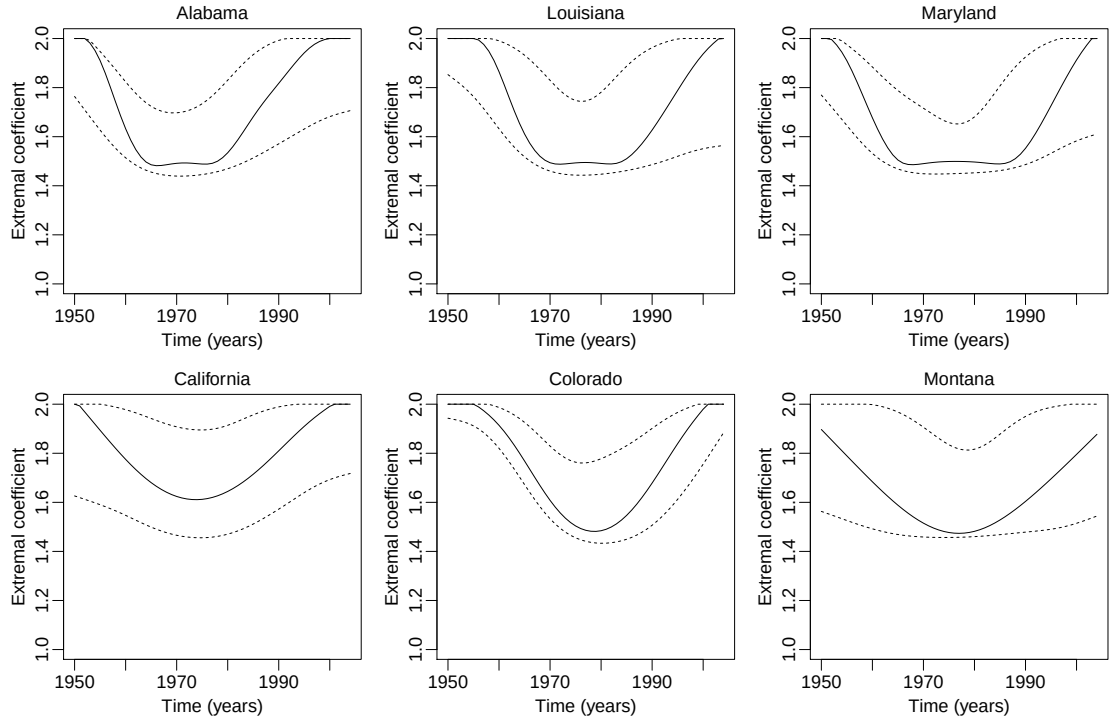

Figure 3.13: Extremal coefficient estimates (smoothed) with the 95% bootstrap variability bands based on 300 bootstrap samples.

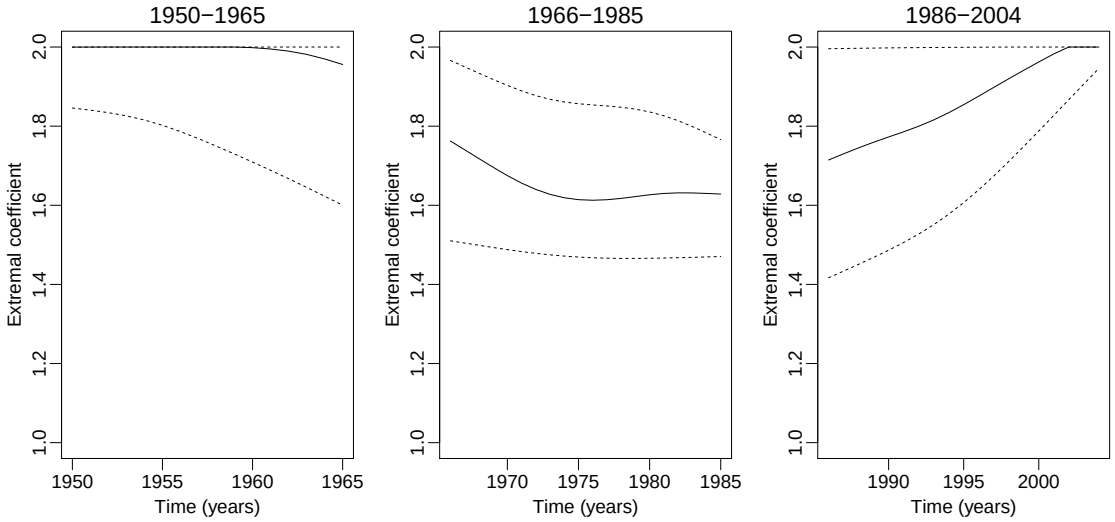

Figure 3.14: Extremal coefficient estimates (smoothed) in Colorado during different time periods. The 95% bootstrap variability bands (dashed lines) are based on 300 bootstrap samples.

A closer examination of the findings in Figure 3.13 shows that the general behaviour of the extremal dependence of the pair of daily maxima and minima temperatures through time might depend on the altitude. We observe that the extremal coefficient behaves similarly through time in states with low altitudes, such as Alabama, Louisiana, and Maryland, while exhibiting a different shape in the states dominated by high-altitude regions, such as California, Colorado, and Montana. Thus, we decide to model the extremal dependence across all of the 5314 weather stations in the U.S. as a function of time and

altitude, which leads us to the second part of the analysis.

The altitudes of the weather stations vary from 0 to 3487 m and we decide to fix six levels and consider the altitude as a categorical covariate in the GAM for the Pickands' function. Table 3.3 shows the different categories for the altitude along with the number of stations in each category.

	Altitude (m)	number of stations
1	0–200	1443
2	200–500	1973
3	500–1000	833
4	1000–2000	837
5	2000–3000	211
6	3000–4000	17

Table 3.3: Altitude levels along with the number of weather stations at each level.

The daily maximum and minimum temperatures in each altitude category are taken over all the stations in that category. Again, we model the marginal monthly maxima for a given level of altitude using non-stationary GEV distributions and transform these margins to be unit Fréchet before we apply our estimation procedure. As mentioned previously, the extremal dependence of the temperatures is now modelled as a function of time (in years) and level of altitude, as fixed in Table 3.3. Figure 3.15 displays the resulting estimates of the non-stationary extremal coefficient.

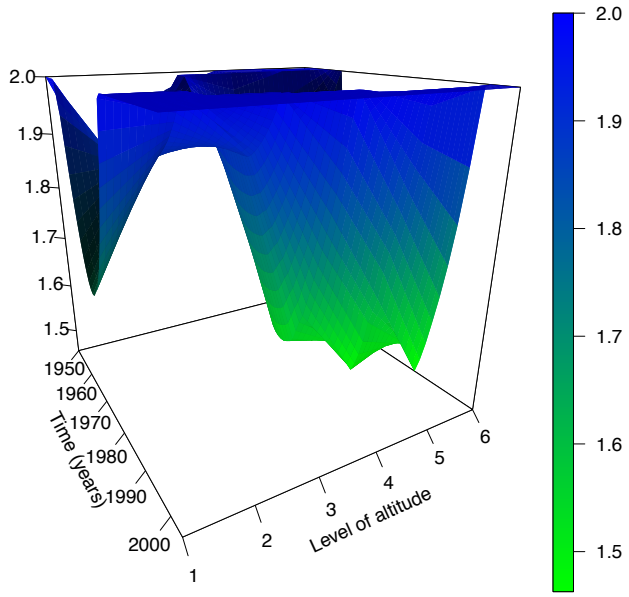

Figure 3.15: Estimates of the time- and altitude-dependent extremal coefficient. For sake of visibility, the variability bands are not shown.

The findings in Figure 3.15 confirm that the extremal dependence of the daily maxima and minima temperatures is altitude-dependent. The substantial changes in the extremal behaviour occurring around 1970 are seemingly more severe at low and high altitudes than

at moderate altitudes. At this stage, we do not have a physical mechanism to explain these climatological features, and future work with atmospheric scientists is needed to determine the causes of our findings.

3.6 Discussion

Recently, many parametric families of stationary max-stable models have been developed. For non-stationary data, we argue that classical non-linear regression techniques, like the GAM, can be adapted to max-stable dependence structures, and we incorporate multiple covariates of any type (continuous, factors, etc.). Our methodology could help to tackle environmental problems in which the dynamics of extreme events are often captured using covariates, such as altitude, longitude, latitude, and time. Moreover, our approach allows for any seasonal or within-year behaviour of extremes to be naturally embedded in the modelling of the extremal dependence. Our method is based on a max-projection technique, allowing an $\boldsymbol{\omega}$-sequential modelling of the Pickands' dependence function. The resulting estimator is then regularized in the bivariate case with a median smoothing technique and shown, in a simulation study, to be consistent in the stationary and non-stationary cases. In the multivariate context with $d > 2$, the max-projection is still convenient in the non-stationary case because the GAM framework is applied at the univariate level, leading to high computational stability based on a fast and convergent Newton–Raphson algorithm commonly found in all statistical software. The smoothing technique step is then applied to obtain valid estimates of non-stationary extremal coefficients, which in a high-dimensional context are more appropriate for assessing the degree of extremal dependence.

Chapter 4

Exceedance-based non-linear regression of tail dependence

As seen in Section 1.1.2.5, when the assumption of asymptotic dependence fails to hold, models for the decay rate of the dependence in the joint tails represent an alternative to the (classical) asymptotic theory of multivariate extremes. Therefore, additionally to the Pickands' dependence function characterizing the structure of the extremal dependence within the class of asymptotically dependent models, the coefficient of tail dependence and its multivariate extension, the angular dependence function, describe the tail dependence within the class of asymptotically independent models. We are interested in this chapter in situations where the probability and structure of co-occurrences of extreme events in multivariate data may critically depend on auxiliary information provided by covariates. We develop a flexible generalized additive modelling framework based on high threshold exceedances for estimating covariate-dependent joint tail characteristics for regimes of asymptotic dependence and asymptotic independence, i.e., covariate-dependent Pickands' dependence functions and covariate-dependent angular dependence functions, respectively. The framework is based on suitably defined marginal pretransformations and projections of the random vector along the directions of the unit simplex, which lead to convenient univariate representations of multivariate exceedances based on the exponential distribution. The link between the proposed projection in the asymptotic dependence case, and the max-projection of max-stable random vectors, defined in Chapter 3, is established. Good performance of the new estimators of a non-parametrically designed covariates' influence on extremal coefficients and tail dependence coefficients is shown through a simulation study. We illustrate the usefulness of our modelling framework on a large dataset of nitrogen dioxide measurements recorded in France between 1999 and 2012, where we use the generalized additive framework for modelling marginal distributions and tail dependence in monthly maxima. Our results imply asymptotic independence of data observed at different stations, and we find that the estimated coefficients of tail dependence decrease as a function of spatial distance and show distinct patterns for different years and for different types of stations (traffic vs. background).

4.1 Introduction

Modelling co-occurrence patterns of extreme values arising in multi-component systems is crucial for an accurate prediction of aggregated risks. The limiting dependence structures for extreme values do not present a simple parametric form, such as the Gaussian

dependence arising from the multivariate central limit theorem, which has spawned an extensive literature covering a wide range of parametric to fully nonparametric dependence models. When asymptotic independence arises, the limit model has a simple form but does not indicate the rate of convergence to this limit, suggesting alternative joint tail representations should be used to model the residual dependence at the observed levels of the process. In this work, we use threshold exceedance data, and we develop nonparametric modelling approaches suitable for characterizing asymptotic dependence and asymptotic independence when the strength of the dependence and more generally its shape may be governed by additional information given by covariate data. Only a few approaches exist in the current literature for modelling and estimating covariate influence on the joint tail structure. In a parametric framework, Mhalla et al. [2017b] proposed integrating covariate information through the parameters indexing angular density models, but we want to avoid such strong parametric assumptions. In the asymptotic dependence case, a nonparametric approach based on a baseline density for the dependence, modified through a density ratio to obtain a set of different dependence models according to covariate information, was developed by de Carvalho and Davison [2014]. More recently, Mhalla et al. [2017a] proposed very flexible modelling of the extremal dependence based on a generalized additive model with shape constraints using blockwise maxima data. In this work, we extend their approach from maxima to threshold exceedance data, and we lift the restriction on asymptotic dependence by proposing a set of tools that work for both asymptotic dependence classes.

Classical extreme value theory and practice are founded on max-stable processes, which are the only non-trivial limits arising from pointwise normalized maxima taken over independent replicates of a continuous stochastic process $\mathbf{X} = \{X(s)\}_{s\in\mathcal{S}}$, with $\mathcal{S}$ a set of indexes. When observed over a finite-dimensional set, the d-dimensional joint distribution $F_{\mathbf{Z}}$ of a max-stable random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)$ is of multivariate extreme value type with marginal generalized extreme value distributions and a max-stable dependence structure [de Haan and Ferreira, 2006, Section 9.2]. A handy representation of $F_{\mathbf{Z}}$ is obtained when the marginal distributions are transformed to a common unit Fréchet distribution $F_{Z_i}(z) = \exp(-1/z)\,\mathbb{1}_{[0,\infty)}(z)$. The multivariate max-stable distribution is then written as $F_{\mathbf{Z}} = \exp\{-V(\mathbf{z})\}\,\mathbb{1}_{[0,\infty)^d}(\mathbf{z})$ with the *exponent function* V measuring the strength and form of the dependence; the max-stable process $\mathbf{Z}$ is then said to be simple. We will assume this property throughout, without loss of generality. The convergence of the extremal dependence of any distribution F in the domain of attraction of $F_{\mathbf{Z}}$ is characterized by the property of multivariate regular variation [Resnick, 1987]:

$$t\{1 - F(t\mathbf{z})\} \to V(\mathbf{z}) = -\log F_{\mathbf{Z}}(\mathbf{z}), \qquad \mathbf{z} > \mathbf{0}, \qquad t \to \infty. \tag{4.1}$$

Max-stable models are useful when data are asymptotically dependent. When this assumption does not hold, limiting models are of little practical use as the property (4.1) fails to distinguish between asymptotic independence and exact independence [Heffernan and Resnick, 2007], and models for the joint tails based on hidden regular variation [Resnick, 2002] are indispensable.

Joint tail characterizations at the interface of asymptotic dependence and independence are often presented in a bivariate setup. In more than two dimensions, some pairs of components may be asymptotically dependent while others are not; for a deeper theoretical treatment, see de Haan and Zhou [2011]. We say here that a stochastic process is asymptotically independent if all of its pairs of components have this property; the only possible max-stable limit in this case is full independence. In view of our aim of modelling

the joint tail decay while abstracting away from the univariate marginal distributions, we now suppose that the d-dimensional random vector $\mathbf{X}^P = (X_1^P, \ldots, X_d^P)$ is nonnegative with Pareto marginal distributions by applying a marginal probability integral transform to $\boldsymbol{X}$ if necessary. The use of unit Fréchet margins (denoted by $\mathbf{X}^F = (X_1^F, \ldots, X_d^F)$), which are tail equivalent to X_i^P, would yield the same representations in the following. With asymptotic independence of X_1^P and X_2^P, their dependence strength vanishes as we move further into the joint tail. A general flexible representation of the joint tail, leading to a broad class of models suitable for asymptotic independence, was introduced by Ledford and Tawn [1996, 1997] for bivariate random vectors and then generalized to the multivariate setup by Wadsworth and Tawn [2013]. We denote by $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$ a *direction* (also called *weight* or *angle*) that is on the unit simplex $S_d = \{u \in \mathbb{R}_+^d \mid u_1 + \cdots + u_d = 1\}$ in $\mathbb{R}_+^d$. Any positive random vector $\mathbf{x} = (x_1, \ldots, x_d) > \mathbf{0}$ can be represented as $\mathbf{x} = (x^{\omega_1}, \ldots, x^{\omega_d}) = x^{\boldsymbol{\omega}}$, for some scalar $x > 0$ and $\boldsymbol{\omega} \in S_d$. The joint tail representation is

$$\Pr\left(X_1^P > x^{\omega_1}, \ldots, X_d^P > x^{\omega_d}\right) = \ell(x; \boldsymbol{\omega}) x^{-\lambda_{\boldsymbol{\omega}}}, \quad x \to +\infty, \tag{4.2}$$

where ℓ is a slowly varying function at infinity for any value of $\boldsymbol{\omega} \in S_d$ held fixed. The function $\lambda_{\boldsymbol{\omega}} \leq 1$ is called the angular dependence function [Wadsworth and Tawn, 2012, 2013]. It must satisfy certain shape constraints and describes the decay rate along rays in direction $\boldsymbol{\omega}$. In the case of asymptotic dependence, we have $\lambda_{\boldsymbol{\omega}} \equiv 1$ and $\ell(x; \boldsymbol{\omega}) \not\to 0$ as $x \to \infty$. If (4.1) holds, then $\ell(x; \boldsymbol{\omega})$ tends to a limit expressed through values of the exponent function V, which for $d = 2$ is $1/x^{\omega_1} + 1/x^{\omega_2} - V(1/x^{\omega_1}, 1/x^{\omega_2})$. In the bivariate case, $\lambda_{\boldsymbol{\omega}}$ generalizes the coefficient of tail dependence η introduced by Ledford and Tawn [1996] and the dependence measure $\overline{\chi}$ [Coles et al., 1999], where $\eta = (1+\overline{\chi})/2 = 1/\{2\lambda(1/2, 1/2)\}$ characterizes the joint tail decay rate along the diagonal of the first hyperoctant. The coefficient η can be defined in d dimensions as $1/\{d\lambda(1/d, \ldots, 1/d)\}$. In a similar way, the extremal coefficient $\theta \in [1, d]$ [Schlather and Tawn, 2003] is related to V via $\theta = V(1, \ldots, 1)$.

In the remainder of the chapter, Section 4.2 introduces projections based on weighted maxima and minima of multivariate random vectors leading to convenient univariate representations with appealing distributional properties for characterizing dependence in the tail. In Section 4.3, we develop nonparametric inference for such dependence based on a generalized additive modelling framework allowing the inclusion of covariate influence. The simulation study in Section 4.4 illustrates the good performance of our methods when the model is exact for the data but also when it represents an asymptotic approximation to the true data distribution. In the application presented in Section 4.5, we use our new techniques to reveal the influence of spatial distance and time on the co-occurrence patterns of extreme values in a large dataset of French air pollution data. Conclusions with an outlook on future work are given in Section 4.6.

4.2 Max- and min-projections

We define the notions of *max-projection* and *min-projection* of a vector $\mathbf{x} = (x_1, \ldots, x_d) \geq \mathbf{0}$ with respect to a weight vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in S_d$. The max-projection is given as $\max_{\boldsymbol{\omega}}(\boldsymbol{x}) = \max_{j=1}^d \omega_j x_j$, and the min-projection is defined as $\min_{\boldsymbol{\omega}}(\boldsymbol{x}) = \min_{j=1}^d x_j/\omega_j$. The link between the two projections is established through the inversion $\max_{\boldsymbol{\omega}}(\boldsymbol{x}) = 1/\min_{\boldsymbol{\omega}}(1/\boldsymbol{x})$ using the convention that $1/0 = \infty$ and $1/\infty = 0$.

4.2.1 Asymptotic dependence

We assume that the random vector $\mathbf{X}^F$ with distribution function F is in the max-domain of attraction of a simple max-stable process $\mathbf{Z}$. The exponent function V in (4.1) characterizing a max-stable random vector $\mathbf{Z}$ is positive, continuous, convex, and homogeneous of order -1 such that $V(t\mathbf{z}) = t^{-1}V(\mathbf{z})$ for $t > 0$. Exploiting the homogeneity of V, an alternative characterization of the extremal dependence is possible through the Pickands' dependence function A [Pickands, 1981], where

$$V(\mathbf{z}) = \left(\frac{1}{z_1} + \cdots + \frac{1}{z_d}\right) A\left(\frac{1/z_1}{1/z_1 + \ldots + 1/z_d}, \cdots, \frac{1/z_d}{1/z_1 + \cdots + 1/z_d}\right). \tag{4.3}$$

Given $\boldsymbol{\omega}$, (4.3) implies $V(1/\boldsymbol{\omega}) = A(\boldsymbol{\omega})$. We denote by $M_{\boldsymbol{\omega}}^{\max} = \max_{i=1}^{d} \omega_i Z_i$ the max-projection of the random vector $\mathbf{Z}$. Then $M_{\boldsymbol{\omega}}^{\max}$ is Fréchet distributed with scale parameter $V(1/\boldsymbol{\omega}) \leq 1$ reflecting the level of dependence in $\mathbf{Z}$ at an angle $\boldsymbol{\omega}$: For $\mathbf{z} > 0$, we get

$$\begin{aligned} \Pr\left(M_{\boldsymbol{\omega}}^{\max} \leq z\right) &= \Pr\left(Z_1 \leq \frac{z}{\omega_1}, \ldots, Z_d \leq \frac{z}{\omega_d}\right) \\ &= \exp\left\{-V\left(\frac{z}{\omega_1}, \ldots, \frac{z}{\omega_d}\right)\right\} = \exp\left\{-\frac{1}{z} V\left(\frac{1}{\omega_1}, \ldots, \frac{1}{\omega_d}\right)\right\}. \end{aligned} \tag{4.4}$$

From (4.3), we conclude that the scale parameter is equal to $A_{\boldsymbol{\omega}} := A(\boldsymbol{\omega})$, and we remark that the corresponding min-projection $M_{\boldsymbol{\omega}}^{\min} = 1/M_{\boldsymbol{\omega}}^{\max}$ follows an exponential distribution with rate $A_{\boldsymbol{\omega}}$. As argued by Ledford and Tawn [1997], a censored version of the multivariate max-stable distribution is a natural model for asymptotic dependent threshold exceedances. In view of convergence (4.1), this corresponds to applying the approximation $F(\mathbf{x}) = \exp\{-V(\mathbf{x})\} \approx 1 - V(\mathbf{x})$ for large values of $\|\mathbf{x}\|$ or equivalently, to replacing $1 - F(t\mathbf{z})$ by $-\log F(t\mathbf{z})$ using the logarithmic series approximation $\log(1-\varepsilon) \approx -\varepsilon$ for small $\varepsilon > 0$. We propose to consider projected data values $M_{\boldsymbol{\omega}}^{\min\downarrow} = 1/\max_{i=1}^{d} \omega_i X_i^F = \min_{i=1}^{d} X_i^{E\downarrow}/\omega_i$ using standard exponentially distributed $X_i^{E\downarrow} = 1/X_i^F$, where we use "$\downarrow$" to emphasize the tail inversion. As left-censoring the upper tail is equivalent to right-censoring the lower tail after inverting the tails, we model, below a small fixed threshold, the projected values by the $\text{Exp}(A_{\boldsymbol{\omega}})$ distribution, while we censor the values above the threshold.

Our max-projection $M_{\boldsymbol{\omega}}^{\max}$ is closely related to the max-projection $Y(\boldsymbol{\omega})$ used by Mhalla et al. [2017a] for estimating max-stable dependence from maxima data where no censoring is applied. They define

$$Y(\omega_1, \omega_2) = \max\left[\exp\{-1/(\omega_2 X_1)\}, \exp\{-1/(\omega_1 X_2)\}\right] \tag{4.5}$$

which corresponds to $\exp\{-\max(\omega_2 X_1, \omega_1 X_2)^{-1}\} = \exp\{-1/M_{(\omega_2,\omega_1)}^{\max}\}$. The link between the distribution functions of the two max-projections arises from the fact that $Z \sim \text{Exp}(\lambda)$ implies $\exp(-Z) \sim \text{Beta}(\lambda, 1)$ for a random variable Z. The components of the direction $\boldsymbol{\omega}$ in (4.5) are inversed as we define the Pickands' dependence function $A_{\boldsymbol{\omega}}$ differently from Mhalla et al. [2017a].

4.2.2 Residual dependence in asymptotic independence

In this section, we consider marginal transformations of the data vector $\mathbf{X} = (X_1, \ldots, X_d)$ to either standard Pareto margins $\mathbf{X}^P = (X_1^P, \ldots, X_d^P)$ or to unit exponential margins $\mathbf{X}^E = (X_1^E, \ldots, X_d^E) = \log(\mathbf{X}^P)$. We define the min-projection of $\mathbf{X}^E$ as $M_{\boldsymbol{\omega}}^{\min} =$

$\min_{i=1}^{d} X_i^E/\omega_i$ for a direction $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$ in S_d. Based on the multivariate tail representation (4.2), the function

$$f(x;\boldsymbol{\omega}) = \Pr\left(X_1^P > x^{\omega_1}, \ldots, X_d^P > x^{\omega_d}\right) = \Pr\left(M_{\boldsymbol{\omega}}^{\min} > \log x\right), \quad \text{for } x \geq 1,$$

is regularly varying at infinity with index $\lambda_{\boldsymbol{\omega}}$. Thus,

$$\begin{aligned} \frac{f(tx;\boldsymbol{\omega})}{f(t;\boldsymbol{\omega})} &= \frac{\Pr\left(M_{\boldsymbol{\omega}}^{\min} > \log x + \log t\right)}{\Pr\left(M_{\boldsymbol{\omega}}^{\min} > \log t\right)} \\ &= \Pr\left(M_{\boldsymbol{\omega}}^{\min} > \log x + u | M_{\boldsymbol{\omega}}^{\min} > u\right) \to x^{-\lambda_{\boldsymbol{\omega}}}, \quad \text{as } u \to +\infty, \end{aligned} \tag{4.6}$$

where $u = \log t \to +\infty$ as $t \to +\infty$. Equivalently, the excesses of the structure variable $M_{\boldsymbol{\omega}}^{\min}$ above a high threshold u are exponentially distributed with rate $\lambda_{\boldsymbol{\omega}}$ in the limit: Setting $\tilde{x} = \log x$ in (4.6) yields

$$\Pr\left(M_{\boldsymbol{\omega}}^{\min} > \tilde{x} + u | M_{\boldsymbol{\omega}}^{\min} > u\right) \to \exp\left\{-\tilde{x}\lambda_{\boldsymbol{\omega}}\right\}, \quad \text{as } u \to +\infty. \tag{4.7}$$

By using the tail structure (4.2) for characterizing asymptotic independence and appropriate marginal pretransformations, we can therefore model the positive excess $M_{\boldsymbol{\omega}}^{\min} - u$ of the min-projection above a fixed high threshold u through an $\text{Exp}(\lambda_{\boldsymbol{\omega}})$-distribution.

A more specific yet still very flexible class of asymptotically independent processes satisfying (4.2) are the *inverted max-stable processes* discussed in Wadsworth and Tawn [2012]. In unit Pareto margins, the limit relation (4.2) is exact for those models and can be written

$$\Pr\left(X_1^P > x^{\omega_1}, \ldots, X_d^P > x^{\omega_d}\right) = x^{-A_{\boldsymbol{\omega}}}, \quad x \geq 1, \tag{4.8}$$

with $A_{\boldsymbol{\omega}}$ the Pickands' dependence function of the associated max-stable process, which takes the role of the angular dependence function $\lambda_{\boldsymbol{\omega}}$. If $\mathbf{X}^P$ is an inverted max-stable process, then the original max-stable process is recovered through the transformation $1/\log(\mathbf{X}^P)$. The slowly varying function $\ell(.;\boldsymbol{\omega})$ in (4.2) is equal to 1 in this special case, such that

$$M_{\boldsymbol{\omega}}^{\min} \sim \text{Exp}\left\{A_{\boldsymbol{\omega}}\right\}. \tag{4.9}$$

Therefore, if we assume an inverted max-stable dependence in the joint tail of $\mathbf{X}^P$, we can model data values of $M_{\boldsymbol{\omega}}^{\min}$ above a fixed high threshold u through the exponential distribution in (4.9) while censoring values below u. This is different from the general setup with arbitrary unknown $\ell(.;\boldsymbol{\omega})$ where only the excesses above the threshold are used, not the information contained in the censoring indicator $\mathbb{1}_{M_{\boldsymbol{\omega}}^{\min} > u}$.

4.2.3 The case of Gaussian dependence

Suppose that X_1^F and X_2^F are unit Fréchet distributed. On one hand, Ledford and Tawn [1996] showed that if X_1^F and X_2^F have bivariate normal dependence with correlation $\rho < 1$, then

$$\Pr(X_1^F > r, X_2^F > r) \sim (1+\rho)^{3/2}(1-\rho)^{-1/2}(4\pi)^{-\rho/(1+\rho)} r^{-2/(1+\rho)} (\log r)^{-\rho/(1+\rho)}, \quad r \to \infty. \tag{4.10}$$

On the other hand, if the distribution function of $\mathbf{X}^F = (X_1^F, X_2^F)$ is a bivariate extreme value distribution $F(x_1, x_2) = \exp\{-V(x_1, x_2)\}$ with exponent function V and extremal coefficient $\theta = V(1,1)$, then

$$\Pr(X_1^F > r, X_2^F > r) \sim \left\{2 - V(1,1)\right\} r^{-1} + \left[\left\{V(1,1)\right\}^2/2 - 1\right] r^{-2}, \quad r \to \infty. \tag{4.11}$$

By assuming asymptotic dependence for data from a Gaussian copula with $\rho < 1$, we would fit an extreme value model at a sub-asymptotic level. Such model misspecification will result in biased extremal coefficient estimates smaller than 2, the value for asymptotic independence, owing to residual dependence in the data at finite levels. We can approximately quantify this bias by equating (4.10) and (4.11). The resulting *subsasymptotic extremal coefficient* $\theta(r)$ is equal to

$$\theta(r) = r - \left\{ r^2 - 4r + 2 + 2(1+\rho)^{3/2}(1-\rho)^{-1/2}(4\pi)^{-\rho/(1+\rho)} r^{-2/(1+\rho)} (\log r)^{-\rho/(1+\rho)} \right\}^{1/2}, \tag{4.12}$$

with r the marginal threshold. This relationship between r, $\theta(r)$, and ρ is displayed in Figure 4.1 where the threshold r is set to high quantiles of the unit Fréchet distribution, $r = -1/\log(1 - 10^{-q})$ with $q = 1, \ldots, 10$. For any $\rho < 1$, the coefficient $\theta(r)$ tends to

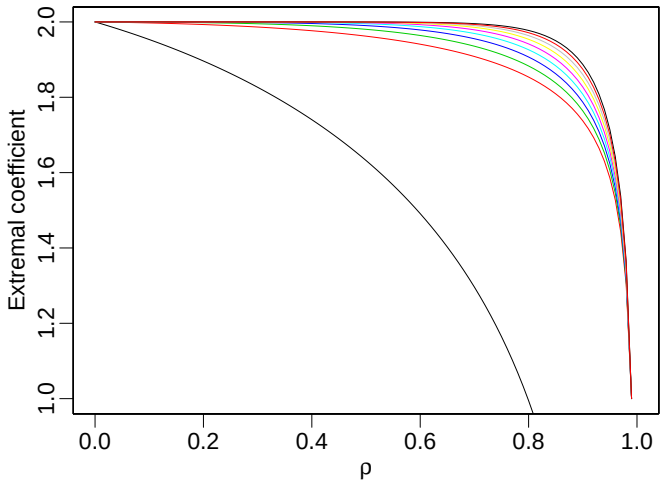

Figure 4.1: Sub-asymptotic extremal coefficient $\theta(r)$ of the Gaussian dependence as a function of the correlation ρ. Threshold levels are $r = -1/\log(1 - 10^{-q})$, with higher curves corresponding to higher levels

2 when r tends to infinity. This is obtained from an application of the binomial series formula to approximate the right-hand side of (4.12).

The extremal coefficient behaves as expected in the limit cases of perfect independence ($\rho = 0$) and perfect dependence ($\rho = 1$). By taking high thresholds such that residual dependence in the exceedance data vanishes, the extremal coefficient is close to 2 unless the correlation $\rho \approx 1$. Detection of asymptotic independence in the Gaussian dependence case with finite sample size was studied in Bücher et al. [2011] where the authors proposed new estimators of the Pickands' dependence function.

4.2.4 An illustration on data

In Sections 4.2.1 and 4.2.2, we developed asymptotically justified univariate exponential models for both classes of asymptotic dependence and independence using projections. Equation (4.2) characterizes the joint tail behaviour of random vectors with

$$\begin{aligned} \lambda_{\boldsymbol{\omega}} &\in \left] \max_{1\le i\le d} \omega_i, 1\right], \quad \text{or } \lambda_{\boldsymbol{\omega}} = \max_{1\le i\le d} \omega_i,\ \ell(x;\boldsymbol{\omega}) \to 0 \quad \text{(asymptotic independence)}, \\ \lambda_{\boldsymbol{\omega}} &= \max_{1\le i\le d} \omega_i \quad \text{and } \ell(x;\boldsymbol{\omega}) \not\to 0 \quad \text{(asymptotic dependence)}, \end{aligned}$$

where $\ell(x; \boldsymbol{\omega})$ has a well-defined positive limit in the asymptotic dependence case if the regular variation property (4.1) holds. If we look at data under the assumption of asymptotic dependence, then we calculate the projections $M_{\boldsymbol{\omega}}^{\min\downarrow}$ to estimate the values of the Pickands' dependence function. If we start with the assumption of asymptotic independence, then we calculate $M_{\boldsymbol{\omega}}^{\min}$ to estimate the values of the angular dependence function. A max-stable model characterized by the Pickands' function A provides an appealing link between the two dependence classes by the possibility of modelling asymptotic independence with the corresponding inverted max-stable model whose function $\ell(\cdot; \boldsymbol{\omega}) \equiv 1$ is known a priori and whose angular dependence function is A. Notice that the link between the two projections is as follows, using probability integral transformations:

$$M_{\boldsymbol{\omega}}^{\min\downarrow} = -\frac{\log\left[1 - \exp\left\{-A(\boldsymbol{\omega}) M_{\boldsymbol{\omega}}^{\min}\right\}\right]}{A(\boldsymbol{\omega})}.$$

Figure 4.2 illustrates these relationships on bivariate data simulated according to an extreme value distribution, where various marginal scales and projections at $\boldsymbol{\omega}_0 = (1/2, 1/2)$ are considered. Empirical thresholds at the 95% and 5% levels are used to censor the upper and lower tails, respectively.

4.3 Inference and regression modelling of dependence

The theory above shows that the projection techniques are appropriate for modelling asymptotic dependence and asymptotic independence in threshold excesses. In either case, we aim to develop a dependence model for the upper joint tail of the data distribution. With asymptotic independence, we transform the data margins to the exponential distribution, calculate the min-projection, and model its excesses above a high threshold through an exponential distribution, whose rate is given by the angular dependence function evaluated at the projection angle. More specifically, the latter corresponds to the Pickands' dependence function if we utilize an inverted max-stable model. With asymptotic dependence, we transform the data margins to the unit Fréchet distribution, calculate the max-projection, invert the latter, and model the resulting deficits below a small fixed threshold through an upper-censored exponential distribution, whose rate is given by the Pickands' dependence function evaluated at the projection angle.

By using min- and max-projections, inference is based on a univariate variable, which frees us from handling multivariate censoring schemes in likelihood-based approaches and the resulting computational burden. A difference between the two dependence regimes arises in the construction of the projection. On one hand, the max-projection in the case of asymptotic dependence retains only the highest value, which makes sense since the asymptotic theory implies that the limit model gives a good approximation for the components that are large. On the other hand, the min-projection in the case of asymptotic independence retains the smallest value, which contains crucial information on the faster tail decay rate along various directions $\boldsymbol{\omega}$, while the max-projection would converge to a unit exponential limit carrying no useful information in this case.

In either case, inference for the dependence structure is a two-step procedure based on a univariate structure variable. In the first step, the margins are transformed to the unit Fréchet scale and then inverted to the unit exponential scale (asymptotic dependence) or directly to the unit exponential scale (asymptotic independence), and we then calculate the min-projection. The second step consists of fitting the appropriate exponential model (4.4) or (4.7) to $M_{\boldsymbol{\omega}}^{\min\downarrow}$ and $M_{\boldsymbol{\omega}}^{\min}$, respectively. In the following, we detail the second

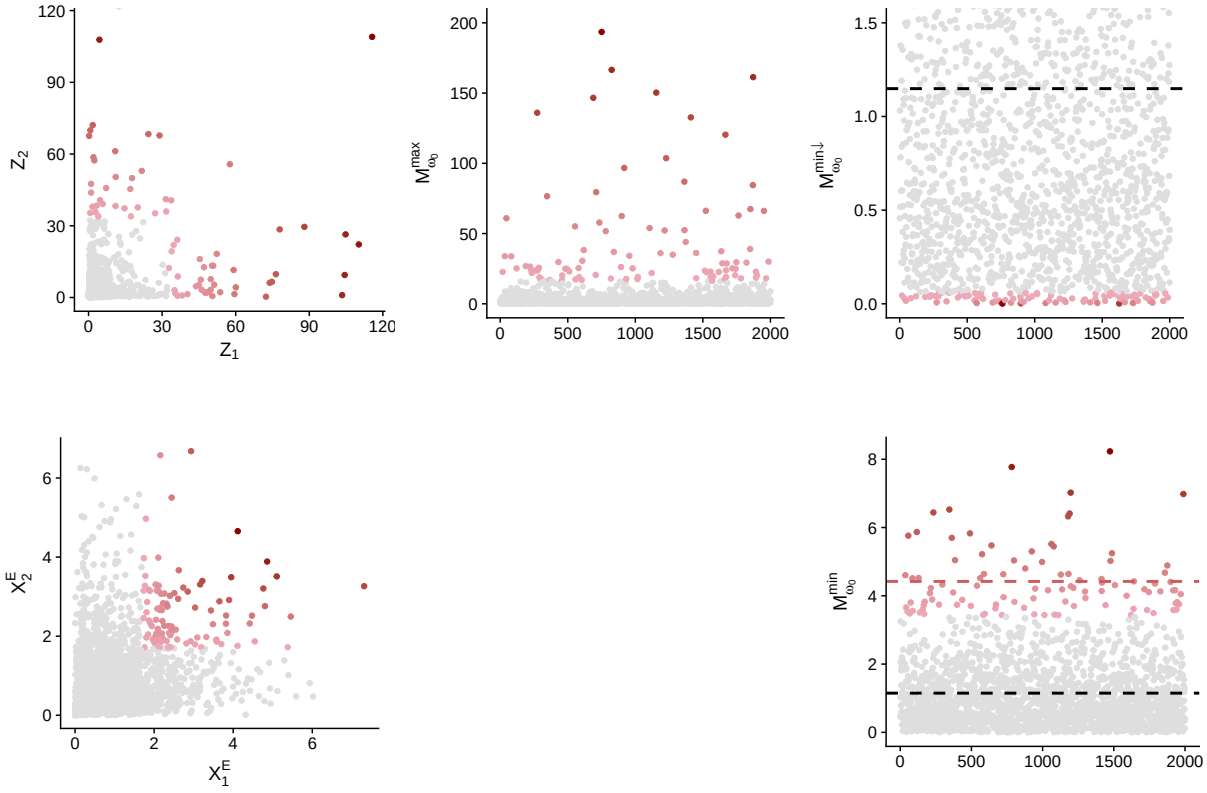

Figure 4.2: Realizations of a bivariate logistic extreme value distribution ($n = 2000$) with parameter 0.8 on the unit Fréchet scale (top left) and the standard exponential scale (bottom left), its corresponding max-projection (middle), and its corresponding min-projection (right). The *red* points correspond to the deficits of $M^{\min\downarrow}_{\boldsymbol{\omega}_0}$ below the 5% quantile (top) and the exceedances of $M^{\min}_{\boldsymbol{\omega}_0}$ above the 95% quantile (bottom). The dashed *black* lines indicate the mean of the random variables $M^{\min\downarrow}_{\boldsymbol{\omega}_0}$ and $M^{\min}_{\boldsymbol{\omega}_0}$, i.e., the inverse of $A_{\boldsymbol{\omega}_0}$ for the logistic model. The dashed *red* line corresponds to the mean of the excesses of $M^{\min}_{\boldsymbol{\omega}_0}$

step of the inference procedure with a view towards estimating the influence of a set of covariates $\mathbf{y} \in \mathbb{R}^q$.

4.3.1 Inference for the case of asymptotic dependence

Given n observations $\{\mathbf{x}_j^F\}_{j=1}^n$ of the random vector $\mathbf{X}^F = (X_1^F, \ldots, X_d^F)$ with unit Fréchet margins, we suppose that $\mathbf{X}^F$ is in the max-domain of attraction of an extreme value distribution with Pickands' dependence function A. We fix a direction $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in S_d$ and calculate the observed structure variables $m_{\boldsymbol{\omega},j}^{\min\downarrow} = \min_{i=1}^d 1/(\omega_i x_{j,i}^F)$, $j = 1, \ldots, n$. To put focus on the dependence of extremes in the second step of the inference procedure, we censor observations $m_{\boldsymbol{\omega},j}^{\min\downarrow}$ that are above a low threshold $u > 0$, for instance, chosen as the empirical 5% quantile of $m_{\boldsymbol{\omega},j}^{\min\downarrow}$. The likelihood function is $L\left(A_{\boldsymbol{\omega}}; m_{\boldsymbol{\omega},1}^{\min\downarrow}, \ldots, m_{\boldsymbol{\omega},n}^{\min\downarrow}\right) = \prod_{j=1}^n L_j\left(A_{\boldsymbol{\omega}}\right)$ with contributions

$$L_j\left(A_{\boldsymbol{\omega}}\right) = \begin{cases} A_{\boldsymbol{\omega}} \exp\left(-A_{\boldsymbol{\omega}} m_{\boldsymbol{\omega},j}^{\min\downarrow}\right) & \text{if } m_{\boldsymbol{\omega},j}^{\min\downarrow} < u, \\ \exp\left(-A_{\boldsymbol{\omega}} u\right) & \text{if } m_{\boldsymbol{\omega},j}^{\min\downarrow} \geq u. \end{cases}$$

When a set of covariates $\mathbf{y} \in \mathbb{R}^q$ is available, we propose using a generalized additive model (GAM) structure [Wood, 2017] to model the dependence in the extremes, similar to the approach of Mhalla et al. [2017a], who estimated the Pickands' dependence function based on the block maxima approach. The dependence of $\mathbf{x}^F$ on $\mathbf{y}$ is assumed to be at the extremal dependence level, i.e., the covariates solely influence the Pickands' dependence function $A_{\boldsymbol{\omega}} \equiv A_{\boldsymbol{\omega}}(\mathbf{y})$. This assumption implies no loss of generality as the marginal inference is performed in a separate step. A very general model for $A_{\boldsymbol{\omega}}(\mathbf{y})$ arises from supposing the semi-parametric form

$$A_{\boldsymbol{\omega}}(\mathbf{y}; \boldsymbol{\Lambda}) \quad = \quad h^{-1}\left\{\mathbf{u}^T\boldsymbol{\beta} + \sum_{k=1}^K h_k(t_k)\right\}, \tag{4.13}$$

where h is a link function and $\mathbf{u} \in \mathbb{R}^s$ and $(t_1, \ldots, t_K)$ are subvectors of $\mathbf{y}$, or products of covariates if interactions between some covariates are considered. The column vector $\boldsymbol{\beta} \in \mathbb{R}^s$ gathers linear coefficients whereas $h_k : \mathbb{H}_k \to \mathbb{R}$ are smooth functions supported on closed intervals $\mathbb{H}_k \subset \mathbb{R}$ and admitting a finite quadratic penalty representation [Green and Silverman, 1994]. The column vector $\boldsymbol{\Lambda}$ gathers all parameters to be estimated in the model, i.e., the vector $\boldsymbol{\beta}$ and the linear basis coefficients of each of the smooth functions h_k. Based on a sample $\{\mathbf{x}_j^F, \mathbf{y}_j\}_{j=1}^n$, we estimate the GAM (4.13) by maximizing the penalized log-likelihood

$$\ell(\boldsymbol{\Lambda}, \boldsymbol{\gamma}) = \ell(\boldsymbol{\Lambda}) - \frac{1}{2}\sum_{k=1}^K \gamma_k \int_{\mathbb{H}_k} h_k''(t_k)^2 \, \mathrm{d}t_k, \tag{4.14}$$

with $\ell(\boldsymbol{\Lambda}) = \sum_{j=1}^n \ell_j(m_{\boldsymbol{\omega},j}^{\min\downarrow}, \mathbf{y}_j, \boldsymbol{\Lambda})$,

$$\ell_j(m_{\boldsymbol{\omega},j}^{\min\downarrow}, \mathbf{y}_j, \boldsymbol{\Lambda}) = \begin{cases} \log A_{\boldsymbol{\omega}}(\mathbf{y}_j) - A_{\boldsymbol{\omega}}(\mathbf{y}_j) m_{\boldsymbol{\omega},j}^{\min\downarrow} & \text{if } m_{\boldsymbol{\omega},j}^{\min\downarrow} < u, \\ -A_{\boldsymbol{\omega}}(\mathbf{y}_j) u & \text{if } m_{\boldsymbol{\omega},j}^{\min\downarrow} \geq u. \end{cases}$$

The integrals in (4.14) are componentwise roughness penalties with smoothing parameters $\gamma = (\gamma_1, \ldots, \gamma_K)$ that balance between the smoothness of the model and its goodness of fit. Higher values of γ_k yield smoother fitted curves. The related effective degrees of

freedom of each smooth function h_k are defined as $\text{trace}\,(I + \gamma_k S_k)^{-1}$, where S_k is the positive definite penalty matrix associated to the basis representation of h_k [Wood, 2017, Chapter 5]. The maximization of the penalized log-likelihood (4.14) is performed based on an outer-iteration procedure. At each iteration, $\boldsymbol{\Lambda}$ and $\boldsymbol{\gamma}$ are estimated separately by penalized iteratively re-weighted least squares (PIRLS) and a prediction error method (Generalized Cross Validation), respectively; see Wood [2017] for a detailed description of the available methods for GAM fitting.
The penalized maximum log-likelihood estimator $\widehat{\boldsymbol{\Lambda}}_n$ then defines the estimate $\hat{A}_{\boldsymbol{\omega}}(\mathbf{y}) = h^{-1}\left\{\mathbf{u}^T\hat{\boldsymbol{\beta}} + \sum_{k=1}^{K}\hat{h}_k(t_k)\right\}$ of the Pickands' dependence function evaluated at $\boldsymbol{\omega}$. A vast amount of literature on GAM-related theory is available and includes Wood [2004, 2006a], Marra and Wood [2011], among others.

4.3.2 Inference for the case of asymptotic independence

Given n observations $\{\mathbf{x}_j^P\}_{j=1}^n$ of the random vector $\mathbf{X}^P = (X_1^P, \ldots, X_d^P)$ with standard Pareto margins, we suppose that $\mathbf{X}^P$ has an asymptotic independent tail structure as in (4.2). We fix a direction $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in S_d$ and calculate the observed structure variables

$$m_{\boldsymbol{\omega},j}^{\min} = \min_{i=1}^{d} \log(x_{j,i}^P)/\omega_i, \quad j = 1, \ldots, n.$$

We fix a high threshold u, for instance, chosen as the empirical 95% quantile of $m_{\boldsymbol{\omega},j}^{\min}$, and extract the sample of positive excesses $\tilde{m}_{\boldsymbol{\omega},j_e}^{\min} = m_{\boldsymbol{\omega},j_e}^{\min} - u > 0$, $e = 1, \ldots, E_u$ with a positive number of excesses $E_u > 0$. Then, we maximize the likelihood composed of contributions

$$L_{j_e}\{\lambda_{\boldsymbol{\omega}}\} = \lambda_{\boldsymbol{\omega}} \exp\left\{-\lambda_{\boldsymbol{\omega}}\left(\tilde{m}_{\boldsymbol{\omega},j_e}^{\min}\right)\right\}, \quad e = 1, \ldots E_u. \tag{4.15}$$

Given covariate vectors $\mathbf{y}_j$, we proceed as for the asymptotic dependence case by proposing a GAM (4.13) for $\lambda_{\boldsymbol{\omega}} \equiv \lambda_{\boldsymbol{\omega}}(\mathbf{y})$, i.e.,

$$\lambda_{\boldsymbol{\omega}}(\mathbf{y}; \boldsymbol{\Lambda}) = h^{-1}\left\{\mathbf{u}^T\boldsymbol{\beta} + \sum_{k=1}^{K} h_k(t_k)\right\},$$

which is fitted by maximizing a penalized version of the likelihood (4.15) similarly to (4.14) and results in the estimate $\hat{\lambda}_{\boldsymbol{\omega}}(\mathbf{y})$ of the tail dependence function evaluated at $\boldsymbol{\omega}$.

4.4 Simulation study

We study the properties of the estimators $\hat{A}_{\boldsymbol{\omega}}(\mathbf{y})$ and $\hat{\lambda}_{\boldsymbol{\omega}}(\mathbf{y})$ when they are related to a covariate vector $\mathbf{y} \in \mathbb{R}^q$ using the link function $h(x) = \log\{(x - 1/2)/(1 - x)\}$, a modification of the logit link resulting in values within $(0.5, 1)$. We focus on the estimation of these two dependence functions in the bivariate case at $\boldsymbol{\omega}_0 = (1/2, 1/2)$, which yields estimates of the covariate-dependent coefficients of extremal dependence $\hat{\theta}(\mathbf{y}) = 2\hat{A}_{\boldsymbol{\omega}_0}(\mathbf{y})$ and tail dependence $\hat{\eta}(\mathbf{y}) = 1/\{2\lambda_{\boldsymbol{\omega}_0}(\mathbf{y})\}$. Realistic sample sizes are chosen, similar to those in the subsequent application. Note that the uncertainty assessment in the following simulation studies does not take into account the uncertainty resulting from the separate marginal fitting.

4.4.1 Case of asymptotic dependence

Suppose that $\mathbf{X}^F$ is an asymptotically dependent random vector with unit Fréchet margins. We first consider the max-domain of attraction (MDA) setting where our model represents an asymptotic approximation of the exact tail behaviour in the data.

4.4.1.1 Data distribution in the maximum domain of attraction

We focus on the bivariate case with $\mathbf{X}^F = (X_1^F, X_2^F)$ a random vector with unit Fréchet margins and an Archimedean copula $C^\star$ [Nelsen, 2006, Chapter 4] with generator $\varphi(t) = (1/t - 1)^{1/\alpha}$ for some $\alpha \in (0, 1)$, i.e., the distribution function of $\mathbf{X}^F$ is

$$F(x_1, x_2) = \varphi\left[\varphi^{-1}\left\{\exp\left(-\frac{1}{x_1}\right)\right\} + \varphi^{-1}\left\{\exp\left(-\frac{1}{x_2}\right)\right\}\right], \quad x_1, x_2 > 0. \tag{4.16}$$

This distribution is in the max-domain of attraction of the logistic bivariate extreme value distribution [Tawn, 1990] with parameter α [Fougères, 2004], and can be simulated using the algorithm from Nelsen [2006, example 4.15, p. 144]. We estimate the covariate-dependent extremal coefficient using deficits of $M_{\boldsymbol{\omega}_0}^{\min\downarrow}$ by setting $\boldsymbol{\omega}_0 = (1/2, 1/2)$ and fixing different threshold levels.

Equation (4.4) holds when at least one variable X_i^F exceeds some high threshold, and the limiting dependence structure in the tails of $\mathbf{X}^F$ is equal to that of a bivariate logistic extreme value distribution [Coles and Tawn, 1991]. We empirically study the bias of the covariate-dependent extremal coefficient estimator resulting from the application of the limiting extreme value model.

We simulate from a covariate-dependent model (4.16) with the following generator

$$\varphi(t; y) = \left(\frac{1}{t} - 1\right)^{1/\alpha(y)},$$

where $y \in [0, 1]$ is observed at 50 equally spaced values and

$$\alpha(y) = \log\left[1 + \frac{\exp\{\sin(2\pi y) + y^2\}}{1 + \exp\{\sin(2\pi y) + y^2\}}\right] / \log(2) \in (0, 1).$$

Thus, the covariate-dependent extremal coefficient is

$$\theta(y) = 2A_{\boldsymbol{\omega}_0}(y) = 1 + \frac{\exp\{\sin(2\pi y) + y^2\}}{1 + \exp\{\sin(2\pi y) + y^2\}}. \tag{4.17}$$

The sample size is chosen between 50000 for a threshold at the 10% level and 500000 at the 1% level, such that an average of 100 threshold deficits of $M_{\boldsymbol{\omega}_0}^{\min\downarrow}$ arises for a fixed value of y. Additionally, we empirically quantify the performance of our estimator in an over parametrized setting where one of the covariates has no influence on the response by considering a dummy categorical covariate I with two levels. In the true model (4.17), the Pickands' dependence function does not depend on I. We simulate the observed values of I at random and include level-dependent predictors in the GAM structure for $A_{\boldsymbol{\omega}_0}(y)$; thus, the estimated model is

$$h\{\hat{A}_{\boldsymbol{\omega}_0}(y, I)\} = \hat{a}_1 + \hat{s}_1(y)\mathbb{1}_{\{I=\text{“Level 1”}\}} + \{\hat{a}_2 + \hat{s}_2(y)\}\,\mathbb{1}_{\{I=\text{“Level 2”}\}},$$

where $\hat{a}_i$, $i = 1, 2$ are the estimated intercepts and $\hat{s}_1$, and $\hat{s}_2$ are the estimated smooth curves describing the dependence of the Pickands' function at $\boldsymbol{\omega}_0$ on y and at each level of I. The Monte Carlo procedure is based on 500 repetitions, and 95% percentile confidence intervals are constructed. Figure 4.3 displays the pointwise mean estimates of $\theta(y, I)$ in both levels of I with a threshold set at the 5% level. As expected, the inclusion of I

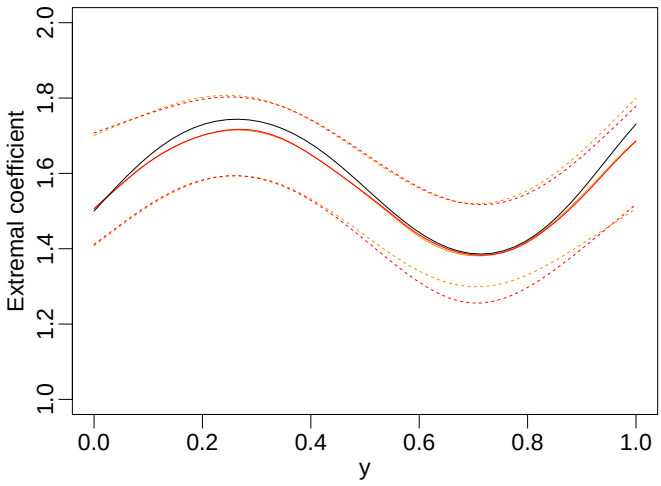

Figure 4.3: Estimation in the max-domain of attraction setting. Estimates (solid) and 95% bootstrap confidence intervals (dashed) of the extremal coefficient $\theta(y)$ in the first and second levels of I in *orange* and *red* lines, respectively. The true values as given by (4.17) are in solid *black* lines

in the model does not prevent it from recovering the true dependence structure of the extremal coefficient on y in both levels. Moreover, when considering the point estimates (i.e., the single-run experiments), the results indicate that the variable I is not statistically significant in the fitted GAM. Next, we remove this artificial covariate from the model and compare the root mean squared error (RMSE) of the resulting extremal coefficient fits obtained for different threshold levels. The RMSE over the 500 samples is defined as

$$\text{RMSE}(y) = \left[\sum_{r=1}^{500} \{\hat{\theta}^r(y) - \theta(y)\}^2 / 500\right]^{1/2},$$

where $\hat{\theta}^r(y)$ is the covariate-dependent extremal coefficient estimate obtained from the rth sample. Figure 4.4 shows the RMSE of $\hat{\theta}(y)$ for different threshold levels. The RMSE tends to decrease when we take lower threshold levels. Similar shapes of the RMSE function are observed for threshold levels below 10%, with a noticeable decrease in the RMSE for values of y corresponding to weak extremal dependence, i.e., for values of y between 0.2 and 0.4.

4.4.1.2 Max-stable data distribution

We now simulate bivariate samples with max-stable dependence where our asymptotic dependence model class contains the exact model. The observations come from the logistic extreme value copula with unit Fréchet margins and dependence parameter $\alpha(t) = t - 0.05$,

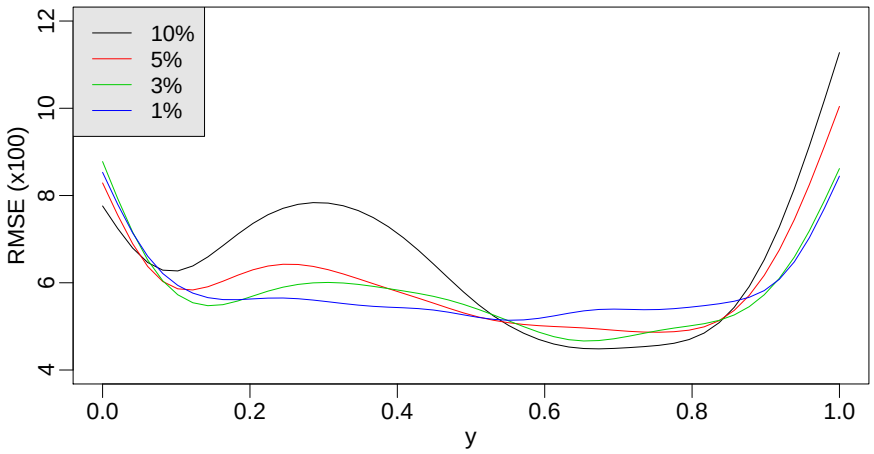

Figure 4.4: Estimation in the max-domain of attraction setting, with the root mean squared error of the estimates of $\theta(y)$ for $y \in [0, 1]$ based on 500 samples and obtained at different threshold levels

$t \in [0.1, 1]$. The covariate-dependent extremal coefficient $\theta(t)$ is $2^{\alpha(t)}$. Figure 4.5 displays the RMSE of the extremal coefficient estimates with respect to t for the 10%, 5%, 3%, and 1% threshold levels. The RMSE is mostly unaffected by the threshold level as our modelling assumption (4.4) holds exactly, as opposed to the sub-asymptotic setting in Section 4.4.1.1. Therefore, we observe a lower RMSE (see Figure 4.4), as there is no estimation bias resulting from penultimate modelling.

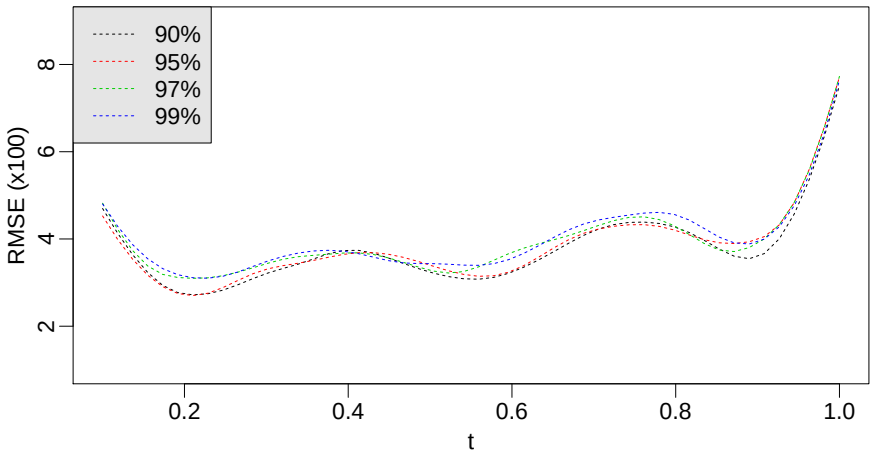

Figure 4.5: Root mean squared error of the estimates of $\theta(t)$ for $t \in [0.1, 1]$ with respect to different threshold levels, for 500 simulations

4.4.2 Case of asymptotic independence

We consider two models for an asymptotically independent random vector $\mathbf{X}^E = (X_1^E, X_2^E)$ with standard exponential margins. The dependence in the tails of $\mathbf{X}^E$ depends on the covariates I and t, where I is a categorical covariate with two levels 1 and 2. Bivariate Gaussian dependence with correlation $\rho(t)$ is observed in the first level of I, and an inverted logistic extreme value dependence with dependence parameter $\alpha(t)$ is observed in the second level of I. The dependence on covariates is as follows:

$$\begin{aligned} \rho(t) &= t, && (4.18)\\ \alpha(t) &= t - 0.05, && (4.19)\\ t &\in [0.1, 1]. \end{aligned}$$

Then,

$$\eta(t, I) = 1/\{2\lambda_{\boldsymbol{\omega}_0}(t, I)\} = \begin{cases} \{1 + \rho(t)\}/2 & \text{if } I = \text{“Level 1”},\\ 1/\{2A_{\boldsymbol{\omega}_0}(t)\} = 2^{-\alpha(t)} & \text{if } I = \text{“Level 2”}. \end{cases}$$

The strength of the tail dependence varies from low (small values of $\rho(t)$, large values of $\alpha(t)$) to strong (large values of $\rho(t)$, small values of $\alpha(t)$) where $\eta(t, I) < 1$ unless $\rho(t) \equiv 1$ or $\alpha(t) \to 0$ (perfect dependence).

For assessing the performance of our generalized additive modelling framework for the coefficient of tail dependence when different strengths of tail dependence are observed, we construct the following model for the angular dependence function:

$$h\{\lambda_{\boldsymbol{\omega}_0}(t, I)\} = \lambda_0 + s_1(t)\mathbb{1}_{\{I=\text{“Level 1”}\}} + s_2(t)\mathbb{1}_{\{I=\text{“Level 2”}\}}, \tag{4.20}$$

with intercept λ_0 and smooth functions s_1 and s_2 of t. Our simulation study is based on an average of 100 threshold exceedances of $M_{\boldsymbol{\omega}_0}^{\min} = \min(2X_1^E, 2X_2^E)$ for a fixed value of the covariate vector (t, I); the sample size varies between 10^5 for a threshold at the 90% level and 10^6 at the 99% level. As before, 500 repetitions are carried out. We consider different threshold levels to quantify the bias resulting from the estimation of $\eta(t, I)$ based on (4.7) at a finite threshold u or equivalently (4.2) at a finite level x.

The RMSE defined as

$$\text{RMSE}(t, I) = \left[\sum_{r=1}^{500}\{\hat{\eta}^r(t, I) - \eta(t, I)\}^2/500\right]^{1/2},$$

with $\hat{\eta}^r(t, I)$ the covariate-dependent tail dependence coefficient estimate obtained from the r-th bootstrap sample, is displayed in Figure 4.6. In the Gaussian case, the RMSE decreases when the threshold level increases but increases when stronger dependence is considered, i.e., when $\rho(t)$ (or t, equivalently) increases. This is due to the slow convergence of tail measures at sub-asymptotic levels for the Gaussian dependence; see Coles et al. [1999]. In the inverted extreme value case, the estimator's performance is largely unaffected by the threshold level as Equation (4.8) entails that the approximation (4.2) is exact at finite levels.

4.5 Application to nitrogen dioxide data

We illustrate the modelling of covariate-dependent tail dependence on a nitrogen dioxide (NO_2) [$\mu g/m^3$] data set, extracted from the European air quality database for pollutants

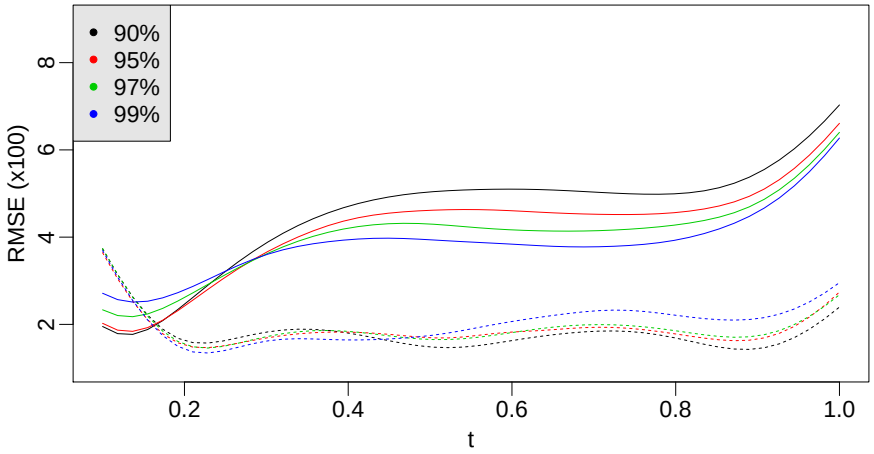

Figure 4.6: Root mean squared error of the bootstrap estimates of $\eta(t, I)$ for $I = 1$, i.e., the Gaussian case (solid lines), and for $I = 2$, i.e., the inverted logistic extreme value case (dashed lines). Different threshold levels are considered

AirBase[1]. It comprises 569 measurement stations in France with hourly records of NO_2 observed over 14 years between 1999 and 2012; see the map of stations in Figure 4.7. With NO_2 produced mostly by the burning of fossil fuel and motor vehicle exhaust, we propose to distinguish traffic stations where the NO_2 level is predominantly determined by nearby traffic (129 stations) from background stations, often located in built-up areas, whose level of NO_2 is influenced by a combination of many sources (440 stations). Figure 4.8 shows NO_2 measurements in 1999 for two background stations located 3 km apart ("Metz-Centre" and "Metz-Borny") and for two traffic stations located 8 km apart ("Auto A1-Saint-Denis" and "Rue Bonaparte"). The measurements at the two background stations seem to follow a similar pattern, with large values recorded around the same time periods and low values observed during the summer season when most of the NO_2 is transformed into ozone through sunlight. The comparison of the NO_2 measurements for the two traffic stations is less straightforward. The magnitude of observations is much higher than for the background sites. Large observations seem to occur mostly locally although the stations being very close.

Such insights justify the distinction between the two station types when investigating the co-occurrence of large concentrations measured at pairs of stations. Moreover, it is natural to ask whether the frequency of co-occurrences of high pollution levels has changed over time, for instance, as a result of regulatory measures. For the present analysis, we reduce the dimension of data by considering monthly maxima at each of the 569 stations, which avoids modelling of hourly patterns in NO_2 concentration levels and of intraday dependence between the measurements [Shi et al., 2014].

At each station, the marginal distribution of monthly maxima is modelled using a generalized extreme value (GEV) distribution. Its parameters for location μ and scale σ

[1]https://www.eea.europa.eu/data-and-maps/data/aqereporting-2

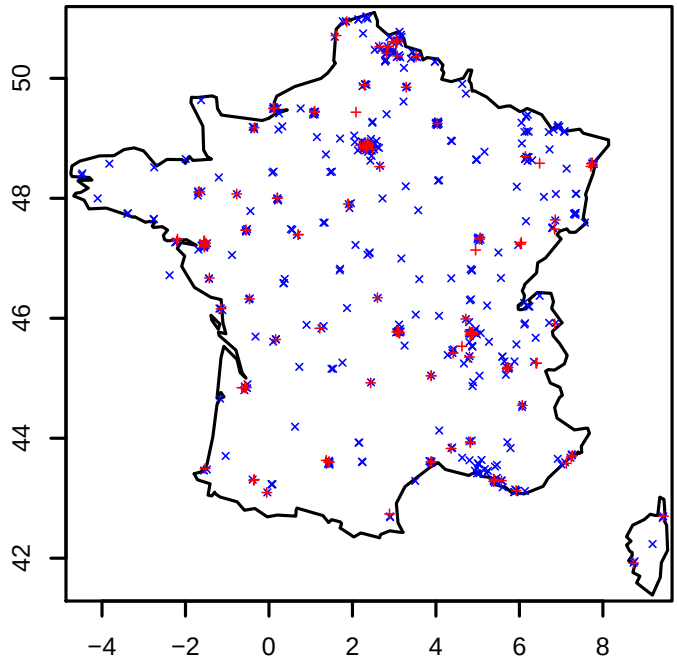

Figure 4.7: Map of nitrogen dioxide measurement stations in France that are "background" (*blue* × symbols) or "traffic" (*red* + symbols) stations

are allowed to vary smoothly with the year and month of the observed maxima as follows:

$$\begin{aligned} \mu(t,m) &= \mu_0 + f_1(t) + f_2(m), \\ \sigma(t,m) &= \sigma_0 + g_1(t) + g_2(m), \end{aligned}$$

where t and m denote the year and month of the observed maxima, respectively, f_1 and g_1 are smooth functions accounting for the trend, and f_2 and g_2 are cyclic smooth functions accounting for the seasonal component in the data [Wood, 2017, Section 5.3]. Likelihood ratio tests are performed to assess whether we need smoothly varying terms or whether parametric and sinusoidal terms already provide a good fit. The final fitted model is then used to transform the data at each station to the unit Fréchet distribution using the probability integral transform.

We focus on modelling the dependence between high NO_2 measurements recorded at pairs of stations. The dependence should naturally tend to decay with the distance between the stations in each pair. Therefore, we include distance (in kilometres) as a covariate in our model along with the type of area (traffic/background) and time (year), as discussed above. We consider the great circle distance between the stations, i.e., the shortest distance over Earth's surface, and fix the distance resolution at 10 km, which represents the minimal distance by which stations must be separated to be distinguishable. This setting leads to 93 distinct values for the distance covariate, 340, 740 pairs of observations for the background stations and 61, 705 pairs for the traffic stations. Both

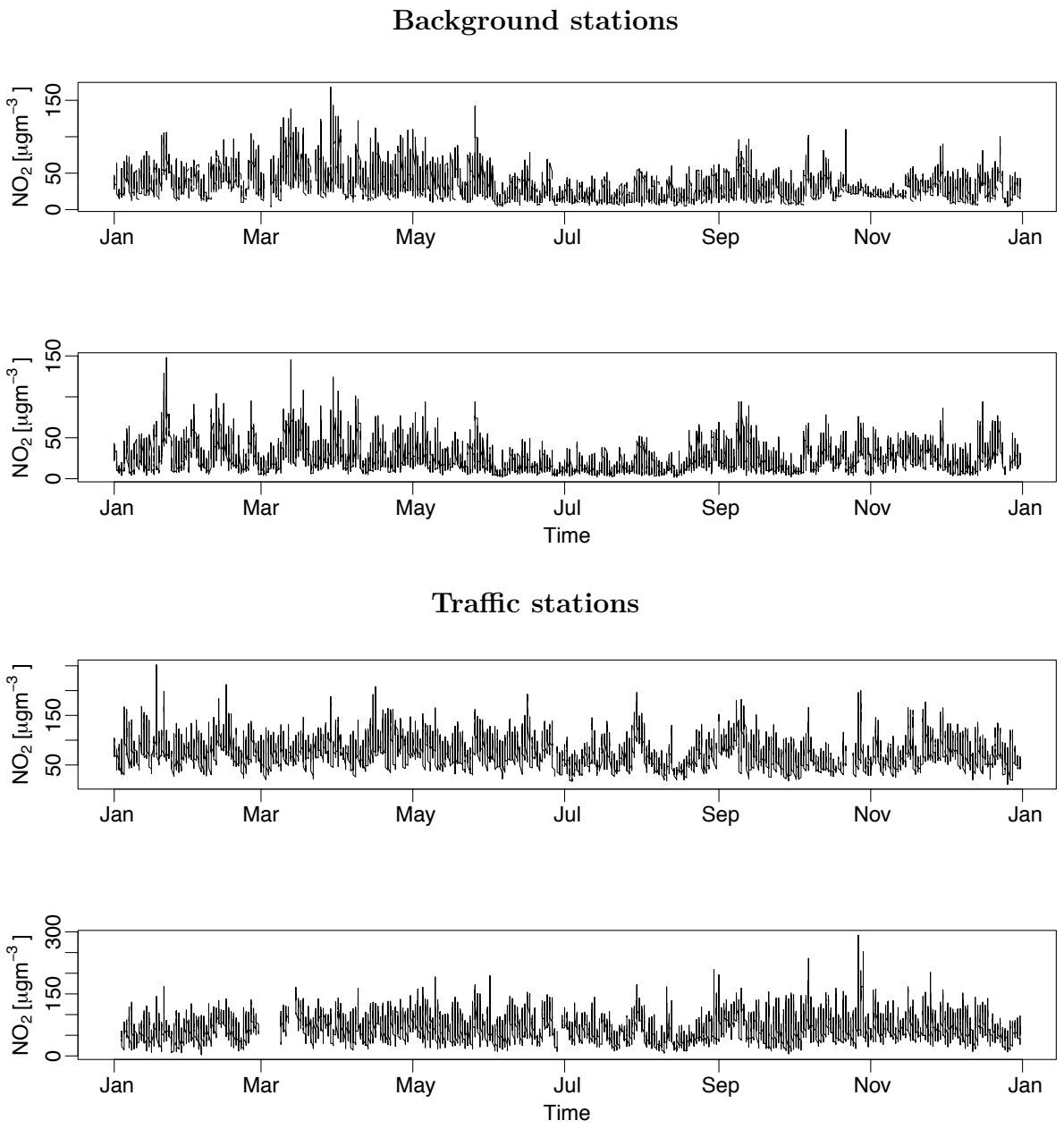

Figure 4.8: Nitrogen dioxide measurements recorded in 1999 at two background stations in Metz (top) and two traffic stations in Paris (bottom)

tail measures, based on the assumption of either asymptotic dependence or asymptotic independence, are modelled. We construct the following "full" models for the extremal coefficient and the tail dependence coefficient:

$$\begin{aligned} h\{\theta(t,d,\text{type})/2\} &= \theta_0 + \mathbb{1}_{\{\text{type="Background"}\}}\{f_1(t) + f_2(d) + f(t,d)\} + \\ &\quad \mathbb{1}_{\{\text{type="Traffic"}\}}\{\theta_1 + g_1(t) + g_2(d) + g(t,d)\}, \qquad (4.21) \\ h\left[\{2\eta(t,d,\text{type})\}^{-1}\right] &= \eta_0 + \mathbb{1}_{\{\text{type="Background"}\}}\{\tilde{f}_1(t) + \tilde{f}_2(d) + \tilde{f}(t,d)\} + \\ &\quad \mathbb{1}_{\{\text{type="Traffic"}\}}\{\eta_1 + \tilde{g}_1(t) + \tilde{g}_2(d) + \tilde{g}(t,d)\}, \qquad (4.22) \end{aligned}$$

where t represents time (in years), d the distance between the stations in each pair (in km), and $h(x) = \log(x - 1/2) - \log(1 - x)$. The interaction between time and distance is

represented using a tensor product basis [Wood, 2017, Section 5.6]. We fit models (4.21) and (4.22) based on the deficits of the 5% quantile of $M^{\min\downarrow}_{\boldsymbol{\omega}_0}$ and the exceedances of the 95% quantile of $M^{\min}_{\boldsymbol{\omega}_0}$ with $\boldsymbol{\omega}_0 = (0.5, 0.5)$, respectively.

To start, we conduct a simpler, purely spatial analysis and consider the models without time effects and such that data are pooled together over the whole period for each pair of stations during the estimation. Estimated summaries $\hat{\theta}(d, \text{type})$ and $\hat{\eta}(d, \text{type})$ are shown in Figure 4.9; they clearly hint at asymptotic independence with estimates and bootstrap-based pointwise confidence intervals of θ very close to 2, while those of η are clearly bounded away from 1. Overall, the joint tail decay rates in asymptotic independence appear to be quite fast with pointwise confidence envelopes of η contained between 0.5 and 0.7 approximately. A partial explanation for this relatively weak dependence is that our univariate models have already appropriately removed seasonal trends in the data, such that dependence in the resulting residuals cannot arise from intermediate-range clustering in space and time. As a matter of fact, as seasonal patterns are typically spatial, filtering out these patterns would result in a spatial de-clustering in the region where the seasonal features are observed. We detect a stronger weakening of dependence with increasing distance in the traffic stations, which makes sense because the peaks in the NO_2 concentrations are strongly influenced by nearby traffic.

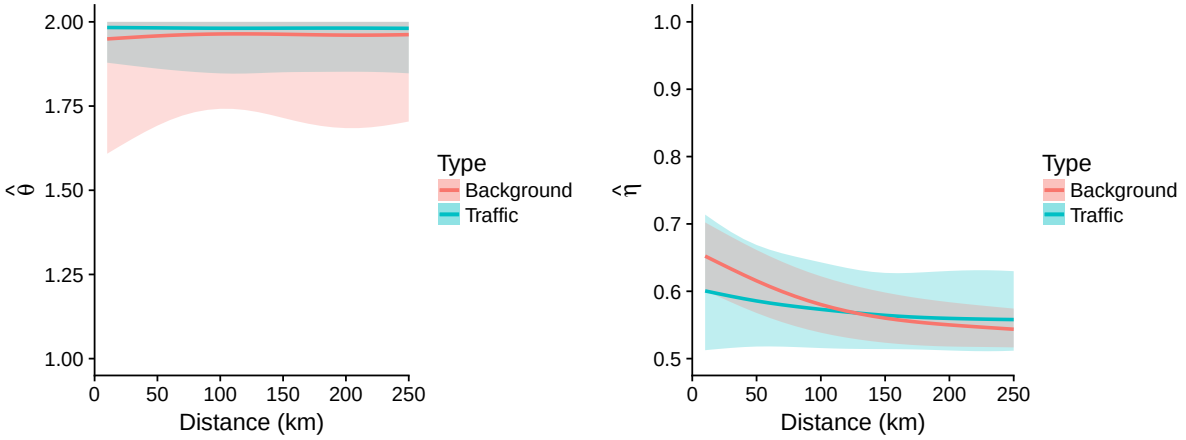

Figure 4.9: Estimation of distance-dependent dependence summaries for the nitrogen dioxide data. Left: $\hat{\theta}(d, \text{type})$, assuming asymptotic dependence. Right: $\hat{\eta}(d, \text{type})$, assuming asymptotic independence

Next, we consider regression fits (4.21) and (4.22) with spatial and temporal components; i.e., we have a more complex, higher-dimensional model for the predictors, and we expect the estimation uncertainty to be higher. Table 4.1 summarizes both fitted models. All the considered covariate effects, except for the distance between traffic stations when the tail dependence coefficient is modelled, are statistically significant.
This confirms our intuition that high NO_2 concentrations are relatively localized for the traffic stations. A block bootstrap procedure treating NO_2 measurements from each month and each year as independent for each type of area is used to assess the uncertainty. Figure 4.10 shows cross sections of the estimates of the extremal coefficient for the traffic and background stations; we fix one of the two continuous covariates to show the smooth effect of the other. For both area types, the estimates are very close to the upper bound of the extremal coefficient for almost all values of t and d. This result implies weak

Table 4.1: Estimates (se) of the intercepts and the degrees of freedom (edf) of the smooth functions in models (4.21) and (4.22)

		$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{f}_1$	$\hat{g}_1$	$\hat{f}_2$	$\hat{g}_2$	$\hat{f}$	$\hat{g}$
$\hat{\theta}(t, d, \text{type})$	estimate/edf	12.76(0.39)	−2.41(0.69)	3.98	3.55	3.82	0.84	15.55	2.92
	p-value	$< 10^{-16}$	4.74×10^{-4}	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	1.21×10^{-2}	$< 10^{-16}$	$< 10^{-16}$
		$\hat{\eta}_0$	$\hat{\eta}_1$	$\hat{\tilde{f}}_1$	$\hat{\tilde{g}}_1$	$\hat{\tilde{f}}_2$	$\hat{\tilde{g}}_2$	$\hat{\tilde{f}}$	$\hat{\tilde{g}}$
$\hat{\eta}(t, d, \text{type})$	estimate/edf	1.86(0.03)	−0.46(0.07)	3.88	3.11	3.82	0	9.62	1.66
	p-value	$< 10^{-16}$	4.49×10^{-10}	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	0.61	$< 10^{-16}$	2.98×10^{-4}

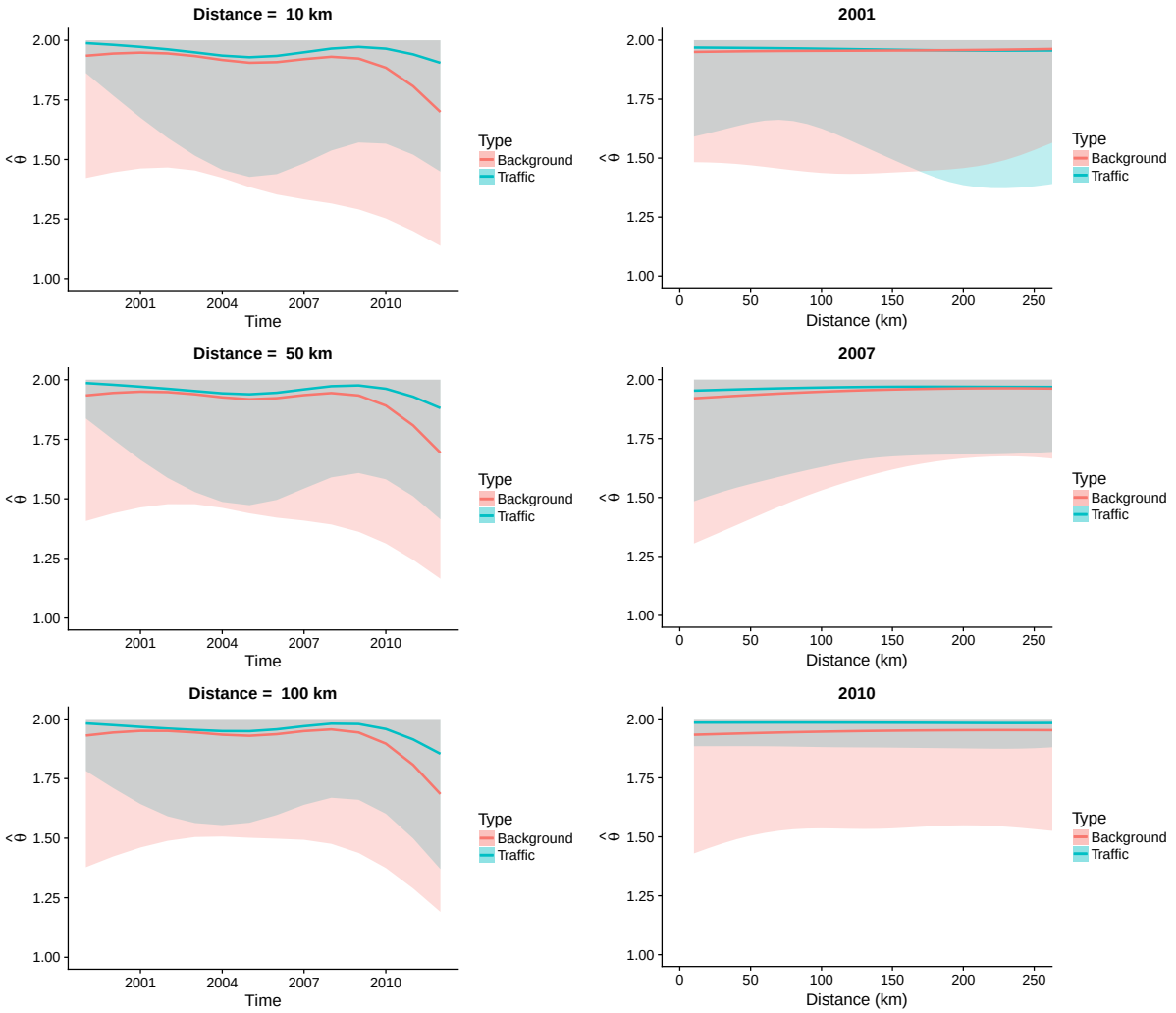

Figure 4.10: Cross sections of the covariate-dependent $\hat{\theta}(t, d)$ with the corresponding 95% bootstrap uncertainty intervals. The bootstrap is based on 300 bootstrap samples

extremal dependence between the NO_2 measurements at the pairs of stations, and we can suppose asymptotic independence in the data. We now focus on the estimates of the tail dependence coefficient; the sub-asymptotic modelling of the tails should be more informative in this case. Figure 4.11 displays cross sections of the estimate of the tail dependence coefficient $\hat{\eta}(t, d)$ for the different types of area. For the considered years, the residual tail dependence is relatively weak and largely unaffected by the distance between traffic stations, owing to the very localized features of traffic pollution inducing high

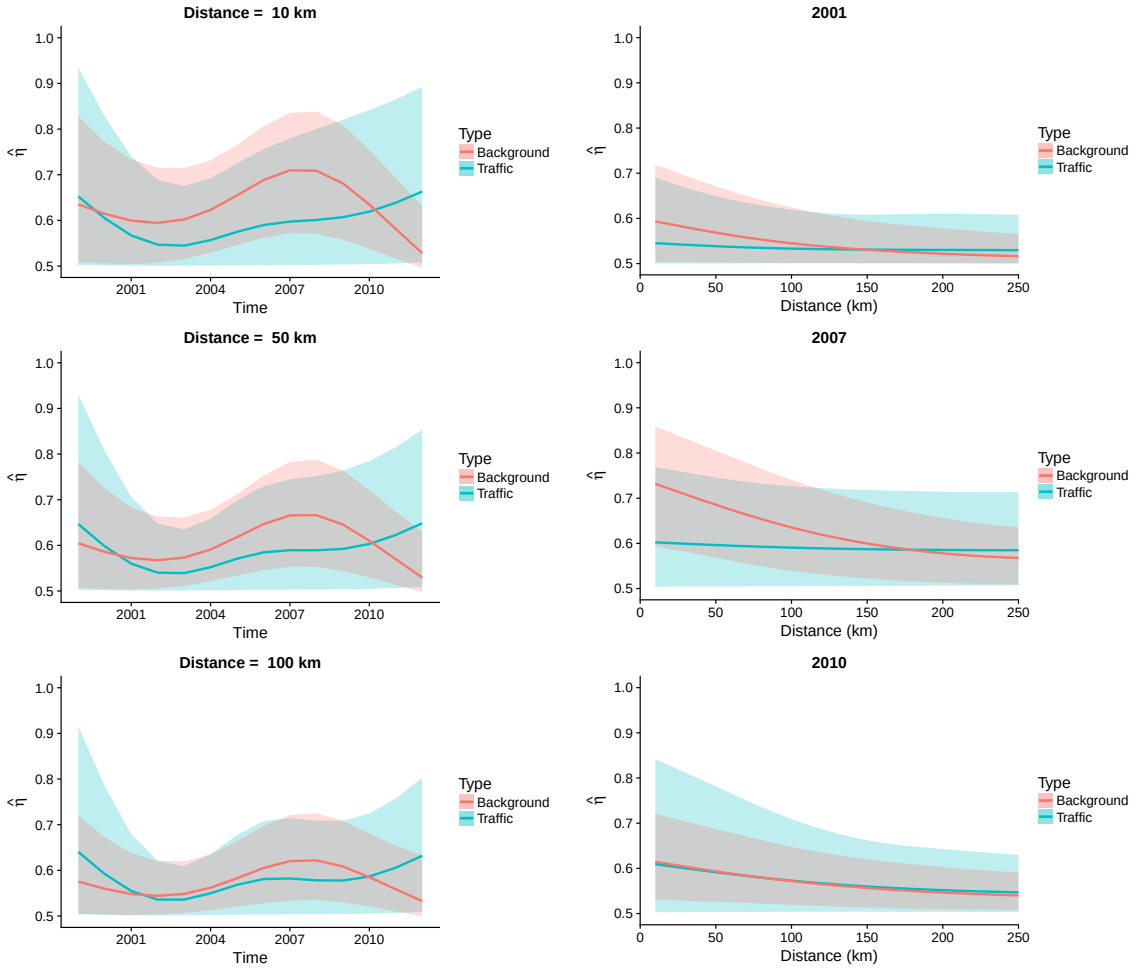

Figure 4.11: Cross sections of the mean bootstrap estimate of $\hat{\eta}(t, d)$ along with the corresponding 95% bootstrap uncertainty intervals. The bootstrap procedure is based on 300 bootstrap samples

concentrations of NO_2. Unreported results with a distance resolution of 5 km have shown globally higher tail dependence estimates (over time) for close traffic stations that are at most 5 km apart. The uncertainty in the estimators $\hat{\eta}(t, d, \text{“Traffic”})$ is relatively high due to the small amount of information available from the traffic stations. For background stations, the smooth effect of distance on the tail dependence is more pronounced with a decrease in the dependence at larger distances. The sharpness of the decrease shows some variation over time, with higher tail dependence for close background stations observed in 2007. The smooth effect of time on the tail dependence is important mainly for small distances with an overall slight increase over time for the traffic stations and a bump (increase) in the dependence around 2005 and 2009 for the background stations. These effects are observable for stations up to 100 km apart although with a lower magnitude of the dependence measure.

4.6 Conclusion

Starting from a d-dimensional random vector $\mathbf{X}$ with either asymptotic dependence or independence, we developed min- and max-projection techniques allowing us to simplify

the joint tail characterization problem to univariate modelling with well-understood exponential distributions for which generalized additive modelling under censoring is feasible and well-known from survival modelling. The exponential rate carries crucial information about the form and the rate of the joint tail decay in different directions. This setup facilitates flexible inference for the tail dependence as it is based on the excesses or deficits of a univariate exponential random variable while censoring observations that do not contribute to the joint tail, and it allows us to include multiple covariates of different types through the GAM framework. Although we focused on estimating the covariate influence for a fixed direction, we can apply the regularization technique of Mhalla et al. [2017a] in different directions to obtain smooth and valid estimates of the Pickands' dependence function or the angular dependence function under shape constraints and for a fixed set of covariates.

Our application demonstrates that it is useful to apply both projection techniques in practice to compare covariate-driven estimates for tail dependence summaries in each of the two asymptotic regimes. Our pairwise modelling of NO_2 measurements in France to investigate the effect of time and spatial distance on the joint tail behaviour showed strong evidence against asymptotic dependence. The results of our application gave strong support for asymptotic independence with estimated extremal coefficients close to 2 and the confidence intervals of tail dependence coefficients bounded away from 1. Our methodology constitutes an important step toward the distinction between asymptotic dependence and independence, when there is no clear evidence for one of the two models or when different model classes arise for different covariate configurations. Formal hypothesis testing of asymptotic dependence against asymptotic independence for a fixed set of covariates would be an important extension. As our censoring mechanisms select different observations according to the two models, likelihood-based tests are not directly applicable, but non-parametric tests of extremal dependence could provide guidance [e.g., Dey and Yan, 2015, Chapters 17 and 18]. Finally, depending on the application context, a threshold that varies with covariates and/or directions could be used to determine excesses and deficits, which could reduce or homogenize estimation bias and uncertainty.

Conclusion

The research presented in this thesis is centred around developing new models and inference methods for covariate-dependent extremal dependence structures, and has two main orientations. The first one concerns the development of non-parametric inference methods for tail dependence measures, and is tackled in Chapters 3 and 4. The second one relates to the incorporation of the effects of covariates on the tail structure based on (vector) generalized additive models, and is addressed in Chapters 2, 3, and 4.

In Chapter 2, we extend the parametric modelling of the extremal dependence, under the assumption of asymptotic dependence, to take into account the effects of a set of covariates. This extension relies on two main ingredients: the spectral representation of multivariate extreme value distributions and the vector generalized additive models. Relying on parametric models for the extremal dependence, we prove, under mild conditions, the consistency and asymptotic normality of the resulting penalized maximum likelihood estimator of the covariate-varying angular density. The novel approach to extremal dependence modelling is illustrated on temperature datasets recorded at two monitoring stations, and interesting dynamics between the considered covariates and the temperatures extremal dependence are uncovered. As an extension, it would be interesting to consider a full model that links the effect of covariates to both the margins and the extremal dependence, at once. Based on the developed inference framework, this is a natural extension of our method that would allow a more accurate uncertainty assessment by taking into account the uncertainty due to the marginal fitting –though the computational complexity would escalate even for moderate dimensions. Furthermore, in high dimensions where the bias in extremal dependence modelling is generally more pronounced, the use of a censored likelihood approach instead of the full Poisson likelihood would certainly lead to a substantial improvement of the performances of our approach. Although it seems entirely feasible to use censored methods probably combined with pairwise or higher-order composite likelihoods [Huser and Davison, 2013] in high dimensions, it is, however, less obvious to see how these inference techniques could be adapted in our VGAM context as second order derivatives of the log-likelihood must be derived. A potential alternative approach would be to consider a Bayesian hierarchical framework [Cooley et al., 2007] where models for the margins and the dependence structure are embedded in different layers. The inclusion of the effect of covariates on the different model parameters would be performed in the same spirit as in Opitz et al. [2018] where a Bayesian generalized additive modelling framework to estimate extreme quantiles over space and time in high dimensions is proposed, and inference is based on integrated nested Laplace approximations (INLA) [Rue et al., 2009] where Gaussian priors on the covariates are imposed.

In Chapter 3, we propose a novel estimator of the Pickands' dependence function. The estimation procedure is based on an original projection technique, termed the max-projection, allowing to handle the dependence in a multivariate max-stable random vector

through a univariate Beta random variable, and a regularization technique yielding valid estimators of the Pickands' function. By adapting the GAM infrastructure to the developed framework, non-stationarity in the Pickands' dependence function and extremal coefficients is flexibly modelled and no a priori assumptions on the structure of the extremal dependence are made. Estimation of the covariate-dependent Pickands' function is performed sequentially along different directions in the unit simplex and we mention that it is more appropriate in high dimensions to focus on the projection at the barycentre as it highly reduces the computational cost of our method while retrieving most of the information about the extremal dependence. The resulting estimators are found to be competitive with standard estimators in the absence of covariates, and we illustrate the approach on a temperature dataset in the U.S. where time and altitude are linked to the dependence between hot days and warm nights. The approach presented in this work relies on a two-step procedure to ensure the validity of the bivariate Pickands' dependence function and the (partial) validity of the extremal coefficient in higher dimensions. In the bivariate case, it would of course be more appealing to model the influence of the covariates on the Pickands' function and to regularize it at once. One can go some way towards this goal, by adapting the framework of the shape constrained additive models (SCAM) [Pya and Wood, 2015] to model the Pickands' function as a function of the covariates and the angular component, i.e., $A(\omega, \mathbf{x})$, and to impose at the same time, the convexity of the marginal smooth related to the angular component. While this approach would free us from imposing the convexity constraint in the cobs regularization step, we would nevertheless still require this second step to ensure the bound and border constraints. Moreover, the SCAM infrastructure dealing uniquely with univariate and bivariate shape constrained smooths, the incorporation of the convexity constraint in the GAM step would only be beneficial in the case where a unique covariate is handled so that the interaction between the angular component and this covariate can be properly modelled.

In Chapter 4, we devise a projection technique, in the same vein as the work presented in Chapter 3, that allows modelling the tail dependence in multivariate settings and under asymptotic dependence and asymptotic independence. We show how the min-projection results in an exponential random variable which carries the information regarding the structure of the extremal dependence under the assumption of asymptotic dependence and the information regarding the tail decay rate under the assumption of asymptotic independence. Building on the projection from the multivariate to the univariate setting, generalized additive models for the Pickands' and the angular dependence functions are constructed to assess the effect of a set of covariates on the tail dependence. We illustrate the approach on a large dataset of pollutants in France where we model the dependence between large nitrogen dioxide concentrations in pairs of stations as a function of the distance between the stations, the type of the stations, and time. The illustration is inspired by the use of spatio-temporal max-stable models in geostatistics, while abstracting away from the curse of dimensionality from which these models suffer, owing to the use of the GAMs at a univariate level. An interesting feature of the proposed min-projection relies on the ability of the resulting random variable to model, depending on whether its upper or its lower tail is considered, the tail dependence under both regimes of asymptotic dependence and asymptotic independence. Therefore, a natural extension of our work would be to consider a testing procedure to discriminate between these two regimes. One could perform, for a fixed set of covariates, two tests of exponentiality [Torabi et al., 2018] on the lower tail and on the exceedances in the upper tail. Thus, when one of the tests rejects the null hypothesis, we can conclude that the assumption of asymptotic depen-

dence or independence might be violated. We need to keep in mind that, as highlighted in the simulation study, the choice of the threshold level is crucial when retrieving the strength of the tail dependence or decay rates, and that consequently we should perform the tests at different threshold levels and combine the results to conclude on the type of the asymptotic regime. Additionally, and in the perspective of discriminating between the asymptotic regimes, we could consider an independence likelihood where the likelihoods obtained under the assumptions of asymptotic dependence and independence are concatenated and the Pickands' and tail dependence functions, at the same fixed direction in the unit simplex, are estimated simultaneously with standard errors adequately adjusted [Chandler and Bate, 2007]. This inference procedure would enable us to include a ridge- or a bayes-inspired penalty to exclude undesirable parameter estimates that prevent us from discriminating between the asymptotic regimes.

Throughout this thesis, we have seen how generalized additive models can be adapted to endow multivariate extreme value modelling with a higher flexibility and a wider applicability. The methods developed in this thesis allow one to model the dependence between extreme events while borrowing information from the working environment. Although we have only considered environmental applications in this thesis, the developed methods are applicable in a wide range of fields including finance where, for example, the dependence between joint extreme losses of stocks can be modelled as a function of their type of sector. When dealing with high-dimensional data, these methods are still applicable but subject to the usual caveats attached to both likelihood-based and non-parametric approaches to extreme value modelling. On one hand, the non-parametric approaches are expected to perform poorly in high dimensions compared to the parametric ones [Vettori et al., 2018] and the non-parametric estimators devised in this thesis are no exception to this curse of dimensionality. As a matter of fact, the proposed projection techniques reduce the information regarding the multivariate joint tail dependence to a univariate random variable requiring therefore a substantial amount of data when the dimension increases. On the other hand, likelihood inference for max-stable distributions is computationally demanding in large dimensions [Castruccio et al., 2016] and the increase of its complexity is more noticeable in the covariate-dependent framework considered in this thesis, due to the additional complexity stemming from the smooth modelling of the parameters. To circumvent these computational challenges, the common approach is to rely on composite likelihoods [Padoan et al., 2010, Varin et al., 2011, Huser and Davison, 2013, Sang and Genton, 2014]. Additionally to the resulting loss of efficiency relative to maximum likelihoods, these misspecified likelihoods require the complete specification of a joint model for the data, which might not be available in some applications, e.g., high-dimensional financial applications.

The dynamic modelling of extremes undertaken in this thesis represents an attempt to fill the gap between the developments of non-stationary marginal models and dependence models, but there is room for future improvements and extensions. For instance, the issue of uncovering the asymptotic class of the data or potentially how transitions between the asymptotic regimes are related to a set of predictors, can be tackled in a more "straightforward" way based on hybrid models. An ongoing work with Raphaël Huser aims at constructing generalized additive models for extremes based on copula families able to capture both asymptotic dependence and independence through a unique parameter ensuring the transition between the two classes [Huser and Wadsworth, 2018]. The Newton–Raphson optimization algorithm, used in this thesis in the presence of a smoothing penalty term, is rather cumbersome to implement when adding the covariate-

dependence component to this type of copulas, and we are considering an alternative approach based on the local likelihood [Davison and Ramesh, 2000], hence avoiding the addition of a penalty term. Finally, an interesting application of the developed methods, that will hopefully be soon undertaken, consists of discovering causal relationships [Pearl et al., 2016] between a tail dependence measure and a set of covariates. In the spirit of the causal inference approach of Peters et al. [2016] which relies on the invariance of a prediction under causal models, the idea would be to compare the regression coefficients when regressing the tail dependence measure (target variable) on a set of covariates, under different changes of the environment (e.g., different time periods), and to retain the causal predictors whose influences on the target variable remain invariant.

List of Tables

List of Figures

References

M. G. Akritas and D. N. Politis. *Recent Advances and Trends in Nonparametric Statistics*. Elsevier Science B.V., 2003. doi: http://dx.doi.org/10.1016/B978-044451378-6/50000-4.

R. S. Anderssen and P. Bloomfield. A time series approach to numerical differentiation. *Technometrics*, 16:69–75, 1974. doi: 10.1080/00401706.1974.10489151.

P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *The Annals of Applied Statistics*, 9:2023–2050, 2015.

A. A. Balkema and L. de Haan. Residual life time at great age. *The Annals of probability*, 2:792–804, 1974.

A. A. Balkema and S. I. Resnick. Max-infinite divisibility. *Journal of Applied Probability*, 14:309–319, 1977.

B. Basrak, R. A. Davis, and T. Mikosch. A characterization of multivariate regular variation. *The Annals of Applied Probability*, 12:908–920, 2002. doi: 10.1214/aoap/1031863174.

J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, D. De Waal, and C. Ferro. *Statistics of Extremes: Theory and Applications*. Wiley, New York, 2004.

M. Beniston. Variations of snow depth and duration in the Swiss Alps over the last 50 years: Links to changes in large-scale climatic forcings. *Climatic Change*, 36:281–300, 1997. doi: 10.1023/A:1005310214361.

M. Beniston. *Climatic Change and Its Impacts: An Overview Focusing on Switzerland*. Advances in Global Change Research. Springer, Netherlands, 2004.

M. Beniston. Warm winter spells in the Swiss Alps: Strong heat waves in a cold season? A study focusing on climate observations at the Saentis high mountain site. *Geophysical Research Letters*, 32:1–5, 2005. doi: 10.1029/2004GL021478.

M. Beniston. Linking extreme climate events and economic impacts: Examples from the Swiss Alps. *Energy Policy*, 35:5384–5392, 2007. doi: 10.1016/j.enpol.2006.01.032.

M. Beniston and M. Rebetez. Regional behavior of minimum temperatures in Switzerland for the period 1979–1993. *Theoretical and Applied Climatology*, 53:231–243, 1996. doi: 10.1007/BF00871739.

A. Bienvenüe and C. Y. Robert. Likelihood inference for multivariate extreme value distributions whose spectral vectors have known conditional distributions. *Scandinavian Journal of Statistics*, 44:130–149, 2017. doi: 10.1111/sjos.12245.

M.-O. Boldi and A. C. Davison. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B*, 69:217–229, 2007.

A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics*, 39:1963–2006, 2011. doi: 10.1214/11-AOS890.

P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84:567–577, 1997.

D. Castro and M. de Carvalho. Spectral density regression for bivariate extremes. *Stochastic Environmental Research and Risk Assessment*, 31:1603–1613, 2017. doi: 0.1007/s00477-016-1257-z.

S. Castruccio, R. Huser, and M. G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25:1212–1229, 2016. doi: 10.1080/10618600.2015.1086656.

R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94:167–183, 2007.

A. Charpentier, A.-L. Fougères, C. Genest, and J. G. Nešlehová. Multivariate Archimax copulas. *Journal of Multivariate Analysis*, 126:118–136, 2014. doi: http://dx.doi.org/10.1016/j.jmva.2013.12.013.

V. Chavez-Demoulin and A. C. Davison. Generalized Additive Modelling of Sample Extremes. *Journal of the Royal Statistical Society, Series C*, 54:207–222, 2005.

V. Chavez-Demoulin and A. C. Davison. Modelling time series extremes. *Revstat Statistical Journal*, 10:109–133, 2012.

V. Chavez-Demoulin, P. Embrechts, and M. Hofert. An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83:735–776, 2015.

Climate and Global Dynamics Division, National Center for Atmospheric Research, University Corporation for Atmospheric Research. Daily minimum and maximum temperature and precipitation for long term stations from the U.S. COOP data. accessed 7 april 2016., 2010. URL http://rda.ucar.edu/datasets/ds510.6/.

S. Coles. Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society, Series B*, 55:797–816, 1993.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.

S. Coles and M. J. Dixon. Likelihood-based inference for extreme value models. *Extremes*, 2:5–23, 1999. doi: 10.1023/A:1009905222644.

S. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B*, 53:377–392, 1991.

S. Coles and J. A. Tawn. Statistical methods for multivariate to structural design extremes: an application to structural design. *Journal of the Royal Statistical Society, Series C*, 43:1–48, 1994.

S. Coles, J. E. Heffernan, and J. A. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:339–365, 1999.

D. Cooley, P. Naveau, and P. Poncet. Variograms for spatial max-stable random fields. In P. Bertail, P. Doukhan, and P. Soulier, editors, *Dependence in Probability and Statistics*, volume 187, pages 373–390. Springer, New York, 2006.

D. Cooley, D. Nychka, and P. Naveau. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102:824–840, 2007. doi: 10.1198/016214506000000780.

D. Cooley, R. A. Davis, and P. Naveau. The Pairwise Beta Distribution: A Flexible Parametric Multivariate Model for Extremes. *Journal of Multivariate Analysis*, 101: 2103–2117, 2010.

E. Cormier, C. Genest, and J. G. Nešlehová. Using B-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes*, 17:633–659, 2014.

P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31:377–403, 1978. doi: 10.1007/BF01404567.

A. C. Davison. *Statistical Models*. Cambridge University Press, Cambridge, UK, 2003.

A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468:581–608, 2011. doi: 10.1098/rspa.2011.0412.

A. C. Davison and D. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

A. C. Davison and R. Huser. Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235, 2015.

A. C. Davison and N. Ramesh. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society, Series B*, 62:191–208, 2000. doi: 10.1111/1467-9868.00228.

A. C. Davison and R. L. Smith. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B*, 52:393–442, 1990.

A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modeling of spatial extremes. *Statistical Science*, 27:161–186, 2012. doi: 10.1214/11-STS376.

A. C. Davison, R. Huser, and E. Thibaud. Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45:511–529, 2013. doi: 10.1007/s11004-013-9469-y.

C. de Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer New York, 1978.

M. de Carvalho. Statistics of extremes: Challenges and opportunities. In F. Longin, editor, *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*. Wiley, Hoboken, 2016.

M. de Carvalho and A. C. Davison. Spectral Density Ratio Models for Multivariate Extremes. *Journal of the American Statistical Association*, 109:764–776, 2014. doi: 10.1080/01621459.2013.872651.

M. de Carvalho, B. Oumow, J. Segers, and M. Warchoł. A Euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics—Theory and Methods*, 42: 1176–1192, 2013. doi: 10.1080/03610926.2012.709905.

M. de Carvalho, D. Castro, and J. L. Wadsworth. Time-varying extreme value dependence with applications to leading European stock markets. *Annals of Applied Statistics*, 2018.

L. de Haan. *On regular variation and its application to the weak convergence of sample extremes*. Mathematical Centre tracts. Mathematisch Centrum, Amsterdam, 1970.

L. de Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12:1194–1204, 1984.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.

L. de Haan and S. I. Resnick. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40:317–337, 1977.

L. de Haan and C. Zhou. Extreme residual dependence for random vectors and processes. *Advances in Applied Probability*, 43:217–242, 2011.

P. Deheuvels. On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters*, 12:429–439, 1991.

D. Dey and J. Yan. *Extreme Value Modeling and Risk Analysis*. Chapman and Hall/CRC, New York, 2015.

C. Dombry, F. Éyi Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100:111–124, 2013. doi: 10.1093/biomet/ass067.

C. Dombry, S. Engelke, and M. Oesting. Asymptotic properties of likelihood estimators for multivariate extreme value distributions. *arXiv preprint arXiv:1612.05178*, 2016. URL http://arxiv.org/abs/1612.05178.

H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75:149–172, 1998.

D. J. Dupuis and C. A. Field. Robust estimation of extremes. *The Canadian Journal of Statistics*, 26:199–215, 1998.

E. F. Eastoe and J. A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society, Series C*, 58:25–45, 2009. doi: 10.1111/j.1467-9876.2008.00638.x.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 1996. doi: 10.1214/ss/1038425655.

J. H. J. Einmahl, L. de Haan, and D. Li. Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics*, 34:1987–2014, 2006. doi: 10.1214/009053606000000434.

J. H. J. Einmahl, J. Li, and R. Y. Liu. Thresholding events of extreme in simultaneous monitoring of multiple risks. *Journal of the American Statistical Association*, 104:982–992, 2009.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, New York, 1997.

S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society, Series B*, 77:239–265, 2015.

M. Escobar-Bach, Y. Goegebeur, and A. Guillou. Local robust estimation of the Pickands dependence function. hal-01340166, 2016.

M. Falk and R.-D. Reiss. A characterization of the rate of convergence in bivariate extreme value models. *Statistics and Probability Letters*, 59:341–351, 2002.

L. Fawcett and D. Walshaw. Markov chain models for extreme wind speeds. *Environmetrics*, 17:795–809, 2006. doi: 10.1002/env.794.

C. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B*, 65:545–556, 2003. doi: 10.1111/1467-9868.00401.

A. Fils-Villetard, A. Guillou, and J. Segers. Projection estimators of Pickands dependence functions. *Canadian Journal of Statistics*, 36:369–382, 2008.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

A.-L. Fougères. Multivariate extremes. *In Extreme Values in Finance, Telecommunications, and the Environment, Ed. Finkenstädt, B. and Rootzén, H., Monographs on Statistics and Applied Probability*, 99, 2004.

C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37:2990–3022, 2009.

B. V. Gnedenko. Sur la distribution limite du terme maximum d'une serie aléatoire. *The Annals of Mathematics*, 44:423–453, 1943.

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979. doi: 10.1080/00401706.1979.10489751.

P. J. Green. Iteratively reweighted least squares for maximum likelihood estimation , and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Series B*, 46:149–192, 1984.

P. J. Green. Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, 55:245–259, 1987.

P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.

J. Greenstadt. On the relative efficiencies of gradient methods. *Mathematics of Computation*, 21:360–367, 1967.

C. Gu. Cross-validating non-Gaussian data. *Journal of Computational and Graphical Statistics*, 1:169–179, 1992. doi: 10.1080/10618600.1992.10477012.

G. Gudendorf and J. Segers. Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis*, 102:37–47, 2011.

G. Gudendorf and J. Segers. Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, 142:3073–3085, 2012.

E. J. Gumbel. Bivariate logistic distributions. *Journal of the American Statistical Association*, 56:335–349, 1961.

P. Hall and N. Tajvidi. Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6:835–844, 2000a.

P. Hall and N. Tajvidi. Nonparametric analysis of temporal trend when fitting parametric models to extreme value data. *Statistical Science*, 15:153–167, 2000b. doi: 10.1214/ss/1009212755.

T. E. Hanson, M. de Carvalho, and Y. Chen. Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics and Probability Letters*, 128:60–66, 2017.

T. Hastie and R. Tibshirani. Generalized additive models. *Statistical Science*, 1:297–310, 1986. doi: 10.1214/ss/1177013604.

T. Hastie and R. Tibshirani. *Generalized Additive Models*. London: Chapman and Hall, 1990.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York, 2 edition, 2009.

X. He and P. Ng. Cobs: Qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14:315–337, 1999.

J. E. Heffernan. A directory of coefficients of tail dependence. *Extremes*, 3:279–290, 2000.

J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17:537–571, 2007. doi: 10.1214/105051606000000835.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society, Series B*, 66:497–546, 2004.

J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29:339, 1987. doi: 10.2307/1269343.

J. R. M. Hosking, J. R. Wallis, and E. F. Woo. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, 27: 251–261, 1985. doi: dx.doi.org/10.1080/00401706.1985.10488049.

T. Hsing, J. Hüsler, and M. R. Leadbetter. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78:97–112, 1988. doi: 10.1007/BF00718038.

X. Huang. *Statistics of Bivariate Extreme Values*. PhD thesis, Tinbergen Institute Research Series, 1992.

H. Hult and F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34:587–608, 2002. doi: 10.1239/aap/1033662167.

R. Huser and A. C. Davison. Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100:511–518, 2013. doi: 10.1093/biomet/ass089.

R. Huser and M. G. Genton. Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21:470–491, 2016.

R. Huser and J. L. Wadsworth. Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 2018. doi: 10.1080/01621459.2017.1411813.

R. Huser, A. C. Davison, and M. G. Genton. Likelihood estimators for multivariate extremes. *Extremes*, 19:79–103, 2016. doi: 10.1007/s10687-015-0230-4.

R. Huser, T. Opitz, and E. Thibaud. Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, 21:166–186, 2017. doi: https://doi.org/10.1016/j.spasta.2017.06.004.

R. Huser, T. Opitz, and E. Thibaud. Penultimate modeling of spatial extremes: statistical inference for max-infinitely divisible processes. *arXiv preprint arXiv:1801.02946*, 2018.

J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, 7:283–286, 1989.

H. Joe. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics and Probability Letters*, 9:75–81, 1990.

P. Jonathan, K. Ewans, and D. Randell. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188, 2014a. doi: 10.1002/env.2262.

P. Jonathan, D. Randell, Y. Wu, and K. Ewans. Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88:520–532, 2014b.

P. Jungo and M. Beniston. Changes in the anomalies of extreme temperature anomalies in the 20th century at Swiss climatological stations located at different latitudes and altitudes. *Theoretical and Applied Climatology*, 69:1–12, 2001. doi: 10.1007/s007040170031.

R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25:1287–1304, 2002. doi: 10.1016/S0309-1708(02)00056-8.

K. Knight. *Mathematical Statistics*. Chapman and Hall/CRC Press, Boca Raton, 2000.

R. Koenker, P. T. Ng, and S. Portnoy. Quantile smoothing splines. *Biometrika*, 81: 673–680, 1994.

S. Kotz and S. Nadarajah. *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press, 2000.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York, 1983.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996. doi: 10.1093/biomet/83.1.169.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B*, 59:475–499, 1997. doi: 10.1111/1467-9868.00080.

J. Luterbacher, M. A. Liniger, A. Menzel, N. Estrella, P. M. Della-Marta, C. Pfister, T. Rutishauser, and E. Xoplaki. Exceptional European warmth of autumn 2006 and winter 2007: Historical context, the underlying dynamics, and its phenological impacts. *Geophysical Research Letters*, 34:1–6, 2007. doi: 10.1029/2007GL029951.

G. Marcon, S. A. Padoan, P. Naveau, and P. Muliere. Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *arXiv preprint arXiv:1405.5228*, 2014.

G. Marcon, S. A. Padoan, and I. Antoniano-Villalobos. Bayesian inference for the extremal dependence. *Electronic Journal of Statistics*, 10:3310–3337, 2016.

G. Marra and S. N. Wood. Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55:2372–2387, 2011. doi: 10.1016/j.csda.2011.02.004.

G. Marra and S. N. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39:53–74, 2012. doi: 10.1111/j.1467-9469.2011.00760.x.

K. Maulik and S. I. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7:31–67, 2004.

A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7: 271–300, 2000.

L. Mhalla, V. Chavez-Demoulin, and P. Naveau. Non-linear models for extremal dependence. *Journal of Multivariate Analysis*, 159:49–66, 2017a.

L. Mhalla, M. de Carvalho, and V. Chavez-Demoulin. Regression type models for extremal dependence. *arXiv preprint arXiv:1704.08447*, 2017b. URL https://arxiv.org/abs/1704.08447.

L. Mhalla, T. Opitz, and V. Chavez-Demoulin. Exceedance-based nonlinear regression of residual dependence in extremes. *arXiv preprint arXiv:1802.01535*, 2018. URL https://arxiv.org/abs/1802.01535.

P. Naveau, A. Guillou, D. Cooley, and J. Diebolt. Modelling pairwise dependence of maxima in space. *Biometrika*, 96:1–17, 2009. doi: 10.1093/biomet/asp001.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, New York, 2nd ed edition, 2006.

W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245, 1994.

P. T. Ng and M. Maechler. A fast and efficient implementation of qualitatively constrained quantile smoothing splines. *Statistical Modelling*, 7:315–328, 2007.

P. T. Ng and M. Maechler. *COBS – Constrained B-splines (Sparse matrix based)*, 2015. URL http://CRAN.R-project.org/package=cobs. R package version 1.3-1.

P. J. Northrop and P. Jonathan. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809, 2011. doi: 10.1002/env.1106.

E. Omey and S. T. Rachev. Rates of convergence in multivariate extreme value theory. *Journal of Multivariate Analysis*, 38:36–50, 1991. doi: https://doi.org/10.1016/0047-259X(91)90030-6.

T. Opitz, R. Huser, H. Bakka, and H. Rue. INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *arXiv preprint arXiv:1802.01085*, pages 1–33, 2018. URL http://arxiv.org/abs/1802.01085.

S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105:263–277, 2010. doi: 10.1198/jasa.2009.tm08577.

F. Pauli and S. Coles. Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics*, 28:547–560, 2001. doi: 10.1080/02664760120047889.

J. Pearl, M. Glymour, and N. P. Jewell. *Causal Inference in Statistics: A Primer*. Wiley, 2016.

J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference by using invariant prediction : identification and confidence intervals. *Journal of the Royal Statistical Society, Series B*, 78:947–1012, 2016.

J. Pickands. The two-dimensional poisson process and extremal processes. *Journal of Applied Probability*, 8:745–756, 1971.

J. Pickands. Multivariate extreme value distributions. In *Proc. 43rd Session of the International Statistical Institute*, pages 859–878, 1981.

S.-H. Poon, M. Rockinger, and J. A. Tawn. Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, pages 929–953, 2003.

N. Pya and S. N. Wood. Shape constrained additive models. *Statistics and Computing*, 25:543–559, 2015. doi: 10.1007/s11222-013-9448-7.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

N. Ramesh and A. C. Davison. Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256:106–119, 2002.

B. Renard and M. Lang. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30:897–912, 2007.

S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 1987.

S. I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.

P. Ressel. Homogeneous distributions–And a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis*, 117: 246–256, 2013.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

D. Ruppert. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11:735–757, 2002. doi: 10.1198/106186002853.

D. Ruppert, M. Wand, and R. Carroll. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003. ISBN 9780521785167.

A. Sabourin and P. Naveau. Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics and Data Analysis*, 71:542–567, 2014.

H. Sang and A. E. Gelfand. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16:407–426, 2009. doi: 10.1007/s10651-007-0078-0.

H. Sang and A. E. Gelfand. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:49–65, 2010. doi: 10.1007/s13253-009-0010-1.

H. Sang and M. G. Genton. Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*, 8:86 – 103, 2014. doi: https://doi.org/10.1016/j.spasta.2013.07.003. URL http://www.sciencedirect.com/science/article/pii/S2211675313000420.

C. J. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Revstat Statistical Journal*, 10:33–60, 2012.

M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002.

M. Schlather and J. A. Tawn. Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes*, 5:87–102, 2002.

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

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

P. Shi, P.-H. Xie, M. Qin, F.-Q. Si, K. Dou, and K. Du. Cluster analysis for daily patterns of SO2 and NO2 measured by the DOAS system in Xiamen. *Aerosol and Air Quality Research*, 14:1455–1465, 2014. doi: 10.4209/aaqr.2013.05.0160.

M. Sibuya. Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11:195–210, 1960.

R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–90, 1985.

R. L. Smith. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4:367–377, 1989. doi: 10.1214/ss/1177012400.

R. L. Smith. *Max-stable processes and spatial extremes*. PhD thesis, University of Surrey, 1990. unpublished thesis.

R. L. Smith and I. Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society, Series B*, 56:515–528, 1994.

J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77:245–253, 1990.

H. Torabi, N. H. Montazeri, and A. Grané. A wide review on exponentiality tests and two competitive proposals with application on reliability. *Journal of Statistical Computation and Simulation*, 88:108–139, 2018. doi: 10.1080/00949655.2017.1379522.

C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42, 2011.

T. Vatter and V. Chavez-Demoulin. Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147–167, 2015.

S. Vettori, R. Huser, and M. G. Genton. A comparison of dependence function estimators in multivariate extremes. *Statistics and Computing*, 28:525–538, 2018. doi: 10.1007/s11222-017-9745-7.

J. L. Wadsworth and J. A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99:253–272, 2012. doi: 10.1093/biomet/asr080.

J. L. Wadsworth and J. A. Tawn. A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714, 2013.

J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101:1–15, 2014. doi: 10.1093/biomet/ast042.

G. Wahba. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 13:1378–1402, 1985. doi: 10.1214/aos/1176349743.

H. Wang and C.-L. Tsai. Tail index regression. *Journal of the American Statistical Association*, 104:1233–1240, 2009.

S. N. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B*, 62:413–428, 2000. doi: 10.1111/1467-9868.00240.

S. N. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99:673–686, 2004. doi: 10.1198/016214504000000980.

S. N. Wood. On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, 48:445–464, 2006a. doi: 10.1111/j.1467-842X.2006.00450.x.

S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Texts in Statistical Science, 2006b.

S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

S. N. Wood, N. Pya, and B. Säfken. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111:1548–1563, 2016. doi: 10.1080/01621459.2016.1180986.

T. W. Yee. *Vector Generalized Linear and Additive Models: With an Implementation in R*. Springer, New York, 1st edition, 2015.

T. W. Yee. Comment on: ≪Smoothing parameter and model selection for general smooth models,≫ by S. N. Wood, N. Pya, and B. Säfken. *Journal of the American Statistical Association*, 111:1565–1568, 2016.

T. W. Yee. *VGAM: Vector Generalized Linear and Additive Models*, 2017. URL https://CRAN.R-project.org/package=VGAM. R package version 1.0-4.

T. W. Yee and A. G. Stephenson. Vector generalized linear and additive extreme value models. *Extremes*, 10:1–19, 2007. doi: 10.1007/s10687-007-0032-4.

T. W. Yee and C. J. Wild. Vector generalized additive models. *Journal of the Royal Statistical Society, Series B*, 58:481–493, 1996.

M.-H. Zhang. Modelling total tail dependence along diagonals. *Insurance: Mathematics and Economics*, 42:73–80, 2008. doi: 10.1016/j.insmatheco.2007.01.002.