Multivariate Threshold Models with Applications to Wind Speed Data

Ph.D. Thesis

Pál Rakonczai

Supervisor:

András Zempléni, Associate Professor, CSc

DOCTORAL SCHOOL OF MATHEMATICS

Director: Professor Miklós Laczkovich, member of HAS

DOCTORAL PROGRAMME OF APPLIED MATHEMATICS

Director: Professor György Michaletzky, DSc

EÖTVÖS LORÁND UNIVERSITY

Faculty of Science, Institute of Mathematics

Department of Probability Theory and Statistics

Budapest, 2012

Acknowledgements

First I would like to thank András Zempléni for his excellent supervision and for spending so much time on correcting my mistakes even at weekends and late evenings. I also thank Nader Tajvidi, Georg Lindgren, Feridun Turkman, Adam Butler and to László Márkus for their very valuable contribution. The financial support of my research work has been provided by the following grants SEAMOCS (Lund, Sweden), HPC2-Europa (Edinburgh, UK), TÁMOP - 4.2.1/B-09/KMR-2010-0003 (Budapest, Hungary), OTKA - MB08A 84576 (Lisbon, Portugal), I thank everybody who helped me to get across. Finally I would like to express my sincere thanks to my mom for her love and support, and to my true friends for the constant encouragement.

Contents

Chapter 1

Introduction

In this thesis I investigate and develop methods for joint modeling of extremely high values (extremes) of multivariate observations. The thesis consists of a short theoretical overview of the multivariate extreme value theory and a detailed presentation of my own scientific contribution during the recent 5 years. The initial motivation for the research was established within the framework of an applied statistical project [1] called "Applied stochastic models for ocean engineering, climate and safe transportation", where modeling simultaneously appearing high wind speeds - monthly maxima or exceedances over high thresholds - at different sites was one of the main research objective. The scope of the presented methods is much wider though, see e.g. the M.Sc. thesis[2] of Krusper (2011) for actuarial and financial applications or the presentation of Zempléni and Rakonczai (2011) for hydrological applications for Danube River, see Figure 1.1.

The most important impact of this thesis is in statistical inference and applications, providing useful material for practitioners of various disciplines working on extreme values, but simultaneously I present some theoretical considerations about certain model properties and conditions of applicability. I devoted substantial work to the development of a new `R` software package called `mgpd` which has been first

[1] The research visit was granted by Lund University, Sweden, see website http://www.maths.lth.se/seamocs for further details.

[2] My research leading to this thesis was partially supported by the TÁMOP project 4.2.1./B – 09/KMR – 2010 – 0003.

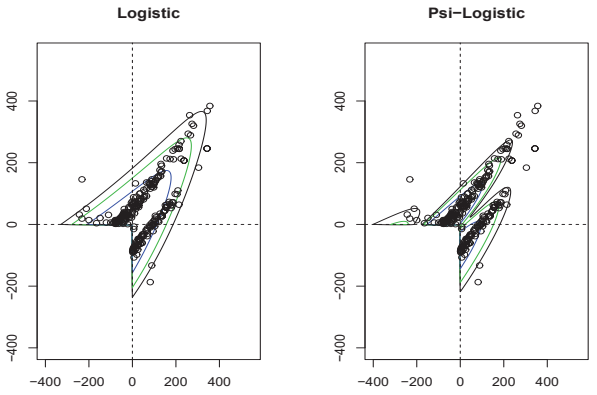

Figure 1.1: Flexibility of the constructed new models. Left panel: density curves of logistic baseline model, right panel: a flexible new BGPD model from the logistic baseline.

published in Rakonczai (2011) for modeling bivariate exceedances. It is freely available at the `Comprehensive R Archive Network` as a part of the `Distributions` task. It is now widely used by researchers of the field, and as the maintainer of the package I develop it following the demands of the users. The practical applications have been carried out by this package as well and all of them are easy to reproduce by running the cited code parts.

I outline the main probabilistic results providing the basis of modeling multivariate extremes in chapter 2. In chapter 3, I propose an universal method of constructing families of asymmetric dependence structures illustrated by some examples. At the end of this chapter I also introduce a novel statistical tool called autocopula. In chapter 4 goodness-of-fit and simulation methods are shown first, and then I present some simulation studies approving the plausibility of the proposed models and methods. Finally, numerous 2 dimensional (2D) and 3 dimensional (3D) applications for

wind data can be found in chapter 5, and further promising ideas in chapter 6.

Related publications with my contribution are the followings:

- Rakonczai, P. and Turkman, F. (2012) Applications of generalized Pareto processes. (Technical report under progress, OTKA outgoing mobility grant, Lisbon, Portugal)
- Rakonczai, P. (2012) Asymmetric dependence models for bivariate threshold exceedance models. *Forum Statisticum Slovacum, ISSN 1336-7420* **1**, p.25-32.
- Rakonczai, P. and Zempléni, A. (2012) Bivariate generalized Pareto distribution in practice: models and estimation. *Environmetrics*, John Wiley & Sons, **23**, p.219-227.
- Zempléni, A. and Rakonczai, P. (2011) New bivariate threshold models with hydrological applications. *Conference on Environmental Risk and Extreme Events*, Ascona, July 10-15
- Rakonczai, P., Márkus, L. and Zempléni, A. (2011) Autocopulas: investigating the interdependence structure of stationary time series. *Methodology and Computing in Applied Probability*, **14**, p.149-167.
- Rakonczai, P. (2011) `Package 'mgpd' manual`.
  see http://cran.r-project.org/web/packages/mgpd/mgpd.pdf
- Rakonczai, P. and Tajvidi, N. (2010) On prediction of bivariate extremes. *International Journal of Intelligent Technologies and Applied Statistics*, **3**(2), p.115-139.
- Rakonczai, P., Butler, A. and Zempléni, A. (2010) Modeling temporal trend within bivariate generalized Pareto models of logistic type.
  (Technical report, HPC-Europa2 Project, Edinburgh, UK, available at http://www.math.elte.hu/∼paulo/pdf/)
- Rakonczai, P. (2009) On Modelling and Prediction of Multivariate Extremes, with applications to environmental data. Centrum Scientiarum Mathematicarum, Licentiate Theses in Mathematical Sciences 2009:05

- Rakonczai, P., Márkus L. and Zempléni, A. (2008a) Goodness of Fit for Auto-Copulas: Testing the Adequacy of Time Series Models, *Proceedings of the 4th International Workshop in Applied Probability* CD-ROM, paper No.73., 6 pages, Compiegne, France

- Rakonczai, P., Márkus L. and Zempléni, A. (2008b) Adequacy of Time Series Models, Tested by Goodness of Fit for Auto-Copulas, *Proceedings of the COMPSTAT2008 conference*, Porto, Portugal

- Rakonczai, P. and Zempléni, A. (2007) Copulas and goodness of fit tests. *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific, Hackensack, NJ, p.198-206.

- Bozsó, D., Rakonczai, P. and Zempléni, A. (2005) Árvizek a Tiszán és néhány mellékfolyóján. *Statisztikai Szemle*, **83**(10-11), p.919-936.

Chapter 2

Extreme Value Theory

2.1 Univariate Extreme Value Theory

Here we outline the main probabilistic results providing the basis of parametric modeling of univariate extremes. These approaches are expanded when our attention turns to the multivariate questions. To reveal the motivation behind extreme value theory (EVT), let $X_1, ..., X_n$ be a sequence of independent random variables with common distribution function F. In addition let $M_n = \max(X_1, X_2, \ldots, X_n)$ be the maximum of the sequence. The variables X_i often represent hourly or daily values of a process and so M_n represents the maximum of the process over n time units. The distribution function of M_n can be computed in a very elementary way as

$$P(M_n \leq z) = P(X_1 \leq, ..., X_n \leq z) = \prod_{i=1}^{n} P(X_i \leq z) = F^n(z). \tag{2.1}$$

However Equation 2.1 is not very useful in practice if F is unknown. Of course, one may suggest to estimate F from the measurements in some way and use this as a plug-in estimate in $\hat{F}^n(z)$. Unfortunately by doing this, even small differences between F and $\hat{F}$ might be multiplied up, leading to large error in the final estimate of F^n. An alternative solution is proposed by EVT, suggesting to look for approximate distribution families for F^n directly, based on the extreme measurements only. Central limit theory for extreme values (without proofs) is provided below.

2.1.1 Limit for Maxima

For the maximum of univariate i.i.d. variables the theory is well-elaborated. Since analogous statements follow for the minimum as

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n),$$

we can limit our attention to the case of maximum. Let

$$z_+ = \sup\{z : F(z) < 1\}$$

denote the upper endpoint of the support of the distribution $F(x)$. Then it is clear that $M_n \to z_+$ a.s. as $n \to \infty$. Thus, in order to get nondegenerate limit for M_n, we consider normalized maxima

$$M_n^* = \frac{M_n - a_n}{b_n},$$

for some sequences of constants $\{a_n\}$ and $\{b_n\} > 0$. The Gnedenko-Fisher-Tippett theorem states that the limit distribution, if exists, is in the class of the so-called extreme value distributions (EVD).

Definition 1. *The extreme value distribution with shape parameter ξ has the following distribution function.*
If $\xi \neq 0$,

$$G_\xi(x) = \exp\left[-(1+\xi x)^{-1/\xi}\right]$$

for $1 + \xi x > 0$ (otherwise 0 if $\xi > 0$ and 1 if $\xi < 0$).
If $\xi = 0$,

$$G_\xi(x) = \exp\left[-e^{-x}\right].$$

The $\xi = 0$ case can also be obtained from the $\xi \neq 0$ case by letting $\xi \to 0$. The limit distribution is called Fréchet for $\xi > 0$, Gumbel or double exponential for $\xi = 0$ and Weibull for $\xi < 0$.

One may also define the corresponding location-scale family $G_{\xi,\mu,\sigma}$ by replacing x above by $(x-\mu)/\sigma$ for $\mu \in \mathbb{R}$ and $\sigma > 0$ and changing the support accordingly. It is straightforward (see for example Embrechts et al.,1997) to check that Gumbel, Fréchet and Weibull families can be combined into a single family as follows.

Definition 2. *The generalized extreme value (GEV) distribution is defined as*

$$G_{\xi,\mu,\sigma}(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}, \tag{2.2}$$

where $1+\xi\frac{x-\mu}{\sigma} > 0$, $\mu \in \mathbb{R}$ *is called the location parameter,* $\sigma > 0$ *the scale parameter and* $\xi \in \mathbb{R}$ *the shape parameter.*

Theorem 1. ***[Fisher and Tippett (1928), Gnedenko (1943)]*** *If there exist* $\{a_n\}$ *and* $\{b_n\} > 0$ *sequences such that*

$$P(M_n^* \leq z) = P\left(\frac{M_n - a_n}{b_n} \leq z\right) \to G(z) \text{ as } n \to \infty \tag{2.3}$$

where G *is a nondegenerate distribution function, then* G *necessarily belongs to the GEV family, defined in Equation 2.2. In this case we say that the distribution of* X_i *belongs to the max-domain of attraction of the GEV distribution* G*.*

This theorem is usually used in practical applications for modeling the maxima of observations appearing in consecutive blocks of time (*block maxima*), as e.g. annual/ monthly/ weekly maxima.

Remark 1. *From the statistical point of view the apparent difficulty is that the normalizing constants are unknown. This can be easily solved in practice, as the distribution of the non-normalized maxima can be approximated by GEV distribution with different location and scale parameters:*

$$P(M_n \leq z) \sim G\left(\frac{z - a_n}{b_n}\right) = G^{\dagger}(z).$$

2.1.2 Limit for Threshold Exceedances

Modeling only the block maxima can be inefficient. As EVT is basically concerned with modeling the tail of an unknown distribution, a natural idea is to model all of those observations X_i, whose values are larger than a considerably high threshold. Due to the results of Balkema and de Haan (1974) and Pickands (1975) it is well-known that if the distribution of X_i lies within the max-domain of attraction of a GEV distribution, then the distribution of the threshold exceedances has a similar limiting representation. The results are summarized in the following theorem.

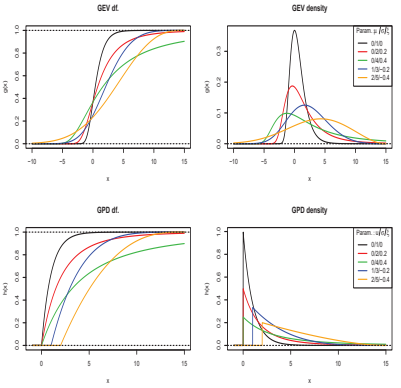

Figure 2.1: Distribution and density functions of GEV distribution (Equation 2.2) and GPD (Equation 2.4).

Theorem 2. *Let $X_1, ..., X_n$ be a sequence of independent random variables with common distribution function F. Suppose that F belongs to the max-domain of attraction of a GEV distribution for some ξ, μ and $\sigma > 0$. Then for high thresholds u*

$$P(X_i - u \leq z | X_i > u) \to H(z) = 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad \text{as } u \to z_+, \tag{2.4}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

The family defined in Equation 2.4 is called generalized Pareto distribution (GPD).

Remark 2. *Note, that both of the above limit results are strongly linked in the sense that, as the threshold tends to the right endpoint of the underlying distribution, the conditional distribution of the exceedances converges to GPD if and only if the distribution of the normalized maxima converges to GEV distribution. For graphical illustrations see Figure 2.1.*

2.1.3 Conditions for the Limit Theorems

The EVT-based statistical procedures implicitly assume that most distributions of practical interest lie within the max-domain of attraction of a GEV distribution (or equivalently, within a GPD). Therefore, a natural question arises: how general is the class of distributions for which the limit results in Theorem 1 and Theorem 2 hold? Although it is not difficult to find counterexamples (e.g. among discrete distributions), the most well-known continuous distributions belong to this class. If the shape parameter is nonzero then there are relatively easily verifiable conditions, using the following definition.

Definition 3. *We say that a distribution tail $\bar{F}$ is regularly varying with index $-\alpha$ for some $\alpha \geq 0$ if for every $x > 0$*

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{\alpha}.$$

In addition if $\alpha = 0$, the function $\bar{F}$ is said to be slowly varying.

Theorem 3 (Max-domain of attraction of Fréchet distribution)**.** *A distribution function F belongs to the max-domain of attraction of a GEV distribution with $\xi > 0$ (Fréchet-type) if and only if the distribution tail $\bar{F}$ is regularly varying with index $-\xi$.*

This condition is satisfied by e.g. the *Pareto*, the *Cauchy* and the *stable* (for $\alpha < 2$) distributions.

Theorem 4 (Max-domain of attraction of Weibull distribution)**.** *A distribution function F belongs to the max-domain of attraction of a GEV distribution with $\xi < 0$ (Weibull-type) if and only if the support of F is bounded to the right (with $x_+ < \infty$ is the right endpoint) and $\bar{F}(x_+ - x^{-1})$ is regularly varying with index $-\xi$.*

In contrast to the heavy-tailed distributions, the Weibull case contains distributions which have a finite right endpoint including e.g. the *uniform* and the *beta* distributions.

The $\xi = 0$ (Gumbel-type) case is more complicated. Although there exist necessary and sufficient conditions here as well, they are hardly used in practice. It can be shown that the max-domain of attraction of the Gumbel distribution covers quite a wide range of families of distribution functions. It contains distributions from heavy-tailed distributions whose all moments are finite (e.g. the *lognormal* distribution) to light-tailed distributions (e.g. the *normal*, the *exponential* or the *gamma* distribution) and even some distributions whose support is bounded to the right are possible. More details and further references can be found about the three above cases e.g. in Section 2.3,2.4 and 2.5 in Beirlant et al. (2004), respectively.

2.2 Modeling Multivariate Maxima

Comparing the multivariate problem with the univariate case the new issue that arises is the dependence structure. In such a case - beyond the marginal distributions - we must be able to determine how the individual variables relate to each other. The main question is describing the class of possible dependence structures and then to investigate how can we estimate them. Modeling multivariate extremes typically consists of two distinct steps: modeling univariate margins and then - after the suitable standardization of margins - modeling the dependence. As the first step involves only applying the univariate models of the previous section, here we focus on the second one, namely on characterizing the dependence structures. The first fundamental question one is confronted with is how to define the multivariate extreme events, as there exists no unique ordering for multivariate observations. Barnett (1976) considers several different categories of order relations. In the subsequent sections, unless mentioned otherwise, all operations and order relations on vectors are understood to be component-wise, i.e. for d-dimensional vectors $\mathbf{x} = (x_1, ..., x_d)$ and $\mathbf{y} = (y_1, ..., y_d)$ the relation $\mathbf{x} \leq \mathbf{y}$ is defined as $x_j \leq y_j$ for all $j = 1, ..., d$. In this case the maximum is defined by taking the component-wise maxima, which is defined as

$$\mathbf{x} \vee \mathbf{y} = (x_1 \vee y_1, \ldots, x_d \vee y_d),$$

where $\vee$ stands for the maximum (analogously, $a \wedge b = \min(a, b)$). By using this notation the maximum of a sample of d-dimensional observations
$\mathbf{X}_i = (X_{i,1}, ..., X_{i,d})$ for $i = 1, ..., n$ is defined as

$$\mathbf{M}_n = (M_{n,1}, ..., M_{n,d}) = \Big(\bigvee_{i=1}^{n} X_{i,1}, ..., \bigvee_{i=1}^{n} X_{i,d}\Big).$$

Finally, it should be mentioned that we can focus on maximum without loss of generality, since the following relation allows us to get the minimum by the help of the maximum of the negatives:

$$\bigwedge_{1}^{n} \mathbf{X_i} = -\bigvee_{1}^{n} (-\mathbf{X_i}).$$

Remark 3. *The sample maximum is not necessarily a sample point. From a practical point of view this means that the maxima we intend to model need not be simultaneous.*

2.2.1 Limit for Multivariate Maxima

Analogously to the univariate case we assume that $\mathbf{X}$ has distribution function $\mathbf{F}$ and there exist $\mathbf{a}_n$ and $\mathbf{b}_n > \mathbf{0}$ sequences of normalizing vectors, such that

$$P\Big(\frac{\mathbf{M}_n - \mathbf{a}_n}{\mathbf{b}_n} \leq \mathbf{z}\Big) = \mathbf{F}^n(\mathbf{b}_n\mathbf{z} + \mathbf{a}_n) \to \mathbf{G}(\mathbf{z}), \tag{2.5}$$

where the G_i margins of the limit distribution $\mathbf{G}$ are nondegenerate distributions. If Equation 2.5 holds then $\mathbf{F}$ is said to be in the domain of attraction of $\mathbf{G}$ and $\mathbf{G}$ itself is said to be a multivariate extreme value distribution (MEVD). Since Equation 2.3 holds for each margin

$$P\Big(\frac{M_{n,j} - a_{n,j}}{b_{n,j}} \leq z_j\Big) \to G_j(z_j) \text{ as } n \to \infty \tag{2.6}$$

for any $j = 1, ..., d$, where the d.f. G_j are nondegenerate by assumption. The margins are necessarily GEV distributions. Hence the essential part of the multivariate extension reduces to handling the dependence structure among the margins. It can be shown that the MEVD cannot be characterized as a parametric family indexed by a finite dimensional parameter vector (in contrary to in the GEV case). Instead, the family of MEVD is usually indexed by the class of the underlying dependence structures. Several authors, among them Resnick (1987) and Pickands (1981),have proposed equivalent characterizations of MEVD, assuming differently standardized margins. The benefit of standardization is that some properties or characterizations are more naturally seen for a given choice (see subsection 2.2.2, subsection 2.2.9 for details.) However, of course, the choice of the marginal distribution itself does not make any difference after transforming the margins back to the original GEV scale.

Remark 4. *Since the margins are continuous,* $\mathbf{G}$ *is continuous, so the convergence in Equation 2.5 holds not only in distribution but also for every* $x \in \mathbb{R}$ *uniformly.*

A useful characterization of MEVD can be given by the next definition.

Definition 4. *A multivariate distribution function* $\mathbf{G}$ *is called max-stable, if for every positive integer* k *there exist* α_k *and* $\beta_k > 0$ *vectors such that*

$$\mathbf{G}^k(\beta_k \mathbf{x} + \alpha_k) = \mathbf{G}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}.$$

It is not difficult to see that the classes of extreme value and max-stable distribution functions coincide (see 8.2.1 in Beirlant et al., 2004).

2.2.2 Exponent Measure

A further consequence of the max-stability is that $\mathbf{G}^{1/k}$ is a valid distribution function for every positive integer k. In such a case we say that the distribution function $\mathbf{G}$ is max-infinitely divisible (Balkema and Resnick, 1977). Specially, there exist a (unique) measure μ on $[\mathbf{q}, \infty) \setminus \{\mathbf{q}\}$, such that

$$\mathbf{G}(\mathbf{x}) = \exp\big(-\mu([\mathbf{q}, \infty) \setminus [\mathbf{q}, \mathbf{x}])\big), \tag{2.7}$$

where $\mathbf{q} = (q_1, \ldots, q_d)$ and $q_i = \inf\{x \in \mathbb{R} : G_i(x) > 0\}$ is the lower end-point of the ith margin. This μ measure is called the exponent measure.

2.2.3 Standardized Margins

When studying the dependence structure of a MEVD, it is convenient to standardize the margins so that they are all the same. Although the choice of marginal distribution might not look very relevant at the first glance, a smart choice can bring further mathematical simplicity. By using unit Fréchet margins with distribution function $\psi(x) = e^{-1/x}$, $x > 0$ the exponent measure satisfies a useful homogeneity property. (Another choice for margins is discussed later in subsection 2.2.9.)

If $(X_1, \ldots, X_d)$ denotes a random vector with a distribution function $\mathbf{G}$ then $(-1/\log G_1(X_1), ..., -1/\log G_d(X_d))$ has unit Fréchet distribution margins. Its joint distribution function $\mathbf{G}_\star$ can be written as

$$\mathbf{G}_\star(\mathbf{y}) = \mathbf{G}(G_1^{\leftarrow}(e^{-1/y_1}), ..., G_d^{\leftarrow}(e^{-1/y_d})), \quad \mathbf{y} \in \mathbb{R}_+^d,$$

where $G_i^{\leftarrow}$ denotes the quantile function of G_i, the ith margin of $\mathbf{G}$. Conversely

$$\mathbf{G}(\mathbf{x}) = \mathbf{G}_\star(-1/\log G_1(x_1), ..., -1/\log G_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $-1/\log(0) = 0$ and $-1/\log(1) = \infty$. The exponent measures μ and $\mu_\star$ of $\mathbf{G}$ and $\mathbf{G}_\star$ are related in the following way. For $\mathbf{x} \in [\mathbf{q}, \infty]$

$$\mu\big([\mathbf{q}, \infty) \setminus [\mathbf{q}, \mathbf{x}]\big) = \mu_\star\big([\mathbf{0}, \infty) \setminus [\mathbf{0}, -\mathbf{1}/\log \mathbf{G}(\mathbf{x})]\big),$$

moreover, for $\mu_\star$, the following homogeneity property is fulfilled

$$s\mu_\star\big(s([\mathbf{0}, \infty) \setminus [\mathbf{0}, \mathbf{y}])\big) = \mu_\star\big([\mathbf{0}, \infty) \setminus [\mathbf{0}, \mathbf{y}]\big), \quad \mathbf{y} \in [\mathbf{0}, \infty); 0 < s < \infty.$$

The above relation actually holds for all Borel subsets of $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$, namely

$$\mu_\star(s\cdot) = s^{-1}\mu_\star(\cdot), \quad 0 < s < \infty.$$

2.2.4 Spectral Representation

Due to this homogeneity a very flexible representation can be given in terms of polar coordinates. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be arbitrary norms on $\mathbb{R}^d$ and $\mathbb{S}_2 = \{\omega \in \mathbb{R}^d : \|\omega\|_2 = 1\}$ the unit sphere with respect to the second norm. Then define the mapping T from $\mathbb{R}^d \setminus \{\mathbf{0}\}$ to $(0, \infty) \times \mathbb{S}_2$ by $T(\mathbf{y}) = (r, \omega)$ where $r = \|\mathbf{y}\|_1$ is the radial part and $\omega = \mathbf{y}/\|\mathbf{y}\|_2$ is the angular part of $\mathbf{y}$. Now define an S measure for any B Borel subsets of $\Xi = \mathbb{S}_1 \setminus [\mathbf{0}, \infty)$ by

$$S(B) = \mu_\star\Big(\{\mathbf{y} \in [\mathbf{0}, \infty) : \|\mathbf{y}\|_1 \geq 1, \mathbf{y}/\|\mathbf{y}\|_2 \in B\}\Big).$$

The measure S - called spectral measure - is determined uniquely by the exponent measure and the chosen norms. Furthermore by the homogeneity of $\mu_\star$ we can see that

$$\mu_\star\Big(\{\mathbf{y} \in [\mathbf{0}, \infty) : \|\mathbf{y}\|_1 \geq r, \mathbf{y}/\|\mathbf{y}\|_2 \in B\}\Big) = r^{-1}S(B)$$

for $0 < r < \infty$ and B Borel subsets in Ξ. This leads to the so-called spectral decomposition of the exponent measure given by de Haan and Resnick (1987)

$$\mu_\star \circ T^{-1}(dr, d\omega) = r^{-2} dr S(d\omega),$$

which can be used to calculate the integral of a real-valued function f on $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$ with respect to $\mu_\star$ as follows

$$\begin{aligned} \int_{[\mathbf{0},\infty)\setminus\{\mathbf{0}\}} f(\mathbf{y})\mu_\star(d\mathbf{y}) &= \int_\Xi \int_0^\infty f(r\omega/\|\omega\|_1) r^{-2} dr S(d\omega) \\ &= \int_\Xi \int_0^\infty f(r\omega) r^{-2} dr \|\omega\|_1^{-1} S(d\omega). \end{aligned}$$

The exponent measure can be written as follows

$$\begin{aligned} V_\star(\mathbf{y}) &= \mu_\star\{[\mathbf{0},\infty) \setminus [\mathbf{0},\mathbf{y})\} = -\log \mathbf{G}_\star(\mathbf{y}) \\ &= \int_{[\mathbf{0},\infty)\setminus\{\mathbf{0}\}} \mathbf{1}\Big\{\bigvee_{i=1}^{d} \frac{z_i}{y_i} > 1\Big\}\mu_\star(d\mathbf{z}) \\ &= \int_\Xi \bigvee_{j=1}^{d} \Big(\frac{\omega_j}{\|\omega\|_1}\frac{1}{y_j}\Big) S(d\omega), \quad \mathbf{y} \in [\mathbf{0},\infty], \end{aligned} \tag{2.8}$$

where $V_\star$ is called the **exponent measure function**, which will be used later for simplicity in the notation. Moreover the constraints for standard Fréchet margins can be formulated as the following d equations

$$\int_\Xi \frac{\omega_j}{\|\omega\|_1} S(d\omega) = 1, \quad j = 1,\dots,d. \tag{2.9}$$

Remark 5. *Conversely, any positive measure S on Ξ satisfying Equation 2.9, is the spectral measure of the d-variate MEVD $\mathbf{G} = \exp(-V_\star)$ given by Equation 2.8.*

2.2.5 Complete Dependence and Independence

Two interesting special cases are those of independence and complete dependence. In general, the dependence structure lies between these cases. Let $\mathbf{G}$ be a multivariate extreme value distribution with spectral measure S and $\mathbf{e}_j$ denote the jth unit vector in $\mathbb{R}^d$. Then the margins of $\mathbf{G}$ are independent

$$\mathbf{G}(\mathbf{x}) = \prod_{j=1}^{d} G_j(x_j), \quad \mathbf{x} \in \mathbb{R}^d,$$

if and only if

$$\int_\Xi f(\omega) S(d\omega) = \sum_{j=1}^{d} \|\mathbf{e}_j\|_1 f(\mathbf{e}_j/\|\mathbf{e}_j\|_2),$$

for any real-valued, S-integrable function f on Ξ. This practically means that S consists of $\|\mathbf{e}_j\|_1$ point masses at the points $\mathbf{e}_j/\|\mathbf{e}_j\|_2$. In addition the margins of $\mathbf{G}$ are completely dependent

$$\mathbf{G}(\mathbf{x}) = \bigwedge_{j=1}^{d} G_j(x_j), \quad \mathbf{x} \in \mathbb{R}^d,$$

if and only if S collapses to a single point mass of size $\|\omega_0\|_1/\omega_0$ at the point ω_0, where $\omega_0 = (\omega_0, \ldots, \omega_0)$ is the intersection of Ξ and the line $\{\mathbf{x} \in \mathbb{R}^d : x_1 = \cdots = x_d\}$. This can be equivalently written as

$$\int_{\Xi} f(\omega) S(d\omega) = \|\omega_0\|_1/\omega_0 f(\omega_0),$$

for any real-valued, S-integrable function f on Ξ.

2.2.6 Spectral Measure with Sum-Norm

The most popular choice for the two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ is the sum-norm, $\|\mathbf{x}\|_1 = \sum_{j=1}^{d} |x_j|$. In that case S will be denoted by W later on and the space Ξ is equal to the unit simplex,

$$S_d = \{\omega \in [0, \infty) : \sum_{j=1}^{d} \omega_j = 1\}.$$

The general representation of $\mathbf{G}$ by Equation 2.8 gets simplified as

$$\log \mathbf{G}(\mathbf{x}) = \int_{S_d} \bigwedge_{j=1}^{d} \{\omega_j \log G_j(x_j)\} W(d\omega), \quad \mathbf{x} \in \mathbb{R}^d$$

and the constraints on W from Equation 2.9 reads as

$$\int_{S_d} \omega_j W(d\omega) = 1, \quad j = 1, \ldots, d. \tag{2.10}$$

Moreover, the entire mass of W equals to the number of dimensions

$$W(S_d) = d. \tag{2.11}$$

Remark 6. *In the case of independence W consists of unit point masses at all vertices $\mathbf{e}_1, ..., \mathbf{e}_d$ of the simplex S_d, while for complete dependence a single point mass of size d is taken at the point $(1/d, ..., 1/d)$.*

Remark 7. *In the bivariate case the unit simplex S_2 can be identified with the unit interval $I = [0, 1]$ by identifying $(\omega, 1 - \omega) \in S_2$ with $\omega \in I$. Hence the spectral measure W can also be defined over I as*

$$W([0, \omega]) = \mu_\star\big(\{(y_1, y_2) \in [0, \infty)^2 : y_1 + y_2 \geq 1, y_1/(y_1 + y_2) \leq \omega\big)\}$$

for $\omega \in I$ and the constraints on W are

$$\int_I \omega W(d\omega) = \int_I (1-\omega) W(d\omega) = 1.$$

2.2.7 Spectral Densities

Let W be a spectral measure on the d-dimensional unit simplex S_d. For an absolutely continuous MEVD $\mathbf{G}_\star$ the densities of W can be reconstructed from the derivatives of the function $V_\star = -\log \mathbf{G}_\star$. We use wording as "densities" and not density, since usually W has a density on the interior and on each of lower-dimensional subspaces of S_d as well. The natural partitions of the unit simplex S_d are the so-called faces, with dimensions ranking from 0 (vertices) up to $d-1$ (interior). In particular, these faces are defined for a non-empty subset a of $\{1, ..., d\}$ as

$$S_{d,a} = \{\omega \in S_d : \omega_j > 0 \text{ if } j \in a; \text{ and } \omega_j = 0 \text{ if } j \notin a\},$$

splitting S_d into $2^d - 1$ subsets. Now let us consider the restriction of the spectral measure W to the face $S_{d,a}$. If a is a singleton $\{j\}$, then $S_{d,a}$ is just the vertex $\{\mathbf{e}_j\}$. Let us denote this mass by $h_a = h_a(\mathbf{e}_j)$, to be considered as the density of W at $\mathbf{e}_j$.

Remark 8. *Even if the $\mathbf{G}_\star$ distribution function is absolutely continuous, the spectral measure W may still allocate positive mass to the vertices. E.g. recalling Remark 6 we see that for independent margins $W(\{\mathbf{e}_j\}) = 1$ for all $j = 1, \ldots, d$.*

If a is a subset of $\{1, ..., d\}$ with at least two elements, then the number of free variables in $S_{d,a}$ is $k = |a| - 1$. Hence $S_{d,a}$ can be identified with an open region as

$$\Delta_k = \{\mathbf{t} \in (0, \infty)^k : \sum_{j=1}^{k} t_j < 1\},$$

similarly to Remark 7 for the bivariate case. Let I_a be the map identifying $\mathbf{t}$ in Δ_k with $I_a(\mathbf{t}) = \omega$ in $S_{d,a}$. If $a = \{j_1, \ldots, j_{k+1}\}$ then $\omega_{j_i} = u_i$ for $i = 1, \ldots, k$, $\omega_{j_{k+1}} = 1 - \sum_{i=1}^{k} u_i$ and $\omega_j = 0$ for $j \notin a$. Assuming that the spectral measure W has a density w_a on $S_{d,a}$ integrals over S_d, a can be calculated as

$$\int_{S_{d,a}} f(\omega) W(d\omega) = \int_{\Delta_k} f(I_a(\mathbf{t})) w_a(\mathbf{t}) dt_1 \ldots dt_k.$$

We can obtain the spectral densities by computing partial derivatives of $V_\star$ as follows. If $a = \{j_1, \ldots, j_{k+1}\} \subset \{1, \ldots, d\}$ and $(y_j)_{j \in a}$ such that $0 < z_j < \infty$, then Coles and Tawn (1991) have shown that

$$\lim_{y_j \to 0, j \notin a} \frac{\partial^{k+1} V_\star}{\partial y_{j_1} \ldots \partial y_{j_{k+1}}}(\mathbf{y}) = -\left(\sum_{j \in a} y_j\right)^{-(k+2)} w_a\left(\frac{y_{j_1}}{\sum_{j \in a} y_j}, \ldots, \frac{y_{j_k}}{\sum_{j \in a} y_j}\right). \quad (2.12)$$

There is a very useful tool introduced by Coles and Tawn (1991) which allows to construct families of dependence models defined by their spectral densities.

Theorem 5. *If $\bar{w}$ is an arbitrary density in the interior of S_d with positive first moments*

$$m_j = \int_{S_d} t_j \bar{w}(t_1, \ldots, t_{d-1}) dt_1 \cdots dt_{d-1}, j = 1, \ldots, d,$$

then the measure W on S_d defined by

$$w(t_1, \ldots, t_{d-1}) = \frac{1}{m_0} \prod_{j=1}^{d} \bar{w}\Big(\frac{m_1 t_1}{m_0} \ldots, \frac{m_{d-1} t_{d-1}}{m_0}\Big), \quad (2.13)$$

where $m_0 = \sum_{j=1}^{d} m_j t_j$, is a valid measure satisfying the constrains in Equation 2.10 and Equation 2.11.

2.2.8 Positive Association

The following properties show practical cases, where the MEVD may be useful. An MEVD G is necessarily "positively quadrant dependent" (Sibuya, 1960 and Tiago de Oliveira, 1962/1963), namely

$$G(\mathbf{x}) \geq G_1(x_1) \ldots G_d(x_d), \quad \mathbf{x} \in \mathbb{R}^d. \quad (2.14)$$

In particular, a random variable Y with distribution function $\mathbf{G}$ as in Equation 2.14 has $cov[f_i(Y_i), f_j(Y_j)] \geq 0$ for any $1 \leq i, j \leq d$ and any pair of non-decreasing functions f_i and f_j such that the relevant expectations exist. MEVD satisfies even stronger concepts of the above positive dependence, e.g. Marshall and Olkin (1983) show that they are "associated" in the sense that $cov[\xi(Y), \eta(Y)] \geq 0$ for every pair of non-decreasing functions ξ and η on $\mathbb{R}^d$ for which the relevant expectations exist.

2.2.9 Pickands' Dependence Function

In two dimensions, all information on the dependence structure is covered by another (equivalent) characterization, due to Pickands (1981). The joint survivor function with standard exponential margins (denoted by $\star\star$ instead of $\star$ which is kept for unit Fréchet margins) $\bar{\mathbf{G}}_{\star\star}$ is given by

$$\bar{\mathbf{G}}_{\star\star}(z_1, z_2) = P(Z_1 > z_1, Z_2 > z_2) = \exp\left\{-(z_1 + z_2)A\left(\frac{z_2}{z_1 + z_2}\right)\right\}, \tag{2.15}$$

where $A(t)$, called the (Pickands) dependence function, is responsible to capture the dependence structure between the margins. It can be shown that the dependence function necessarily satisfies the following properties (P):

1. $(1-t) \vee t \leq A(t) \leq 1$ for $t \in [0,1]$ $(\Rightarrow A(0) = A(1) = 1)$;
2. $A(t)$ is convex.

Remark 9. *In the first property of (P) the lower and upper bounds correspond to the two limiting cases of subsection 2.2.5. If $A(t) = (1-t) \vee t$ then we get complete dependence and if $A(t) = 1$ then independence. For graphical illustration see the left panel in Figure 2.2.*

Of course, the representation using exponent measure function and unit Fréchet margins can be written by the dependence function as well:

$$-\log G_\star(y_1, y_2) = V_\star(y_1, y_2) = \left(\frac{1}{y_1} + \frac{1}{y_2}\right) A\left(\frac{y_1}{y_1 + y_2}\right). \tag{2.16}$$

Furthermore, there is a connection between the Pickands dependence function and the spectral measure W of $\mu_\star$ with respect to the sum-norm as in subsection 2.2.6

$$A(t) = 1 - t + \int_0^t W([0,\omega])d\omega, \quad t \in [0,1]$$

Conversely W can be computed from A as

$$W([0,\omega]) = 1 + A'(\omega) \text{ if } \omega \in [0,1),$$

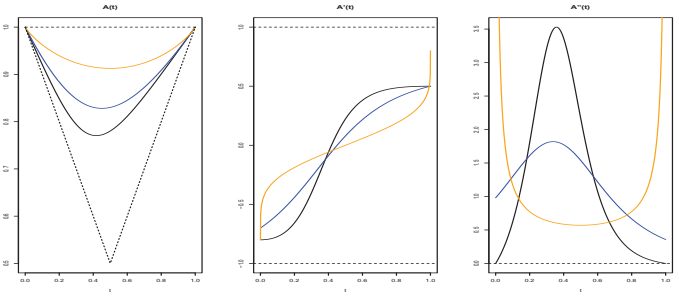

Figure 2.2: Differentiable dependence functions, and their derivatives.

and $W([0,1]) = 2$, where A' is the derivative of A. The point masses in the endpoints are

$$W(\{0\}) = 1 + A'(0) \text{ and } W(\{1\}) = 1 - A'(1). \tag{2.17}$$

Remark 10. *If A' is absolutely continuous, then W is absolutely continuous on the interior of the unit interval with density $w = A''$.*

For higher dimensions Equation 2.15 could be generalized as

$$\bar{G}_{\star\star}(\mathbf{z}) = \exp\left\{-\left(\sum_{i=1}^{d} z_i\right) A\left(\frac{z_1}{\sum_{i=1}^{d} z_i}, ..., \frac{z_{d-1}}{\sum_{i=1}^{d} z_i}\right)\right\},$$

for some dependence function A, defined on the d-dimensional simplex. Further characteristic properties of the dependence function can be found in Falk and Reiss (2005).

2.2.10 Density Function of BEVD Models

Let (X_1, X_2) denote a bivariate random vector representing the component-wise maxima of an i.i.d. sequence over a given period of time. Due to subsection 2.2.1, under the appropriate conditions the distribution of (X_1, X_2) can be approximated by a bivariate extreme-value distribution (BEVD) with cdf G. The

BEVD is determined by its margins G_1 and G_2 respectively, which are necessarily GEV, and by its Pickands dependence function A as

$$G(x_1, x_2) = \exp\left\{ \log\big(G_1(x_1)G_2(x_2)\big) A\left(\frac{\log\big(G_2(x_2)\big)}{\log\big(G_1(x_1)G_2(x_2)\big)} \right) \right\}. \tag{2.18}$$

Alternatively, for standardized margins it can be also given as in Equation 2.15 or in Equation 2.16. Due to the probability integral transform $U_i = G_i(X_i)$, $i = 1, 2$ we can obtain uniformly distributed variables on the unit interval, which can easily be further transformed to any desired distribution. The benefit of the simpler forms (beyond the handiness in computing) will be clear later when we derive the nonparametric estimator of $A(t)$. To obtain the density of the BEVD one needs the first and second derivatives of the dependence function denoted by $A'(\cdot)$ and $A''(\cdot)$ respectively. A useful formula for the BEVD density is presented in (8) of Hall and Tajvidi (2004):

Theorem 6. *The BEVD density g (on the original scale) can be expressed by the dependence function A as*

$$\begin{aligned} g(x_1, x_2) &= \frac{\partial^2}{\partial x_1 \partial x_2} G(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} \bar{G}_*(t_1, t_2) = \\ \bar{G}_*(t_1, t_2) t_1' t_2' &\times \left(\Big(A(\zeta) + (1-\zeta)A'(\zeta) \Big) \Big(A(\zeta) - \zeta A'(\zeta) \Big) + \eta A''(\zeta) \right), \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} t_i &= t_i^{\star\star}(x_i) = -\log G_i(x_i) = \left(1 + \xi_i \frac{x_i - \mu_i}{\sigma_i}\right)^{-\frac{1}{\xi_i}}, \; i = 1, 2 \\ t_i' &= t_i^{\star\star\prime}(x_i) = -\frac{1}{\sigma_i} \left(1 + \xi_i \frac{x_i - \mu_i}{\sigma_i}\right)^{-\frac{1}{\xi_i} - 1}, \; i = 1, 2 \\ \zeta &= \frac{t_2}{t_1 + t_2} \\ \eta &= \frac{t_1 t_2}{(t_1 + t_2)^3}. \end{aligned}$$

In order to ensure that the BEVD is absolutely continuous, we must assume that $A(t)$ is two times differentiable, which is not included in the necessary requirements for a dependence function (P). Although the well-known parametric models satisfy this extra property as well, it turns out to be a more problematic issue when using nonparametric estimates for the dependence function $A(t)$.

2.3 Modeling Multivariate Threshold Exceedances

As taking component-wise maxima can hide the time structure within the given period, we do not know if the different components of the maxima occurred simultaneously (in the same day for instance) or not. To avoid this problem, instead of considering only the highest value of a time period, we investigate all observations exceeding a given high threshold. Since this method usually uses more data (depending on the threshold level) it usually leads to more accurate estimation than the other method using only the maxima. In this case the first fundamental question is which observations we should consider as exceedances. In a classical framework, only those observations are involved in modeling that exceed the threshold in all components, simultaneously. The distribution function of these exceedances can be written analogously to MEVD, replacing the GEV margins by GPD margins. The methods that uses this definition are widely investigated in the literature, see Smith (1994) for one of its first applications or the PhD thesis of Michel (2006) for a review. However, this definition can be very restrictive in some cases, as it still ignores many potentially important observations: those with only some components beyond the threshold.

2.3.1 Limit for Multivariate Threshold Exceedances

Here we give the following definition by Rootzén and Tajvidi (2006) which follows the latter idea.

Definition 5. *A distribution function H is a multivariate generalized Pareto distribution if*

$$H(\mathbf{x}) = \frac{-1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})} \tag{2.20}$$

for some MEVD G with nondegenerate margins and with $0 < G(\mathbf{0}) < 1$. In particular, $H(\mathbf{x}) = 0$ for $\mathbf{x} < \mathbf{0}$ and $H(\mathbf{x}) = 1 - \log G(\mathbf{x}) / \log G(\mathbf{0})$ for $\mathbf{x} > \mathbf{0}$.

Remark 11. *Note, that this definition does not imply that the lower dimensional margins of an MGPD are GPDs. However, if Z is distributed as H then the con-*

ditional distribution of $Z_i|Z_i > 0$ is GPD. This property holds for all marginal distributions in any dimensions less than d.

Let $\mathbf{X}$ be a d-dimensional random vector with distribution function F, $\{\mathbf{u}(t) : t \in [1, \infty)\}$ a d-dimensional curve starting at $\mathbf{u}(1) = \mathbf{0}$ and $\sigma(\mathbf{u}) = \sigma(\mathbf{u}(t)) > 0$ be a function with values in $\mathbb{R}^d$. Then the normalized exceedances at level $\mathbf{u}$ can be defined as

$$\mathbf{X}_\mathbf{u} = \frac{\mathbf{X} - \mathbf{u}}{\sigma(\mathbf{u})}.$$

Theorem 7 (Rootzén and Tajvidi, 2006)**.** *(i) Suppose, that G is a d-dimensional MEVD with $0 < G(\mathbf{0}) < 1$. If $F \in D(G)$ then there exist an increasing continuous curve $\mathbf{u}$ with $F(\mathbf{u}(t)) \to 1$ as $t \to \infty$, and a function $\sigma(\mathbf{u}) > \mathbf{0}$ such that*

$$P(\mathbf{X}_\mathbf{u} \leq \mathbf{x} | \mathbf{X}_\mathbf{u} \not\leq \mathbf{0}) \to \frac{-1}{\log G(\mathbf{0})} \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}$$

as $t \to \infty$, for all $\mathbf{x}$.

(ii) Suppose there exists an increasing continuous curve $\mathbf{u}$ with $F(\mathbf{u}(t)) \to 1$ as $t \to \infty$, and a function $\sigma(\mathbf{u}) > 0$ such that

$$P(\mathbf{X}_\mathbf{u} \leq \mathbf{x} | \mathbf{X}_\mathbf{u} \not\leq \mathbf{0}) \to H(\mathbf{x}), \tag{2.21}$$

for some function H, as $t \to \infty$, for $\mathbf{x} > \mathbf{0}$, where the marginals of H on $\mathbb{R}^+$ are nondegenerate. Then the left-hand side of Equation 2.21 converges to a limit $H(\mathbf{x})$ for all $\mathbf{x}$ and there is a unique MEVD G with $G(\mathbf{0}) = \exp(-1)$ such that

$$H(\mathbf{x}) = \log \frac{G(\mathbf{x})}{G(\mathbf{x} \wedge \mathbf{0})}.$$

This G satisfies $G(\mathbf{x}) = \exp(-\bar{H}(\mathbf{x}))$ for $\mathbf{x} > 0$, and $F \in D(G)$.

Theorem 8 (Rootzén and Tajvidi, 2006)**.** *(i) Suppose $\mathbf{X}$ has MGPD. Then there exists an increasing continuous curve $\mathbf{u}$ with $P(\mathbf{X} \leq \mathbf{u}(t)) \to 1$ as $t \to \infty$, and a function $\sigma(\mathbf{u}) > 0$ such that*

$$P(\mathbf{X}_\mathbf{u} \leq \mathbf{x} | \mathbf{X} \not\leq \mathbf{0}) = P(\mathbf{X} \leq \mathbf{x}), \tag{2.22}$$

for $t \in [1, \infty)$ and all x.

(ii) If there exists an increasing continuous curve u with $P(\mathbf{X} \leq \mathbf{u}(t)) \to 1$ as

$t \to \infty$, and a function $\sigma(\mathbf{u}) > 0$ such that Equation 2.22 holds for $\mathbf{x} > \mathbf{0}$ and $\mathbf{X}$ has nondegenerate margins, then $\mathbf{X}$ has MGPD.

Remark 12. *We can see that similarly to the univariate case, if the limit distribution for the normalized maxima converges to MEVD then the normalized exceedances converges to MGPD.*

2.3.2 An Invariance Property of MGDP

Focusing first on the bivariate case we can index G in Equation 2.20 by $\mu_i, \sigma_i, \gamma_i$ marginal parameters for $i = 1, 2$ and some function representing the dependence structure (e.g. exponent measure function V as below). For $x_1 > 0, x_2 > 0$ the BGPD H can be written as

$$H(x_1, x_2) = 1 - \frac{\log G(x_1, x_2)}{\log G(0,0)} = 1 - \frac{V(x_1, x_2)}{V(0,0)} = 1 - \frac{V_\star(\tau_1(x_1), \tau_2(x_2))}{V_\star(\tau_1(0), \tau_2(0))}, \quad (2.23)$$

where

$$\tau_i(x_i) = \frac{-1}{\log G_i(x_i)} = \left(1 + \gamma_i \frac{x_i - \mu_i}{\sigma_i}\right)^{-\frac{1}{\gamma_i}}, \ i = 1, 2,$$

and V and $V_\star$ are the exponent measure functions on the original scale and the unit Fréchet scale, respectively.

Theorem 9 (Rakonczai and Turkman, 2012)**.** *For any (X_1, X_2) bivariate generalized Pareto random vector there is a continuous, increasing curve ϱ starting from point $(0,0)$ for which the value of distribution function $H(x_1, x_2)$ at $(x_1, x_2) \in \varrho$ is invariant under changing the underlying dependence structure.*

Proof: For simplicity we use the following notations:

$$\tau_1 = \tau_1(x_1), \ \tau_2 = \tau_2(x_1), \ c_{\tau_1} = \tau_1(0), \ c_{\tau_2} = \tau_2(0)$$

and take a closer look on the fraction on the right side of Equation 2.23. It can also be written in terms of the dependence function as

$$H(x_1, x_2) = 1 - \frac{\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right) A\left(\frac{\tau_1}{\tau_1 + \tau_2}\right)}{\left(\frac{1}{c_{\tau_1}} + \frac{1}{c_{\tau_2}}\right) A\left(\frac{c_{\tau_1}}{c_{\tau_1} + c_{\tau_2}}\right)}. \quad (2.24)$$

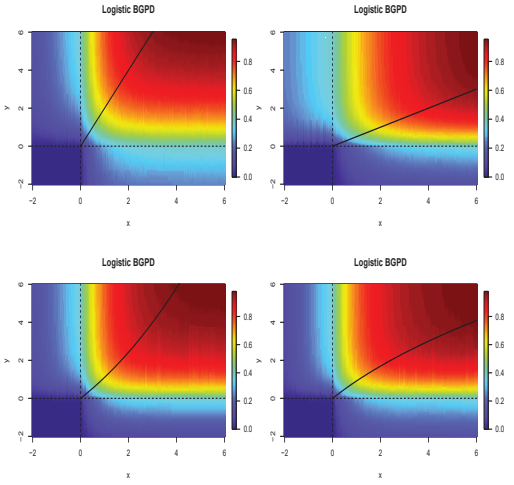

Figure 2.3: Distribution functions of symmetric logistic BGPD (introduced later in subsection 3.1.1) with different marginal parameters (dependence parameter $\alpha = 1.5$). On the black curve the distribution functions are invariant under changing the dependence structure.

Obviously, if the arguments on the dependence functions are equal, then we can simplify by them and the rest will not be dependent on A any more. This case occurs if $\tau_2 = \frac{c_{\tau_2}}{c_{\tau_1}}\tau_1$ defining a linear function on the transformed scale. In general, for arbitrary marginal parameters

$$x_2 = \tau_2^{-1}\left(\frac{c_{\tau_2}}{c_{\tau_1}}\tau_1(x_1)\right), \tag{2.25}$$

which is a increasing function in x_1 as stated in the theorem. □

See Figure 2.3 for illustration.

Corollary 1. *The curve ϱ defined in Equation 2.25 is linear if the shape parameters of the marginal distributions are equal. Specifically, it is identity if $\tau_1(\cdot) = \tau_2(\cdot)$,*

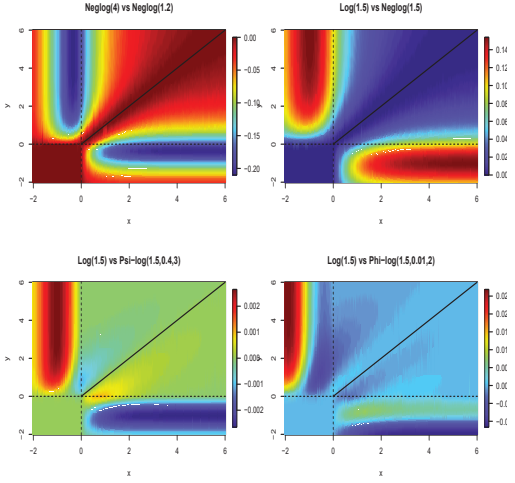

Figure 2.4: Differences between some parametric BGPD distribution functions (introduced later in subsection 3.1.1 and subsection 3.1.2) are zero on the diagonal for identical location, scale and shape parameters. The dependence structures are different, whereas the marginal parameters are $\mu = 0$, $\sigma = 1$ and $\gamma = 0.2$ in any case.

namely

$$\mu_1 = \mu_2 \, , \sigma_1 = \sigma_2, \; \gamma_1 = \gamma_2.$$

Corollary 2. *Let (X_1, X_2) and (X_1', X_2') be bivariate generalized Pareto random vectors with distribution functions H and H', respectively. Let us denote the curve function in Equation 2.25 as $f(\cdot) = \tau_2^{-1}(c_{\tau_2}/c_{\tau_1}\tau_1(\cdot))$. If the marginal parameters $\mu_i, \sigma_i, \gamma_i$ for $i = 1, 2$ are identical for H and H' then*

$$H(x, f(x)) = H'(x, f(x)).$$

See Figure 2.4 for illustration.

Exactly the same considerations can be applied for $d > 2$ dimensions. If $H(x_1, \ldots, x_d)$ is a d-dimensional MGPD then Equation 2.24 for $x_1 > 0, \ldots, x_d > 0$ modifies as

$$H(x_1, x_2, \ldots, x_d) = 1 - \frac{\sum_{i=1}^{d} \frac{1}{\tau_i} A\left(\frac{\tau_1}{\sum_{i=1}^{d}\tau_i}, \ldots, \frac{\tau_{d-1}}{\sum_{i=1}^{d}\tau_i}\right)}{\sum_{i=1}^{d} \frac{1}{c_{\tau_i}} A\left(\frac{c_{\tau_1}}{\sum_{i=1}^{d} c_{\tau_i}}, \ldots, \frac{c_{\tau_{d-1}}}{\sum_{i=1}^{d} c_{\tau_i}}\right)}.$$

For further simplicity let us denote $c_j = \frac{c_{\tau_j}}{\sum_{i=1}^{d} c_{\tau_i}}$, for $j = 1, \ldots, d-1$. We can simplify by A if its arguments appearing in the nominator and denominator are equal componentwise. This happens if the following linear equation system is fulfilled

$$\begin{aligned} \tau_1 &= c_1 \sum_{i=1}^{d} \tau_i \\ \tau_2 &= c_2 \sum_{i=1}^{d} \tau_i \\ &\vdots \\ \tau_{d-1} &= c_{d-1} \sum_{i=1}^{d} \tau_i. \end{aligned} \tag{2.26}$$

The solution of Equation 2.26 is a straight line in the d-dimensional space as

$$\tau_1 = \frac{c_{\tau_1}}{c_{\tau_2}} \tau_2 = \cdots = \frac{c_{\tau_1}}{c_{\tau_d}} \tau_d. \tag{2.27}$$

The coordinates of the curve ϱ on the original space can be then computed recursively by solving

$$x_j = \tau_j^{-1}\left(\frac{c_{\tau_j}}{c_{\tau_1}} \tau_{j-1}(x_{j-1})\right)$$

for $j = 2, \ldots, d$.

Corollary 3. *If $\tau_1(\cdot) = \cdots = \tau_d(\cdot)$ then $c_j = 1/d$ for $j = 1, \ldots, d-1$, and Equation 2.27 leads to a diagonal line on the original scale. Hence for any pairs of MGPD distribution functions H and H' with the above property $H(x, \ldots, x) = H'(x, \ldots, x)$ for $x > 0$.*

2.3.3 Conditions for an MGPD to be absolutely continuous

For practical applications it is reasonable to require that in the MGPD model there is no positive probability mass on the boundary of the distribution, because otherwise the model would not remain absolutely continuous. A model, being not absolutely continuous, is hardly realistic and causes further complications for maximum likelihood estimation. The following statement (Rakonczai and Zempléni, 2012) can be used to construct an absolutely continuous MGPD model from a known absolutely continuous MEVD:

Theorem 10. *Let H be a MGPD represented by an absolutely continuous MEVD G with spectral measure W. H is absolutely continuous $\Leftrightarrow W(int(\mathcal{S}_d)) = d$ holds, i.e. all mass is put on the interior of the simplex.*

Corollary 4. *In the bivariate case, the above statements are also equivalent with $W(\{0\}) = W(\{1\}) = 0$ and so, due to (Equation 2.17), with*

$$-A'(0) = A'(1) = 1, \tag{2.28}$$

where A is the Pickands dependence function.

Proof: Here we limit ourself to a special bivariate case, but for higher dimensions analogous calculations can be made. Suppose that $(X_1, X_2) \sim H$ is a BGPD with lower endpoints l_{x_1}, l_{y_2} for its margins. The marginal transformations mapping X_1 and X_2 into the unit Fréchet scale are $t_i = -1/\log G_i(x_i)$ for $i = 1, 2$ as in subsection 2.2.3. Let the value of the first component of (X_1, X_2) converge to l_{x_1}

assuming fixed $x_2 > 0$ value for the second one, then

$$\begin{aligned}
\lim_{x_1 \to l_{x_1}} H(x_1, x_2) &= \lim_{x_1 \to l_{x_1}} \frac{\log G(x, y) - \log G(x \wedge 0, y \wedge 0)}{G(0,0)} = \\
&= \lim_{x_1 \to l_{x_1}} \frac{\log G_\star(t_1, t_2) - \log G_\star(t_1, t_{2,0})}{G_\star(t_{1,0}, t_{2,0})} = \\
&= \underbrace{\frac{1}{G_\star(t_{1,0}, t_{2,0})}}_{=c_0} \times \lim_{x_1 \to l_{x_1}} [V_\star(t_1, t_2) - V_\star(t_1, t_{2,0})] = \\
&= \frac{1}{c_0} \times \lim_{x_1 \to l_{x_1}} \left\{ \int_{S_2} \max\left(\frac{w}{t_1}, \frac{1-w}{t_2}\right) W(d\mathbf{w}) \right. \\
&- \left. \int_{S_2} \max\left(\frac{w}{t_1}, \frac{1-w}{t_{2,0}}\right) W(d\mathbf{w}) \right\} = \\
&= \frac{1}{c_0} \times \left\{ \frac{1}{t_2} W(\{0,1\}) + \lim_{x_1 \to l_{x_1}} \int_{S_2 \setminus \{0,1\}} \frac{w}{t_1} W(d\mathbf{w}) - \right. \\
&- \left. \frac{1}{t_{2,0}} W(\{0,1\}) - \lim_{x_1 \to l_{x_1}} \int_{S_2 \setminus \{0,1\}} \frac{w}{t_1} W(d\mathbf{w}) \right\} = \\
&= \underbrace{\frac{1}{c_0}}_{\neq 0} \times \underbrace{\left\{ \frac{1}{t_2} - \frac{1}{t_{2,0}} \right\}}_{\neq 0} \times W(\{0,1\}) = \\
&= 0 \Leftrightarrow W(\{0,1\}) = 0. \ \square
\end{aligned} \tag{2.29}$$

Remark 13. *By Remark 6 the class of absolutely continuous MGPD models does not include the case of independent margins. This case is discussed in detail in the Section 3 of Rootzén and Tajvidi (2006).*

2.3.4 Density Function of BGPD Models

Following the notation of subsection 2.3.1 let (Y_1, Y_2) be the observed random variable, (u_1, u_2) a given threshold vector and $(X_1, X_2) = (Y_1 - u_1, Y_2 - u_2)$ the random vector of exceedances. Due to Theorem 7 we can define the bivariate generalized Pareto distribution (later BGPD) for the exceedances by its distribution function

$$H(x_1, x_2) = \frac{-1}{\log G(0,0)} \log \frac{G(x_1, x_2)}{G(x_1 \wedge 0, x_2 \wedge 0)}, \tag{2.30}$$

for some BEVD G with nondegenerate margins and with $0 < G(0,0) < 1$. This is the bivariate version of (Equation 2.20). So practically the probability measure is *positive* in the upper three quarter planes and *zero* in the bottom left one. The main message of this definition is that the BGPD distribution models those observations too, which are extremes merely in one component. Similar formula to Equation 2.19 using dependence function A for the BGPD density h can be obtained as in (10) of Rakonczai and Tajvidi (2010):

Theorem 11. *The BGPD density h can be written by the dependence function A as follows*

$$h(x_1, x_2) = \frac{t_1' t_2'}{c_0} \times \eta A''(\zeta), \tag{2.31}$$

where $t_1, t_2, t_1', t_2', \zeta, \eta$ are the same as previously in Theorem 6 and

$$c_0 = -(t_1^{\star\star}(0) + t_2^{\star\star}(0))A\left(\frac{t_2^{\star\star}(0)}{t_1^{\star\star}(0) + t_2^{\star\star}(0)}\right).$$

Proof: It is easy to see, that for the regions where $x > 0, y < 0$ and $x < 0, y > 0$ the second derivatives are the same as for $x > 0, y > 0$. Viz. in the mentioned regions $\frac{\partial^2 G(x\wedge 0, y\wedge 0)}{\partial x \partial y} = 0$. Taking this into account we see that the density is

$$h(x,y) = \frac{\partial^2 H(x,y)}{\partial x \partial y} = \frac{\partial^2}{\partial x \partial y}\Big(1 - \frac{\log G(x,y)}{\log G(0,0)}\Big),$$

where considering the usual marginal standardization $\log G(x,y)$ can be written as

$$\begin{aligned} \log G(x,y) &= \log G_{\star\star}(-\log G_1(x), -\log G_2(y)) = \\ \log G_{\star\star}(t_1,t_2) &= -(t_1+t_2)A\left(\frac{t_2}{t_1+t_2}\right). \end{aligned}$$

Then the second mixed partial derivative of the above form look as

$$\begin{aligned} \frac{\partial^2}{\partial t_1 \partial t_2}\left((t_1+t_2)A\Big(\frac{t_2}{t_1+t_2}\Big)\right) &= \frac{\partial}{\partial t_2}\left(A(\zeta) + (t_1+t_2)A'(\zeta)\frac{-t_2}{(t_1+t_2)^2}\right) = \\ \frac{\partial}{\partial t_2}(A(\zeta) - \zeta A'(\zeta)) &= A'(\zeta)\zeta_{t_2}' - \zeta_{t_2}' A'(\zeta) - \zeta A''(\zeta)\zeta_{t_2}' = \\ \frac{-t_1 t_2}{(t_1+t_2)^3} A''(\zeta). \square \end{aligned}$$

Illustration for some BGPD density functions can be found in Figure 2.5 and Figure 2.6. In these figures there is the logistic dependence model chosen as baseline and both previously proposed examples of $\Psi-$ and $\Phi-$models are applied. A simple formula for the density of the d-variate MGPD, applying the exponent measure, can be found in (7) of Rakonczai and Zempléni (2012).

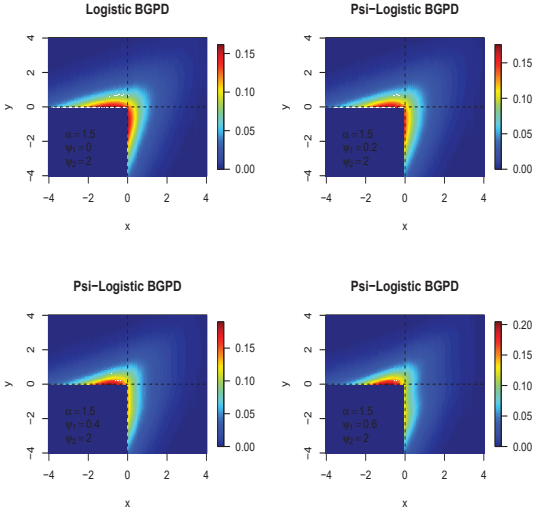

Figure 2.5: Ψ-logistic BGPD density functions, see subsection 3.1.2 for description and Figure 3.5 for the applied dependence models.

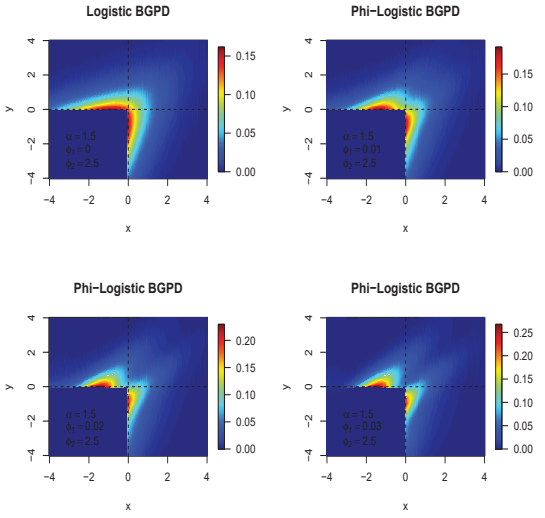

Figure 2.6: Φ-logistic BGPD density functions, see see subsection 3.1.2 for description and Figure 3.7 for the applied dependence models.

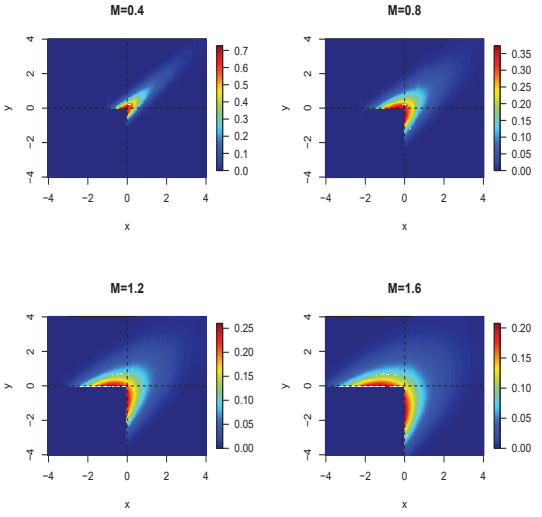

Figure 2.7: BGPD density functions based on Smith dependence models, see Figure 2.8 for the applied dependence models.

2.3.5 Nonstationary BGPD models

What happens if we intend to use the BGPD model for probabilistic statements about the future characteristics of extreme events? The stationary BGPD model may produce biased predictions for the values of future extremes, if the underlying process actually contains some form of systematic trend over time. This may occur even when the fitted BGPD model appears to have a good fit to the observed data.

This motivates a generalization of the stationary BGPD model in which the parameters of the BGPD density (in Theorem 11) are assumed to be time dependent (see Rakonczai et al., 2010). Extensions of univariate extreme value models to allow the values of the parameters to depend upon time or other covariates are well developed (e.g. Yee & Stephenson, 2007), and bivariate models can be extended in an analogous way. There are good arguments for filtering out possible trends and periodicity from the bulk of the data before fitting a model to the extremes (Eastoe and Tawn, 2009) , but the extremes of the filtered series may still be nonstationary so the models to be introduced are definitely interesting for practitioners. It is also possible to directly filter out trends in the extremes, but the outputs obtained from fitting an extreme value model to the filtered data - e.g. prediction regions, which we will introduce in the next section - can be difficult to transform back onto the original scale.

By considering the marginal and the dependence parameters to be functions of time, rather than constant, more flexibility can be added into the model. Here we focus on the symmetric logistic dependence structure and linear trends over time for the parameters of the BGPD model. Within this context the density of the stationary BGPD model (Equation 2.31) is modified by transforming the parameters to be

$$\begin{aligned} \mu_i &= \mu_i(t) = a_{\mu_i} t + b_{\mu_i} \\ \sigma_i &= \sigma_i(t) = a_{\sigma_i} t + b_{\sigma_i} \\ \xi_i &= \xi_i(t) = a_{\xi_i} t + b_{\xi_i} \\ \alpha &= \alpha(t) = a_\alpha t + b_\alpha, \end{aligned} \tag{2.32}$$

where $i = 1, 2$ and t denotes time. We refer to this as the nonstationary BGPD model. Each parameter of the stationary BGPD model is replaced by a pair of

parameters: an intercept, and a linear slope (trend) over time. Of course, these new parameters are subject to constraints, because the usual conditions that $\sigma_i(t) > 0$, $i = 1, 2$ and $\alpha(t) > 1$ must remain fulfilled across the entire range of observed times t (and possibly beyond this range). The slope parameters (later: trend) monitor the rates of change of the individual BGPD parameters over time, so the existence of a statistically significant slope parameter indicates the presence of time-dependence within the BGPD model.

Remark 14. *If the assumption of linearity is inappropriate for a given application, then the approach can be extended to include polynomial, exponential, logarithmic or even periodic functions. It should also be noted that explanatory variables other than time can be incorporated into the analysis in an analogous way.*

2.4 Infinite Dimensional Generalization of EVT

This section describes the latest area of extreme value theory, which provides further possible field of research and applications. Some parametric subclasses of max-stable processes lead to new dependence models, which could be considered (if satisfying Theorem 10) in MGPD models, see Theorem 12 and Theorem 13.

2.4.1 Max-stable Processes

Max-stable processes (de Haan, 1984 and Vatan, 1985) can be thought of an infinite-dimensional generalization of extreme value theory. Due to Smith (1990) this idea is extended as a universal approach to modeling extremes of processes with spatial dependence. In spatial applications there are observations collected on the points of a space grid and the joint distribution of extreme values at different places is of interest. Of course, for a finite grid we can consider this problem as another application of multivariate extreme value theory as in section 2.2, but usually this leads to untractable explosion in model parameters and hence to further problems in estimation and prediction.

Definition 6 (de Haan, 1984)**.** *Let T be an index set and $\{Y_i(t)\}_{t\in T}$, $i = 1, \ldots, n$ be n independent replications of a continuous stochastic process. Assume that there are sequences of continuous functions $a_n(t) \in \mathbb{R}$ and $b_n(t) > 0$ such that*

$$Z(t) = \lim_{n\to\infty} \frac{\bigvee_{i=1}^{n} Y_i(t) - a_n(t)}{b_n(t)} \text{ for any } t \in T.$$

If this limit exists, the limit process $Z(t)$ is a max-stable process.

The univariate margins are GEV distributions as in Equation 2.2 and for any $d = 2, 3, \ldots$ the d-dimensional marginal distribution are MEVD. Specially, if $a_n(t) = 0$ and $b_n(t) = n$, then the corresponding process, $\{Z_\star(t)\}_{t\in T}$, has unit Fréchet margins. Similarly to subsection 2.2.3 the standardized process is available by

$$\{Z_\star(t)\}_{t\in T} = \left\{ \left[1 + \frac{\xi(t)(Y(t) - \mu(t))}{\sigma(t)} \right]_+^{1/\xi(t)} \right\}_{t\in T},$$

where $\mu(t)$, $\sigma(t) > 0$ and $\sigma(t)$ are continuous functions. The standardized process $Z_\star$ is still a max-stable process. Moreover, if $Z_\star$ is stationary as well, it can be expressed by the spectral representation of de Haan and Pickands (1986).Let $\{\pi_j\}_{j=1}^{\infty}$ be a Poisson process on $\mathbb{R}_+$ with intensity ds/s^2 and $\{\Lambda_j\}_{j=1}^{\infty}$ be independent copies of a stationary process $\Lambda(x)$ on $\mathbb{R}^d$ satisfying

$$E\{0 \vee \Lambda_j(0)\} = 0, \tag{2.33}$$

where 0 denotes the origin. Then

$$Z_\star = \bigvee_{j=1}^{\infty} \{\pi_j \times \{0 \vee \Lambda_j(x)\}\} \tag{2.34}$$

is a stationary max-stable process on $\mathbb{R}^d$ with unit Fréchet margins. (For more details see Smith, 1990.) Different choices for the process $\Lambda(x)$ lead to some useful max-stable models. Due to stationarity the joint distribution of $Z_\star(x)$ is well described by the joint distribution of $Z_\star(0)$ and $Z_\star(h)$, where 0 and h are the two sites. Here we recall the two most well-known models, more detailed model summary can be found in Davison et al. (2011).

Smith model

One possibility is to take

$$\Lambda_j(x) = r(x - C_j),$$

where r is a probability density function and $\{C_j\}$ is a homogeneous Poisson process, both on $\mathbb{R}^d$. In this case the value of the max-stable process at x may be interpreted as the maximum over a infinite number of storms, centered at random point C_j and of strength π_j. The case, where r is multivariate normal density was considered by Smith (1990). The exponent measure function for $Z_\star(0)$ and $Z_\star(h)$

$$\begin{aligned} V_\star(z_1, z_2) &= \frac{1}{z_1}\Phi\left(\frac{m(h)}{2} + \frac{1}{m(h)}\log\left(\frac{z_2}{z_1}\right)\right) \\ &+ \frac{1}{z_2}\Phi\left(\frac{m(h)}{2} + \frac{1}{m(h)}\log\left(\frac{z_1}{z_2}\right)\right), \end{aligned} \tag{2.35}$$

where $m^2(h) = m^T\Omega^{-1}m$ is the so-called Mahalanobis distance between h and the origin, and Φ is the standard normal distribution function. This is an exchangeable model, having $M = m(h) > 0$ as dependence parameter. Independence or complete dependence between the margins corresponds to the limit $M \to 0^+$ or $M \to \infty$, respectively.

Theorem 12 (Rakonczai and Turkman, 2012)**.** *The BGPD model defined by the Smith model (having exponent measure as in Equation 2.35) is absolutely continuous.*

Proof: In order to prove the above statement it is useful to switch to the dependence function representation.

$$\begin{aligned} V_\star(z_1, z_2) &= \left(\frac{1}{z_1} + \frac{1}{z_2}\right) \times \left(\frac{z_2}{z_1 + z_2}\Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{\frac{z_2}{z_1+z_2}}{\frac{z_1}{z_1+z_2}}\right)\right)\right. \\ &+ \left.\frac{z_1}{z_1 + z_2}\Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{\frac{z_1}{z_1+z_2}}{\frac{z_2}{z_1+z_2}}\right)\right)\right) \\ &= \left(\frac{1}{z_1} + \frac{1}{z_2}\right)A\left(\frac{z_1}{z_1 + z_2}\right), \end{aligned}$$

where the dependence function is

$$\begin{aligned} A(t) &= (1-t) \times \Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{1-t}{t}\right)\right) \\ &+ t \times \Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{t}{1-t}\right)\right), \end{aligned} \tag{2.36}$$

For the BGPD density, we need to check the conditions in Equation 3.2, namely the left and right limit of the first order derivative of $A(t)$ in Equation 2.36 should be computed at 0 and 1 respectively. Straightforward calculations shows that the limits of $A'(t)$

$$\begin{aligned} A'(t) &= -\Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{1-t}{t}\right)\right) \\ &+ (1-t) \times \Phi'\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{t}{1-t}\right)\right) \times \frac{1}{M(1-t)t} \\ &+ \Phi\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{1-t}{t}\right)\right) \\ &+ t \times \Phi'\left(\frac{M}{2} + \frac{1}{M}\log\left(\frac{t}{1-t}\right)\right) \times \frac{-1}{M(1-t)t} \end{aligned}$$

are $\lim_{t\to 0^+} A'(t) = -1 + 0 + 0 + 0 = -1$ and $\lim_{t\to 1^-} A'(t) = 0 + 0 + 1 + 0 = 1$, hence Equation 3.2 ensures that BGPD constructed by this dependence model has a density function. □

For graphical illustration of $A(t)$ and its derivatives see Figure 2.8. (Some other plots of BGPD density functions based on Smith dependence function are presented later, in Figure 2.7.)

Remark 15. *The Brown-Resnick (Section 6 in Davison et al., 2011) model leads to the same bivariate dependence model as in Equation 2.36.*

Schlather model

Another possibility is to take $\{\Lambda_j\}$ to be a stationary standard Gaussian process with correlation function $\varrho(h)$, scaled so that Equation 2.33 remain valid. The

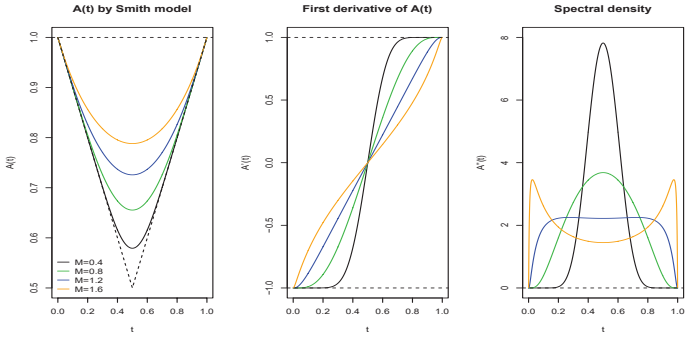

Figure 2.8: Dependence functions of the Smith model with different parameters and the derivatives verifying Equation 3.2.

exponent measure function for $Z_\star(0)$ and $Z_\star(h)$ is

$$V_\star(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1}+\frac{1}{z_2}\right)\left(1+\left[1-2(\varrho(h)+1)\frac{z_1 z_2}{(z_1+z_2)^2}\right]^{1/2}\right). \quad (2.37)$$

This model is appealing as it allows the use of the rich variety of correlation functions. Unfortunately the positive definiteness of $\varrho(h)$ imposes constraints on the level of asymptotic dependence (e.g. in some cases asymptotic independence can not be approached).

Theorem 13 (Rakonczai and Turkman, 2012)**.** *The BGPD model defined by the Schlather model (having exponent measure as in Equation 2.37) is not absolutely continuous.*

Proof: In order to prove the above statement it is useful to switch to the dependence function representation. Let us define the dependence parameter $M = 2(\varrho(h)+1) \in (0,4)$. Hence, independence or complete dependence between the margins corresponds to the limit $M \to 0^+$ or $M \to 4^-$, respectively. (M could be defined

at the endpoints 0 and 4 as well, but in these cases the BGPD is degenerate anyway.)

$$\begin{aligned} V_\star(z_1, z_2) &= \left(\frac{1}{z_1} + \frac{1}{z_2}\right) \times \left(\frac{1}{2} + \frac{1}{2}\left[1 - M\frac{z_1}{z_1 + z_2}\frac{z_2}{z_1 + z_2}\right]^{1/2}\right) \\ &= \left(\frac{1}{z_1} + \frac{1}{z_2}\right) A\left(\frac{z_1}{z_1 + z_2}\right), \end{aligned}$$

where

$$A(t) = \frac{1}{2} + \frac{1}{2}\left[1 - Mt(1-t)\right]^{1/2}. \tag{2.38}$$

By checking the conditions Equation 3.2 for $A'(t)$

$$A'(t) = -\frac{1}{2}\left(\left(1 - M \times t(1-t)\right)^{-\frac{1}{2}} \times \left(\frac{1}{2}M \times (1-2t)\right)\right)$$

we found that at the two endpoints the first derivative is dependent on the dependence parameter: $\lim_{t\to 0^+} A'(t) = -1/4 \times M$ and $\lim_{t\to 1^-} A'(t) = 1/4 \times M$ (for illustration see the right panel in Figure 2.9). Therefore Equation 3.2 can not be fulfilled for any $M \in (0, 4)$. This means that the Schlather model does not lead to an absolutely continuous BGPD model. □

Additional literature related to modeling extreme events can be found e.g. in Coles (2001), Finkenstadt and Rootzén (2004), Galambos (1987), Gumbel (1958), de Haan and Ferreira (2006), Leadbetter et al. (1983), Reiss and Thomas (2007) and Resnick (1987). An add-on `R` packages as `evd` by Stephenson (2002) and `SpatialExtremes` by Ribatet (2012) are also available.

2.4.2 Connection to Generalized Pareto Processes

The generalization of the GPD concept, called the generalized Pareto process, is defined in (3.1) of the very recent paper of Ferreira and de Haan (2012). In the same paper it is shown (Theorem 4.1) that if a process X is in the domain of attraction of a max-stable process then it is necessarily in the domain of attraction

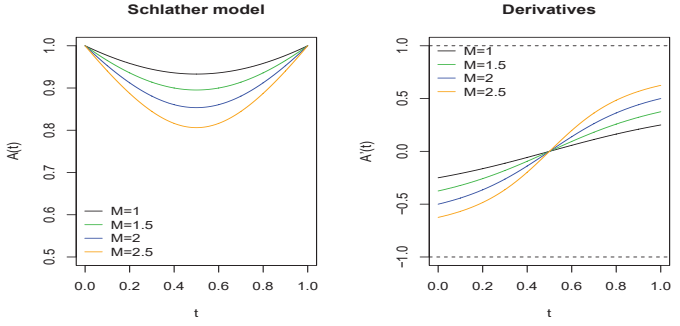

Figure 2.9: Dependence functions of the Schlather model with different parameters and the derivatives, Equation 3.2 is not fulfilled.

of the corresponding generalized Pareto process and both limiting processes share the same exponent measure. This results of the above authors are new and provide various possibilities for further applications of threshold exceedance models in spatial analysis of extremes.

Chapter 3

Parametric Families and Extensions

3.1 Models in 2D

In the previous chapter when representing the characterizations of MEVD and MGPD models there has been no parametric assumption made. Although, of course, for statistical inference it is very useful to have rich and flexible parametric families satisfying the constraints of the representations. In this section we show numerous subclasses of dependence structures, and also propose a construction method for building further asymmetric models based on the existing ones. We take special care on those models which lead to absolutely continuous MGPD models (see Theorem 10), as we intend to use maximum likelihood inference as well when applying them to wind speed data later on.

3.1.1 Classes of Bivariate Dependence Models

The most popular and well-studied dependence models are collected below. The list is not exhaustive but covers a rather wide range of families. Further details about the presented dependence models can be found in the respective papers indicated, giving the first appearance of the related models. Further models can be found e.g. in Section 3.4 of Kotz and Nadarajah (2000) or in Section 9.2.2 of Beirlant

et al. (2004).We present the exponent measure function $V_\star$ from Equation 2.8 as well as the spectral density $w(t)$ from Equation 2.12 for every model. (Although dependence functions are not explicitly given below, some graphical illustrations by the `evd` package of `R` can be found in Figure 3.1, Figure 3.2, and Figure 3.3).

Asymmetric logistic model (Tawn, 1988)

$$\begin{aligned} V_\star(x,y) &= (1-\psi_1)/x + (1-\psi_2)/y + ((\psi_1/x)^\alpha + (\psi_2/y)^\alpha)^{1/\alpha} \\ w(t) &= (\alpha-1)\psi_1^\alpha\psi_2^\alpha(t(1-t))^{\alpha-2}((\psi_2 t)^\alpha + (\psi_1(1-t))^\alpha)^{1/\alpha-1}, \end{aligned}$$

where $\alpha > 1$ and $0 \leq \psi_1; \psi_2 \leq 1$. It allows exchangeability unless $\psi_1 = \psi_2$. In the special case if $\psi_1 = \psi_2 = 1$, it is called symmetric logistic model. This is the only case when the model has all its mass in the interior of the simplex W, otherwise $W(\{0\}) = 1 - \psi_2$ and $W(\{1\}) = 1 - \psi_1$.

Asymmetric negative logistic model (Joe, 1990)

$$\begin{aligned} V_\star(x,y) &= 1/x + 1/y - ((\psi_1/x)^\alpha + (\psi_2/y)^\alpha)^{-1/\alpha} \\ w(t) &= (1-\alpha)\psi_1^\alpha\psi_2^\alpha(t(1-t))^{\alpha-2}((\psi_2 t)^\alpha + (\psi_1(1-t))^\alpha)^{1/\alpha-1}, \end{aligned}$$

where $\alpha > 0$ and $0 < \psi_1; \psi_2 \leq 1$. $W(\{0\}) = 1 - \psi_2$, $W(\{1\}) = 1 - \psi_1$ and so $\psi_1 = \psi_2 = 1$ gives the only symmetric version with all its mass in the interior. We refer to it as the symmetric negative logistic model.

Bilogistic model (Smith, 1990)

$$V_\star(x,y) = \int_{[0,1]} \max\left\{\frac{(\psi_1-1)z^{-1/\psi_1}}{\psi_1 x}, \frac{(\psi_2-1)(1-z)^{-1/\psi_2}}{\psi_2 y}\right\} dz,$$

where $\psi_1; \psi_2 > 1$. A disadvantage is though, that there is only an implicit formula for its spectral density on (0, 1) in terms of the root of an equation as

$$w(t) = \frac{(1-\psi_1)(1-q)q^{1-\psi_1}}{(1-t)t^2((1-q)\psi_1 + q\psi_2)},$$

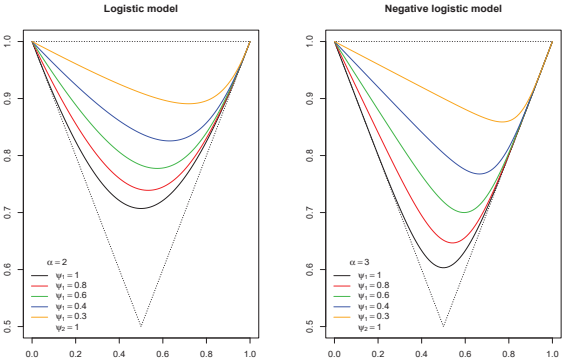

Figure 3.1: Asymmetric dependence functions: logistic and negative logistic models.

where $q = q(x, y; \psi_1, \psi_2)$ is the root of the equation

$$(1 - \psi_1)x(1 - q)^{\psi_2} - (1 - \psi_2)yq^{\psi_1} = 0.$$

For valid input parameters the equation has a unique root in [0,1], what makes the numerical root finding quite handy in this case.

Negative bilogistic model (Coles and Tawn, 1994)

The negative bilogistic model has the same exponent measure function as the bilogistic model except that in this case $\psi_1; \psi_2 < 0$. The spectral density is $w_{\text{Negbilog}}(t) = -w_{\text{Bilog}}(t)$ and similarly $W(\{0\}) = W(\{1\}) = 0$.

Tajvidi's generalized symmetric logistic model (Tajvidi, 1996)

$$V_\star(x, y) = \left(\left(\frac{1}{x}\right)^{2\alpha} + 2(1 + \psi)\left(\frac{1}{xy}\right)^{\alpha} + \left(\frac{1}{y}\right)^{2\alpha} \right)^{\frac{1}{2\alpha}},$$

where $1 \leq \alpha$ and $1 < \psi \leq 2\alpha - 2$. Complete dependence arises when $\alpha \to \infty$, while independence occurs as $\alpha = 1$ and $\psi = 0$. The model has all its mass in the interior

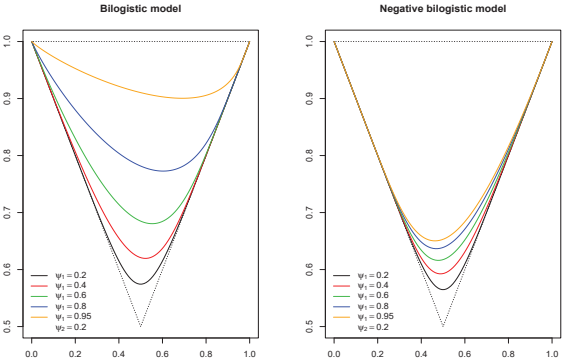

Figure 3.2: Asymmetric dependence functions: bilogistic and negative bilogistic models.

of the simplex, $W(\{0\}) = W(\{1\}) = 0$.

Polynomial model (Klüppelberg and May, 1999)

$$\begin{aligned} V_\star(x,y) &= 1/x + 1/y - \sum_{k=2}^{m} \psi_k \sum_{l=0}^{m-k} \binom{m-k}{l} \frac{x^{l+k-1} y^{m-k-l-1}}{(x+y)^{m-1}} \\ w(t) &= m(m-1)\psi_m t^{m-2} + (m-1)(m-2)\psi_{m-1} t^{m-3} + \ldots + 2\psi_2, \end{aligned}$$

where $\psi_2 > 0, \sum_{k=2}^{m} \psi_k \geq 0, 0 \leq \sum_{k=2}^{m} (k-1)\psi_k \leq 1$ and $\sum_{k=2}^{m} k(k-1)\psi_k \geq 0$. The spectral measure W on the boundary points is $W(\{0\}) = 1 - \sum_{k=2}^{m} \psi_k$ and $W(\{1\}) = 1 - \sum_{k=2}^{m} (k-1)\psi_k$.

A special case is the asymmetric mixed model (Tawn, 1988)

$$\begin{aligned} V_\star(x,y) &= 1/x + 1/y - xy(x+y)^{-2}((\psi_1 + \psi_2)/x + (\psi_1 + 2\psi_2)/y) \\ w(t) &= \psi_1 t^3 + \psi_2 t^2 - (\psi_1 + \psi_2)t + 1, \end{aligned}$$

where $\psi_1 \geq 0$, $\psi_1 + \psi_2 \leq 1$, $\psi_1 + 2\psi_2 \leq 1$ and $\psi_1 + 3\psi_2 \geq 0$. When $\psi_1 = 0$, the asymmetric mixed model reduces to the symmetric mixed model, in such a case

$W(\{0\}) = 1 - \psi_2$ and $W(\{1\}) = 1 - \psi_2$. Consequently $\psi_2 = 1$ is the only symmetric case when all mass is in the interior of [0,1].

Beta model (Coles and Tawn, 1991)

The beta model is asymmetric and satisfies $W(\{0\}) = W(\{1\}) = 0$ like the two bilogistic models above.

$$V_\star(x, y) = 1/x(1 - \beta(\psi_1 + 1, \psi_2; q)) + 1/y\beta(\psi_1, \psi_2 + 1; q) \tag{3.1}$$

and

$$w(t) = \frac{\psi_1^{\psi_1}\psi_2^{\psi_2}\Gamma(\psi_1 + \psi_2 + 1)}{\Gamma(\psi_1)\Gamma(\psi_2)} \frac{t^{\psi_1 - 1}(1 - t)^{\psi_2 - 1}}{(\psi_1 t + \psi_2(1 - t))^{1+\psi_1+\psi_2}}, t \in (0, 1),$$

where $q = \psi_1 y/(\psi_1 y + \psi_2 x)$ $\psi_1, \psi_2 > 0$ and $\beta()$ is a normalized incomplete beta function.

Remark 16. *Although this model family origins from the Beta(ψ_1, ψ_2) density function*

$$\bar{w}(t) = \frac{\Gamma(\psi_1 + \psi_2)}{\Gamma(\psi_1)\Gamma(\psi_2)} t^{\psi_1 - 1}(1 - t)^{\psi_2 - 1}$$

applying the construction form in Equation 2.13, there is also an explicit formula for the exponent measure.

3.1.2 Construction of new asymmetric models in 2D

Here we focus on those specific dependence models, which can be considered in absolutely continuous BGPD models. Since BGPD models are defined by the underlying BEVD model, and since BEVD models are defined by their dependence structures, the most popular parametric families for describing the dependence structures of BEVD models might be considered for the BGPD models. Unfortunately not all of them can be applied without further complications. E.g. we pointed out earlier that a class of BGPD, represented by an absolutely continuous class of BEVD, is not necessarily absolutely continuous. Conditions providing absolutely continuous MGPD models can be found in Theorem 10 and equivalent forms for BGPD are presented in Corollary 4. Some important properties of these models are summarized in

Table 3.1: Summary of bivariate dependence models. The asymmetric logistic and negative logistic are not absolutely continuous. Asymmetric models with valid density are the bilogistic, negative bilogistic and Dirichlet models, but all of them have further complications in calculation. New models denoted by * are proposed later in subsection 3.1.2.

Model	Asym.	Density	Complications
Sym. Logistic	–	✓	–
Asym. Logistic	✓	–	–
Ψ-logistic*	✓	✓	convexity constraints
Φ-logistic*	✓	✓	convexity constraints
Sym. Neg. Logistic	–	✓	–
Asym. Neg. Logistic	✓	–	–
Ψ-negative logistic*	✓	✓	convexity constraints
Φ-negative logistic*	✓	✓	convexity constraints
Bilogistic	✓	✓	not explicit
Neg. Bilogistic	✓	✓	not explicit
Tajvidi's	–	✓	–
C-T	✓	✓	only $w(t)$ is given
Mixed	–	if $\psi_2 = 1$	–
Asym. Mixed	✓	–	–

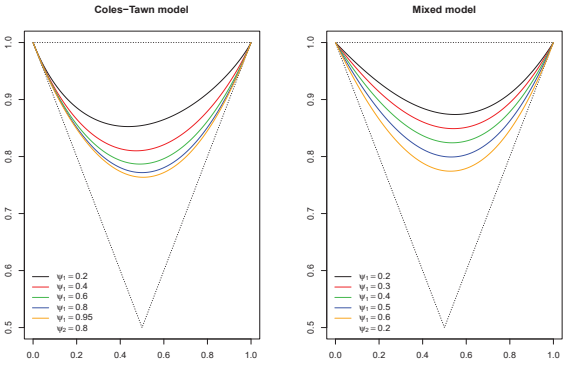

Figure 3.3: Asymmetric dependence functions: Coles-Tawn (or Beta) and asymmetric mixed models.

Table 3.1. All of the presented parametric models have dependence function which are differentiable in $(0, 1)$, hence without mentioning specifically it is assumed by default. The first and second column together show which specific parametric cases include an absolutely continuous and asymmetric model. From the third column of Table 3.1 we can see that there are a lack of easily computable asymmetric models, especially if all probability mass is required to be put on the interior of the simplex $\mathcal{S}_2$, ensuring the absolute continuity of the model. The bilogistic and negative bilogistic models do not have an explicit formula for their exponent measure function. The C-T model only have an explicit formula for the spectral density.

In order to extend the possible classes of models we propose a methodology, which allows for the construction of new dependence models with extra asymmetry parameter(s) from any valid models (four examples are denoted by * in Table 3.1). As the result of this method we may obtain more flexible asymmetric models defining a new class of absolutely continuous BGPD. We illustrate the method - only for mathematical simplicity - in the bivariate setting, but the same methodology can be eliminated to the higher dimensional cases as well (see subsection 3.2.2). As the

method is based on the representation of the BGPD by the dependence function, we recall that if the extra condition Equation 3.2

$$-A'(0) = A'(1) = 1 \tag{3.2}$$

is fulfilled, then the corresponding BGPD is absolutely continuous. Hence we need to find a suitable transformation $\Upsilon(x)$ as the next algorithm shows:

1. Take a parametric **baseline dependence model** satisfying the extra condition in Equation 3.2 as well (e.g. symmetric logistic/negative logistic/mixed, etc.);

2. Take a strictly monotonic **transformation** $\Upsilon(x) : [0,1] \to [0,1]$, such that $\Upsilon(0) = 0$, $\Upsilon(1) = 1$;

3. Construct a **new dependence model** from the baseline model $A_\Upsilon(t) = A(\Upsilon(t))$;

4. **Check** whether A_Υ is still a valid dependence function satisfying even the extra condition of Equation 3.2.

For the construction it is natural to assume that it has a form of

$$\Upsilon(t) = t + f(t),$$

hence

$$\begin{aligned} \Big(A(\Upsilon(t))\Big)' &= A'(\Upsilon(t))\Upsilon'(t) = A'(\Upsilon(t)) \times [1 + f'(t)] \\ \Big(A(\Upsilon(t))\Big)'' &= A''(t + f(t))(1 + 2f'(t)) + A'(t + f(t))f''(t). \end{aligned} \tag{3.3}$$

Then by choosing f such that $f'(0) = f'(1) = 0$, (Equation 3.2) remains automatically fulfilled. although the boundary condition and convexity need to be checked. Particulary, convexity (the positivity of the second derivative) ensures the existence of the density. In the following there are two different examples shown, which ones in general lead to a wide class of valid models.

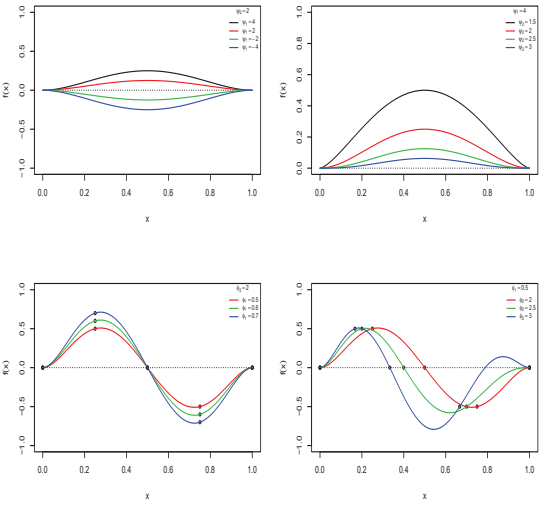

Figure 3.4: Two-parametric class of transformations for model extension. (The chosen parameters are arbitrary, and do not necessarily fulfill any constrains for the resulted dependence function.) Upper panels: some f_{ψ_1,ψ_2} functions; Lower panels: some f_{ϕ_1,ϕ_2} functions.

3.1.3 Example 1: Ψ-transformation

Rakonczai and Zempléni (2012) found that the following functional form leads to a wide class of valid models

$$f_{\psi_1,\psi_2}(t) = \psi_1(t(1-t))^{\psi_2}, \text{ for } t \in [0,1], \tag{3.4}$$

where $\psi_1 \in \mathbb{R}$ and $\psi_2 \geq 1$ are asymmetry parameters. The transformation involving f_{ψ_1,ψ_2} will be denoted by Ψ in the following. Applications on logistic dependence are shown in Figure 3.5. Numerically computed parameter regions verifying convexity can be found in Figure 3.6 for $\Psi-$logistic (on the left) and $\Psi-$negative logistic (on

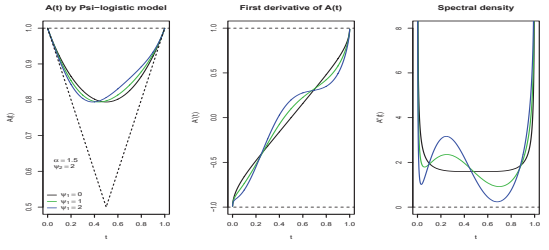

Figure 3.5: Ψ-logistic dependence functions and their derivatives.

the right).

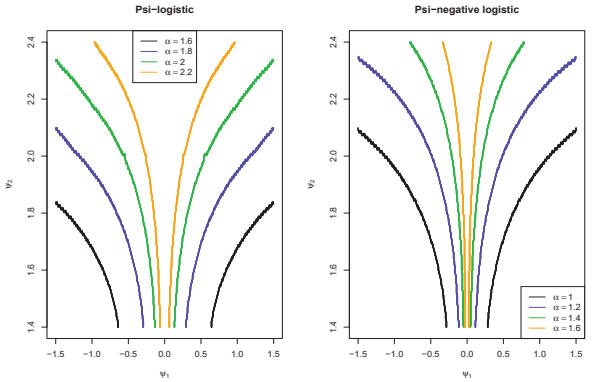

Figure 3.6: Parameters ranges of ψ_1, ψ_2 and α fulfilling convexity constraints. The valid parameters are between the identically colored curves for a given α.

3.1.4 Example 2: Φ-transformation

As the function f in Equation 3.4 takes only non-positive or non-negative values for a given set of parameters, it might be useful to find another functional form which

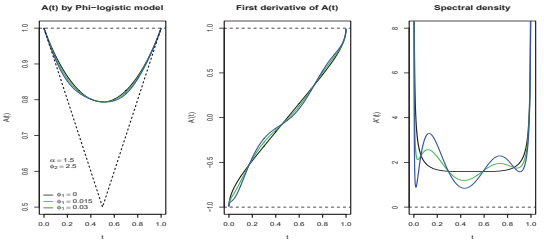

Figure 3.7: Φ-logistic dependence functions and their derivatives.

allows positive and negative values as well. (By doing so the graph of transformation goes below and over the identity). We may try to find a function with two parameters ϕ_1, ϕ_2 which satisfies the following constraints (C):

- $f_{\phi_1,\phi_2}(0) = f_{\phi_1,\phi_2}(1) = 0$ and $f'_{\phi_1,\phi_2}(0) = f'_{\phi_1,\phi_2}(1) = 0$,
- $f_{\phi_1,\phi_2}\left(\frac{1}{\phi_2}\right) = 0$,
- $f_{\phi_1,\phi_2}\left(\frac{1}{2\phi_2}\right) = \phi_1$ and $f_{\phi_1,\phi_2}\left(\frac{\phi_2+1}{2\phi_2}\right) = -\phi_1$.

By using Hermite interpolation it is easy to find a polynomial of degree 6 according to (C)

$$\begin{aligned} f_{\phi_1,\phi_2}(t) \;=\; & \frac{64\phi_1(\phi_2^5-\phi_2^4-2\phi_2^3-2\phi_2^2+5\phi_2-2)\phi_2^5}{(\phi_2+1)^2(\phi_2-1)^2(-1+2\phi_2)(1+2\phi_2^2-3\phi_2)}t^6 \\ & + \frac{-32\phi_1(5\phi_2^6-2\phi_2^5-13\phi_2^4-12\phi_2^3+15\phi_2^2+6\phi_2-6)\phi_2^4}{(\phi_2+1)^2(\phi_2-1)^2(-1+2\phi_2)(1+2\phi_2^2-3\phi_2)}t^5 \\ & + \frac{32\phi_1(4\phi_2^7+3\phi_2^6-14\phi_2^5-15\phi_2^4+23\phi_2^2-8\phi_2-2)\phi_2^3}{(\phi_2+1)^2(\phi_2-1)^2(-1+2\phi_2)(1+2\phi_2^2-3\phi_2)}t^4 \\ & + \frac{-32\phi_1(\phi_2^7+4\phi_2^6-5\phi_2^5-10\phi_2^4-5\phi_2^3+12\phi_2^2+2\phi_2-4)\phi_2^3}{(1+2\phi_2^2-3\phi_2)(\phi_2+1)^2(4\phi_2-1+2\phi_2^3-5\phi_2^2)}t^3 \\ & + \frac{32\phi_1(\phi_2^6-3\phi_2^4-\phi_2^2+4\phi_2-2)\phi_2^3}{(1+2\phi_2^2-3\phi_2)(\phi_2+1)^2(4\phi_2-1+2\phi_2^3-5\phi_2^2)}t^2. \end{aligned}$$

(See Figure 3.7 for examples with logistic baseline.) The transformation involving f_{ϕ_1,ϕ_2} will be denoted by Φ in the following. The same transformation is discussed

in section 2.3 of Rakonczai (2012). Applications on logistic dependence are shown in Figure 3.7.

Finally, solving the convexity inequations provides the valid range for the asymmetry parameters. We should note, however, that this range depends on the parameters of the baseline dependence model. As the set of valid parameters are often too complicated to determine beforehand, we would rather suggest to check to convexity for any set of parameters one by one and then accept or reject them accordingly. The shape of the valid parameter set - verifying convexity - is investigated by numerical calculations in the MSc. thesis of Kusper (2010), where e.g. Figure 12 shows 3 regions of (ψ_1, ψ_2) asymmetry parameters assuming different dependence parameters $\alpha = 0.3, 0.5$ and 0.8 within the $\Phi-$logistic model, respectively.

Remark 17. *We should also note that the baseline model is embedded in the new classes as for $\psi_1 = 0$ or $\phi_1 = 0$ the transformations coincide with identity. This fact helps in estimating the new parameters setting the baseline parameters as initial parameters for further optimization.*

In the next two sections we will see how the bivariate densities can be computed from the characterization by the dependence function. These densities will be the base for the maximum likelihood estimation, prediction regions and goodness-of-fit.

3.2 Models in $d > 2$ dimension

In this section we investigate some well-known parametric dependence structures which can be possibly considered in the MGPD framework. Beyond that we will generalize the procedure of subsection 3.1.2 for higher dimensions. In subsection 3.2.2 we present how even a very simple symmetric dependence model with one single dependence parameter can be expanded to a very flexible family with additional asymmetry parameters. These novel structures have been applied for wind data in subsection 5.4.3.

3.2.1 Classes of Multivariate Dependence Models

There are four multivariate models presented in this section. The first two models generally do not fulfill the condition of Theorem 10, but in the simplest symmetric case both of them do, allowing us to consider these symmetric models as baseline models for model construction. The third and fourth model are rather popular asymmetric models for MEVD. Although both of them satisfy even the condition of Theorem 10 their application in the MGPD framework is cumbersome as there is no explicit formula available for their exponent measure. (Some inference can be done though using some advanced statistical estimation methods presented in subsection 4.1.2 later.)

Multivariate Asymmetric Logistic Model (Tawn, 1990)

Let B be the set of all non-empty subsets of $\{1, \dots, d\}$, let $B_1 = \{b \in B : |b| = 1\}$, where $|b|$ denotes the number of elements in the set b, and let $B_i = \{b \in B : i \in b\}$. The exponent measure function assuming unit Fréchet margins is

$$V_\star(x_1, \dots, x_d) = \sum_{b \in B} \left\{ \sum_{i \in b} \left(\frac{\psi_{i,b}}{x_i} \right)^{\alpha_b} \right\}^{1/\alpha_b},$$

where the dependence parameters $\alpha_b > 1$ for all $b \in B \setminus B_1$, and the asymmetry parameters $\psi_{i,b} \in [0, 1]$ for all $b \in B$ and $i \in b$. Further constraints and associated densities can be found in Kotz and Nadarajah (2000), see Section 3.5.1. It is easy to see (based on the bivariate case) that any nontrivial asymmetric setting of parameters leads to positive mass on the boundary, and so violates the constraints of Theorem 10. Only the multivariate symmetric logistic model fulfills the above requirements. Unfortunately this version is very limited, as it has one single parameter to capture the strength of dependence among all components. Its dependence function in the trivariate case is the following

$$A_{\text{Triv.Sym.Log}} = \left(t_1^\alpha + t_2^\alpha + (1 - t_1 - t_2)^\alpha \right)^{\frac{1}{\alpha}}, \tag{3.5}$$

where $\alpha > 1$. Graphical illustration of this dependence function is in Figure 3.8.

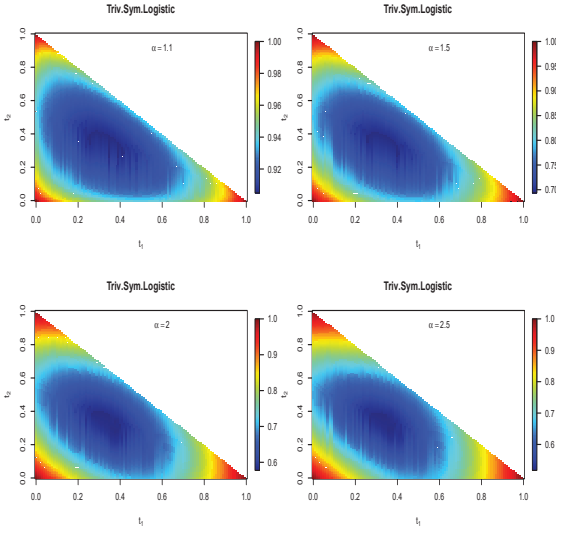

Figure 3.8: Dependence functions of trivariate symmetric logistic dependence models.

Multivariate Asymmetric Negative Logistic Model (Joe, 1990)

The exponent measure function assuming unit Fréchet margins is

$$V_\star(x_1,\ldots,x_d) = \sum_{i=1}^{d} \frac{1}{x_i} - \sum_{b\in B:|b|\geq 2} (-1)^{|b|} \left\{ \sum_{i\in b} \left(\frac{\psi_{i,b}}{x_i}\right)^{\alpha_b} \right\}^{1/\alpha_b},$$

where the dependence parameters $\alpha_b > 0$ for all $b \in B \setminus B_1$, and the asymmetry parameters $\psi_{i,b} \in [0,1]$ for all $b \in B$ and $i \in b$. Further constraints and associated densities can be found in Kotz and Nadarajah (2000), see Section 3.5.2. Similarly to the logistic structure only the symmetric version fulfills the requirements of Theorem 10. Its dependence function in the trivariate case is the following

$$A_{\text{Triv.Sym.Neg.Log}} = 1 - \left(t_1^{-\alpha} + t_2^{-\alpha} + (1-t_1-t_2)^{-\alpha} \right)^{-\frac{1}{\alpha}}, \tag{3.6}$$

where $\alpha > 0$. Graphical illustration of this dependence function is in Figure 3.9.

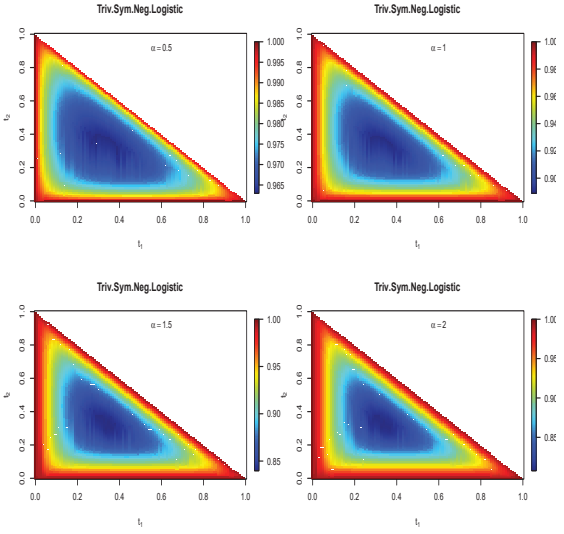

Figure 3.9: Dependence functions of trivariate symmetric negative logistic dependence models.

Dirichlet Model (Coles and Tawn, 1991)

Applying Equation 2.13 for the p-variate Dirichlet density function we can construct the following spectral density

$$w(\mathbf{t}) = \frac{\Gamma(\psi_1 + \cdots + \psi_d)}{(\psi_1 t_1 + \ldots \psi_d t_d)^{d+1}} \prod_{j=1}^{d} \frac{\psi_j}{\Gamma(\psi_j)} \prod_{j=1}^{d} \Big(\frac{\psi_j t_j}{\psi_1 t_1 + \ldots \psi_d t_d} \Big)^{\psi_j - 1}, \tag{3.7}$$

where $\psi_i > 0$ for $j = 1, \ldots, d$ and $\mathbf{t} \in S_d$. This leads to an asymmetric structure, symmetry arises when $\psi_1 = \cdots = \psi_d$. The form for the exponent measure is complicated, only numerically computable. For more details see Section 3.5.4 in Kotz and Nadarajah (2000).

Remark 18. *We should notice a useful property of this model. If* $(X_1, \ldots, X_d)$ *has Dirichlet dependence structure with* $(\psi_1, \ldots, \psi_d)$*, then* (X_i, X_j) *has bivariate Dirichlet (see Beta model in section 3.1.1) structure with parameters* (ψ_i, ψ_j) *for each* $i \neq j$*.*

Pairwise Beta Model (Cooley et al.,2010)

Similarly to the Dirichlet model the pairwise beta model is also defined by its spectral density. This is based on the pairwise beta density below

$$w_{i,j}(\mathbf{t}; \psi) = \frac{d-1}{\sqrt{d}} \frac{\Gamma(2\psi)}{\Gamma(\psi)^2} \left(\frac{t_i}{t_i + t_j} \right)^{\psi - 1} \left(\frac{t_j}{t_i + t_j} \right)^{\psi - 1}, \tag{3.8}$$

where $\mathbf{t} \in S_d$, $\psi > 0$ and $1 \leq i < j \leq d$. Using this structure for every $\frac{d(d-1)}{2}$ pairs, after the necessary standardization we get the pairwise beta model as follows

$$\begin{aligned} w(\mathbf{t}; \alpha, (\psi)_{i,j}) &= \frac{1}{\binom{d}{2}} \frac{1}{(d-1)(d-2)} \frac{\Gamma(\alpha d + 1)}{\Gamma(2\alpha + 1)\Gamma\{\alpha(2-\alpha)\}} \\ &\times \sum_{1 \leq i < j \leq d} (t_i + t_j)^{2\alpha - 1} \{1 - (t_i + t_j)\}^{(\alpha-1)(d-2)} w_{i,j}(\mathbf{t}; \psi_{i,j}), \end{aligned} \tag{3.9}$$

where $\alpha > 0$ is an additional global parameter. This model allows for asymmetry and the parameters are easy to interpret as well.

Remark 19. *This model is actually proven to be a special case of a more general construction principle, for more details see Ballani and Schlather (2011). The beta distributions may be replaced by any other distribution on* $[0, 1]$ *with expectation* $1/2$*. Further distributions constructed in the above paper are the pairwise exponential, weighted Dirichlet and weighted exponential models.*

3.2.2 New asymmetric models in higher dimensions

The idea of subsection 3.1.2 can be applied successfully in higher than 2 dimensional cases as well. Finding a baseline model according to Theorem 10 (as e.g. in Equation 3.5 or in Equation 3.6) and a feasible transformation preserving the necessary constraints of a (multivariate) dependence function

1. $\max(t_1, \ldots, t_d) \leq A(t_1, \ldots, t_{d-1}) \leq 1$;

2. $A(t_1, \ldots, t_{d-1})$ is convex;

may lead to a more flexible class of dependence structures. Although by increasing dimensions, checking convexity becomes more and more computationally intensive, we may check the second-order condition for convexity, i.e. A is convex if and only if the Hessian matrix

$$\nabla^2 A(t)_{i,j} = \frac{\partial^2 A(t)}{\partial t_i \partial t_j}, \tag{3.10}$$

for $i, j = 1, \ldots, d-1$ is positive semi-definite for all $t \in S_{d-1}$. Expanding the idea of the $\Psi-$transformation in subsection 3.1.3 we can consider $\Psi(t_1, \ldots, t_{d-1})$ as

$$\left(t_1 + \psi_{1,1}\left(t_1 \left[1 - \sum_{i=1}^{d-1} t_i \right] \right)^{\psi_{1,2}}, \ldots, t_{d-1} + \psi_{d-1,1}\left(t_{d-1} \left[1 - \sum_{i=1}^{d-1} t_i \right] \right)^{\psi_{d-1,2}} \right).$$

For instance we can take the trivariate symmetric logistic dependence model (Equation 3.5) and apply the following transformation

$$\Psi(t_1, t_2) = (t_1 + \psi_{1,1}(t_1(1 - t_1 - t_2))^{\psi_{2,1}}, t_2 + \psi_{2,1}(t_2(1 - t_1 - t_2))^{\psi_{2,2}}),$$

by choosing $\psi_{1,1} = 0.1, \psi_{1,2} = 2, \psi_{2,1} = 0.2, \psi_{2,2} = 2$. The differences between the new models and the baseline models are illustrated in Figure 3.10 for different α dependence parameters. Further comparisons between the second order derivatives of the above models can be found in Figure 3.11.

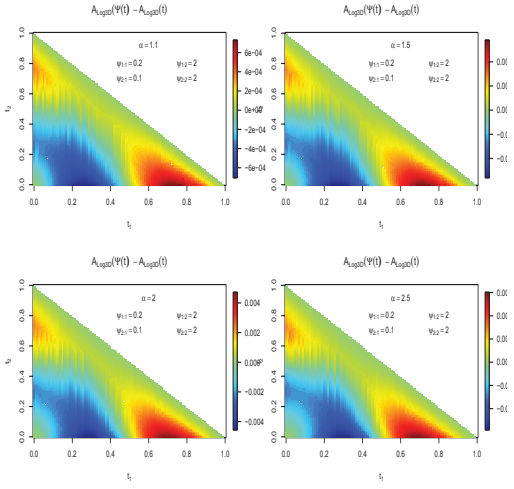

Figure 3.10: Differences between the trivariate $\Psi-$ logistic ($\psi_{1,1} = 0.1, \psi_{1,2} = 2, \psi_{2,1} = 0.2, \psi_{2,2} = 2$) and the baseline logistic models with different α dependence parameters.

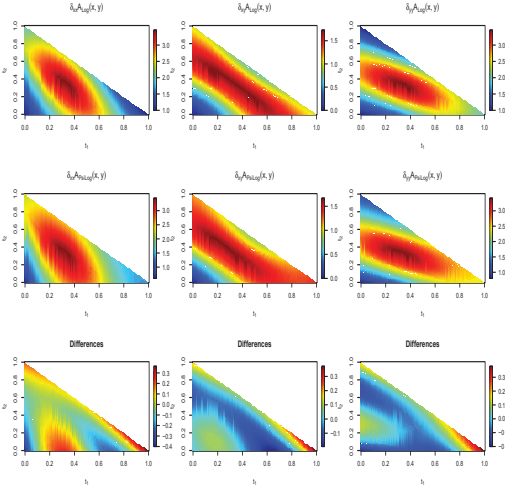

Figure 3.11: Second order derivatives for the trivariate $\Psi-$logistic ($\psi_{1,1} = 0.1, \psi_{1,2} = 2, \psi_{2,1} = 0.2, \psi_{2,2} = 2$), the baseline logistic model with $\alpha = 2$ dependence parameter and the differences between the above two functions.

3.3 Copula Models

This approach for describing dependence is more general than those mentioned above as copulas are applicable for any multivariate distribution. The main theorem is proven by Sklar (1959) who has shown that for *any* d-variate distribution function H, with univariate margins F_i there exits a copula C, a distribution over the d-dimensional unit cube, with uniform margins, such that $H(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$. Moreover the copula C is unique if the marginal distributions are continuous. For more details see Chapter 4 in Cherubini et al. (2004). In the following we present the characterization of MEVD by copulas and give a brief introduction of some parametric copula families which are not strictly linked to MEVD.

Extreme Value Copulas

The extreme value copula family is used to represent the MEVD by uniformly distributed margins.

Definition 7. *The copula C_{MEV} is called to be an extreme value copula if it is a copula of a MEVD G as*

$$G(\mathbf{x}) = C_{MEV}(G_1(x_1), ..., G_d(x_d)). \tag{3.11}$$

This family is well characterized by the *stability property* defined as follows.

Definition 8. *A copula C is called to be stable if $C^s(u) = C(u_1^s, ..., u_d^s)$.*

It can be shown that an extreme value copula of a MEVD must necessarily be stable and conversely, if a copula is stable then it is an extreme value copula. As the copula representation is also an equivalent form for a MEVD, it is possible to write it in terms of the other dependence concepts. For instance a bivariate extreme value copula can be written in terms of Pickands dependence function as

$$C_{\mathrm{BEV}}(u_1, u_2) = \exp\left\{ \log(u_1 u_2) A\left(\frac{\log u_2}{\log u_1 u_2} \right) \right\}, \ (u_1, u_2) \in [0, 1]^2.$$

Elliptical Copulas

Elliptical copulas are the copulas of elliptical distributions as multivariate Gauss or Student-t distributions (see 4.8 in Cherubini et al. 2004).The main advantage of this class is that one can specify different levels of correlation for every pairs of margins. Unfortunately they do not have closed form expressions and are restricted to have radial symmetry. The Gaussian copula family can be derived from the multivariate Gaussian distribution function with mean zero and correlation matrix Σ, transforming the margins by the inverse of the standard normal distribution function Φ as

$$\begin{aligned} \mathbf{C}_{\Sigma}^{Ga}(\mathbf{u}) &= \Phi_{\Sigma}(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} ... \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}} dx_1...dx_d. \end{aligned} \tag{3.12}$$

See Figure 3.12 for bivariate distribution and density functions of the Gaussian copula. Another member of the elliptical copula family is the Student's t copula, is similar to Equation 3.12, but the Gaussian distributions are replaced by t-distributions in the formula. The Student's t copula is defined as

$$\begin{aligned} \mathbf{C}_{\Sigma,\upsilon}^{T}(\mathbf{u}) &= t_{\Sigma,\upsilon}(t_{\upsilon}^{-1}(u_1), ..., t_{\upsilon}^{-1}(u_d)) \\ &= \int_{-\infty}^{t_{\upsilon}^{-1}(u_1)} ... \int_{-\infty}^{t_{\upsilon}^{-1}(u_d)} \frac{\Gamma(\frac{\upsilon+n}{2})|\Sigma|^{\frac{1}{2}}}{\Gamma(\frac{\upsilon}{2})(\upsilon\pi)^{\frac{n}{2}}} \Big(1 + \frac{1}{\upsilon}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\Big)^{\frac{\upsilon+n}{2}} dx_1...dx_d, \end{aligned} \tag{3.13}$$

where υ is the number of degrees of freedom.

Archimedean Copulas

Another broad class of copulas is called the Archimedean copula-family. Its structure is based on a so-called *generator function*: $\phi(u) : [0, 1] \to [0, \infty]$, which is continuous and strictly decreasing with $\phi(1) = 0$. Then a d-variate Archimedean copula function is

$$\mathbf{C}_{\phi}(\mathbf{u}) = \phi^{-1}\Big(\sum_{i=1}^{d} \phi(u_i)\Big).$$

Although this family has a very simple construction, it unfortunately suffers from considerable limitations. For instance there is only one parameter (or just few) to

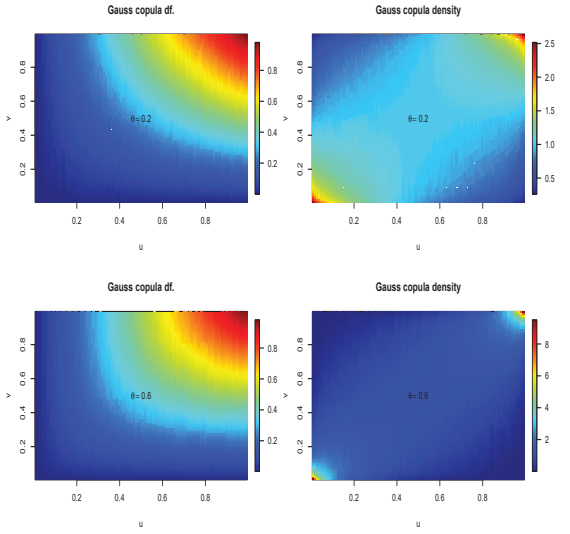

Figure 3.12: Bivariate distribution and density functions of the Gaussian copula with different dependence parameters. In the upper panels $\theta = 0.2$, in the lower panels $\theta = 0.6$.

capture the entire dependence structure and by its construction all of these copulas are permutation symmetric (exchangeable). This means that all $s < d$ dimensional margins are identically distributed, which is usually a too strict assumption in many applications. Here we present two examples from this family, namely the Clayton and the Gumbel copula.

The generator function of the *Clayton copula* (also known as Cook and Johnson's family) is given by $\phi_\theta(u) = u^{-\theta} - 1$, where $\theta > 0$. Thus, the Clayton d-copula function is the following

$$\mathbf{C}_{Clayton}(\mathbf{u}) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta}}. \tag{3.14}$$

The *Gumbel copula* has the generator $\phi_\theta(u) = [-ln(u)]^\theta$, where $\theta \in [1, +\infty)$. Thus, the Gumbel d-copula function is given by

$$\mathbf{C}_{Gumbel}(\mathbf{u}) = e^{-(\sum_{i=1}^{d} - \log(u_i)^\theta)^{\frac{1}{\theta}}}. \tag{3.15}$$

See Figure 3.13 for bivariate distribution and density functions of the Gumbel copula.

Remark 20. *It should be noted that the Gumbel copula also satisfies the stability property in Equation 3.11 and so it belongs to both of the Archimedean and extreme value copula families at the same time. It is actually equivalent with the symmetric logistic structure in section 3.1.1.*

Additional literature related to copula modeling can be found e.g. in Cherubini et al. (2004), Joe (1997), McNeil et al. (2005) and Nelsen (2006).

3.4 From Copulas to Autocopulas

In this section we extend the use of copulas to the interdependence structure of stationary time series, to the analogy of the autocorrelation function. We give a general definition and also a simplified (bivariate) version, which is often largely sufficient for practical purposes.

Definition 9 (Rakonczai et al. 2008a,2008b,2011)**.** *Given a strictly stationary time series Y_t and $\mathcal{L} = \{l_i \in \mathbb{Z}^+, i = 1, ..., d\}$ a set of lags, the autocopula $C_{Y,\mathcal{L}}$ is defined as the copula of the $d+1$-dimensional random vector $(Y_t, Y_{t-l_1} ..., Y_{t-l_d})$.*

Let us remark that the supposed strict stationarity implies that the autocopula does not depend on t.

Definition 10 (Rakonczai et al. 2008a,2008b,2011)**.** *Given a strictly stationary time series Y_t and $l \in \mathbb{Z}^+$ the l-lag autocopula $C_{Y,l}$ is the copula of the bivariate random vector (Y_t, Y_{t-l}). The l-lag autocopulas as the function of the lag l give the autocopula function.*

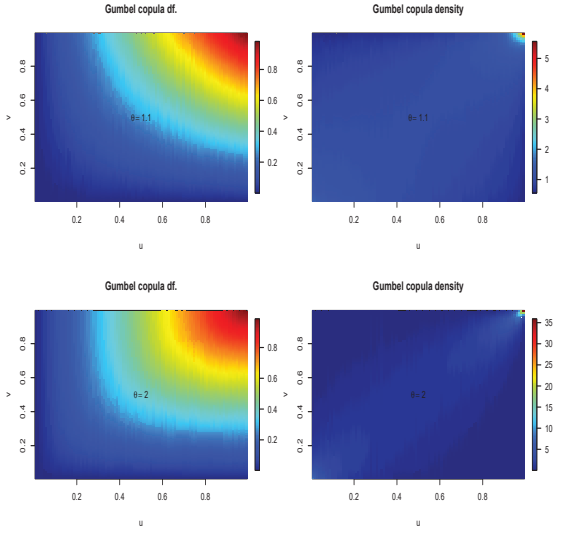

Figure 3.13: Bivariate distribution and density functions of the Gumbel copula with different dependence parameters. In the upper panels $\theta = 1.1$, in the lower panels $\theta = 2$.

One of the most commonly used parametric copula family is the Gauss copula family (see Equation 3.12), for which case the dependence structure is completely determined by the covariances. In the time series setup, the Gaussian copula plays a very specific role, as linear processes with *Gaussian* innovations give rise to Gaussian copulas as both their serial copulas and autocopulas. We emphasize here the dependence of the autocopula on the marginal distribution of the noise by mentioning that linear processes with non-Gaussian innovations do not, in general, have Gaussian autocopulas. Gaussian linear time series with autocorrelation decaying quicker than $1/\log(t)$ do have Gaussian-like interdependence, e.g. their extremes cannot appear to be synchronized. This property can be quantified by the tail-dependence χ (see Equation 5.1 later), which is 0 for the Gaussian copula. While covariances are fully

representative of interdependence in the Gaussian case, they are insufficient for non-Gaussian distributions, especially fail to represent the extremal dependence. Later, in subsection 4.3.2 we present goodness-of-fit tests for copula models, which will be very useful in section 4.6 for testing the autocopulas.

Chapter 4

Estimation, Goodness-of-Fit and Simulation

4.1 Parametric Estimation

In this section we summarize the most important results about maximum likelihood approaches. After the standard results we introduce a method known from the area of spatial statistics providing a very useful tool for estimating models in higher dimensions.

4.1.1 Maximum Likelihood Estimation

In the univariate case, Theorem 1 and Theorem 2 can be used to construct estimation methods for the distribution of maxima or exceedances. It is pointed out in Section 6.3.1 of Embrechts et al. (1997) that there is no explicit solution to the maximum likelihood equations. However in regular cases, when $\xi > -1/2$, there are reliable numerical procedures to find the maximum likelihood estimators. These estimators are efficient, consistent and asymptotically normal. Full discussion about the properties of the estimators, including non-regular ($\xi \leq -1/2$) cases can be found in Smith (1985).

Remark 21. *For applications in insurance, finance or quite a few environmental*

data sets the cases of non-negative shape parameter $\xi \geq 0$ are the most relevant, as these data can rarely be supposed to be bounded to the right.

As it was proven in Smith (1985), maximum likelihood estimators behave regularly in the multivariate case, if the above condition is fulfilled marginally, that is $\xi_i \geq -1/2$ for $i = 1, \ldots, d$. In some cases estimators for the dependence parameters can be supperefficient. For more details, see Section 3.6 in Kotz and Nadarajah (2000). An explicit algebraic formula of the Fisher information matrix is given in Oakes and Manatunga (1992) for the BEVD and in Shi (1995) for the MEVD, both assuming symmetric logistic dependence structure and simultaneous estimation of the marginal and dependence parameters. Even for this simple case of dependence models the density function and score statistic are extremely complicated, hence the elements of the Fisher information matrix are available by numerical computations only. Although similar tedious computations might be possible to carry out for further (and even more complicated) dependence models or MGPD as well, there is no particular advantage presenting the rather untractable results, mainly because they have to be computed numerically anyway. In order to get information about the standard error of estimates we used parametric or nonparametric bootstrap methods depending on the purpose of application, see e.g. Table 4.6, Table 5.6, Table 5.12 or Table 6.1. Maximum likelihood estimation of the time-dependent parameters in the nonstationary BGPD model (introduced in subsection 2.3.5) does not lead to any additional theoretical difficulties over those encountered with the stationary BGPD model, since the equations for linear (or non-linear) trends in the parameters can be easily substituted into the bivariate density. Numerical optimization of the likelihood becomes technically more challenging though, as the number of parameters is increased. This increase in the number of parameters can, at least partially, be addressed by using the parameter estimates from a simpler model (e.g. stationary BGPD) to provide some of the starting values for optimizing the log-likelihood of the nonstationary BGPD model.

4.1.2 Maximum Composite Likelihood Estimation

Unfortunately joint density for the maximum likelihood estimation can be often unavailable, e.g. for max-stable process models only the pairwise marginal distributions are known. Or in some other cases it may happen that the analytical form of the joint distribution $\exp\{-V_\star(z_1,\dots,z_d)\}$ is available (see Equation 2.8), but to obtain the density function from it is computationally infeasible for a high dimension d. Later in subsection 5.4.3 we show a special example of an asymmetric MGPD for $d = 3$ fitted by the "full"-likelihood and another MGPD is presented in Table 6.1 for $d = 5$ fitted by the "pairwise"-likelihood. In the above circumstances it might be useful to base the inference on lower dimensional marginal densities. E.g. if we have a model with a parameter vector θ, assuming that it can be identified from the bivariate marginal densities, θ can be estimated by maximizing the so-called composite log-likelihood function of the form

$$\ell_p(y_1, y_2, \dots, y_d; \theta) = \sum_{1\le i<j\le d} \log f_{i,j}(y_i, y_j; \theta) \tag{4.1}$$

(Lindsay, 1988; Cox and Reid, 2004; Varin, 2008). based on pairwise likelihoods $f_{i,j}(\theta)$. Of course, the composite likelihood can be also written analogously to Equation 4.1 for arbitrary subsets of marginal events.

Theorem 14. *Under appropriate conditions the maximum composite likelihood estimator (MCLE) is consistent and asymptotically normally distributed as*

$$\hat{\theta}_{MCLE} \sim N(\theta, I(\theta)^{-1}),$$

for

$$I(\theta) = H(\theta)J(\theta)^{-1}H(\theta),$$

where $H(\theta) = E(-\nabla\ell(\theta))$ is the observed information matrix and $J(\theta)$ is the estimated variance of the composite score vector. $I(\theta)$ is often called as Godambe information matrix (Godambe, 1960).

The maximum composite likelihood estimator may be unbiased (see Cox and Reid, 2004), but it may not be asymptotically efficient, namely the inverse of the

Godambe information matrix may not attain the Cramér-Rao bound. As numerical maximization of a composite log-likelihood function is not always straightforward, in the later applications we used multiple starting points obtained by preliminary models for the lower dimensional margins. We use this estimation method for a 5 dimensional wind data set later in Table 6.1.

4.2 Nonparametric Dependence Functions

For nonparametric estimation of $A(t)$ we introduce first the modified version of the Pickands estimator. Let (X_1, X_2) denote a bivariate random vector representing the componentwise maxima of an i.i.d. sequence over a given period of time and let (Z_1, Z_2) denote the standardized version of (X_1, X_2) as in subsection 2.2.9. Then $\min\Big\{Z_1/(1-t), Z_2/t\Big\}$ has exponential distribution with mean $1/A(t)$ for any $t \in [0,1]$. The approximation of $1/A(t)$ by the sample mean provides a natural estimation method. Let $(Z_{1,j}, Z_{2,j})$, $j = 1, ..., n$ denote a sample from (Z_1, Z_2). The estimator proposed by Pickands (1981,1989) can be written to the following form

$$\frac{1}{\hat{A}_P(t)} = \frac{1}{n}\sum_{j=1}^{n} \min\Big\{\frac{Z_{1,j}}{1-t}, \frac{Z_{2,j}}{t}\Big\}.$$

However the estimator has the drawback that it might not be a dependence function according to (P). In order to propose some appropriate marginal adjustments, let us define $\bar{Z}_i = n^{-1}\sum_{j=1}^{n} Z_{i,j}$, $i = 1, 2$. The estimator of Hall and Tajvidi (2000)

$$\frac{1}{\hat{A}_{HT}(t)} = \frac{1}{n}\sum_{j=1}^{n} \min\Big\{\frac{Z_{1,j}/\bar{Z}_1}{1-t}, \frac{Z_{2,j}/\bar{Z}_2}{t}\Big\}, \tag{4.2}$$

satisfies $\hat{A}_{HT}(0) = \hat{A}_{HT}(0) = 1$ as well as $\hat{A}_{HT}(t) \geq \max(t, 1-t)$. Although it is still not necessarily convex, by replacing it with its greatest convex minorant $\breve{A}_{HT}$, we can obtain an estimator, which already satisfies all the necessary criteria of (P). The theoretical properties of the above estimators have been shown in Hall and Tajvidi (2000), some of them are summarized below.

Theorem 15. *Both $\hat{A}_{HT}$ and its greatest convex minorant $\breve{A}_{HT}$ are uniformly $\sqrt{n}$ consistent estimators for A in the full interval $[0,1]$. Moreover, if A is twice differ-*

entiable then $\hat{A}'_{HT}$ *and* $\breve{A}'_{HT}$ *are uniformly* $\sqrt{n}$ *consistent estimators for* A' *on any interval* $[\epsilon, 1-\epsilon]$*, for* $\epsilon > 0$.

Theorem 16. *Both* $\hat{A}_{HT}$ *and its greatest convex minorant* $\breve{A}_{HT}$ *have biases of order* n^{-1} *as estimators of* A*. Similarly,* $\hat{A}'_{HT}$ *and* $\breve{A}'_{HT}$ *have biases of order* n^{-1} *as estimators of* A'.

At the expense of its flexibility even $\breve{A}_{HT}$ might not fulfill the "extra" property of differentiability, so the BEVD density function is still not available by assuming $\breve{A}_{HT}$ to be A in (Equation 2.19). To tackle this problem there has been another modification suggested in Hall and Tajvidi (2000),namely that $\hat{A}_{HT}$ can be approximated by smoothing splines, constrained to satisfy (P). By choosing an appropriately fine $0 = t_0 < t_1 < \cdots < t_m = 1$ division of the interval $[0, 1]$ and a given smoothing parameter $\lambda > 0$, we can take $\breve{A}_\lambda$ to be the polynomial smoothing spline of degree 3 or more which minimizes

$$\sum_{j=0}^{m} \left(\hat{A}_{HT}(t_j) - \breve{A}_\lambda(t_j) \right)^2 + \lambda \int_0^1 \breve{A}''_\lambda(t)^2 dt, \tag{4.3}$$

subject to $\breve{A}_\lambda(0) = \breve{A}_\lambda(1) = 1$, $\breve{A}'_\lambda(0) \geq -1$, $\breve{A}'_\lambda(1) \leq 1$ and $\breve{A}''_\lambda(t) > 0$ on [0,1]. By solving the non-linear optimization problem above we can obtain a proper nonparametric estimator, suitable for density estimation. Although for $\breve{A}_\lambda$ it is possible to preserve the $\sqrt{n}$ consistency by choosing an appropriate smoothing parameter, the approach will not lead to a consistent estimation of the second derivative of A. For the best choice of the smoothing parameter Hall and Tajvidi (2000) suggest a cross-validation method.

Remark 22. *In reality of course we do not observe* $(Z_{1,j}, Z_{2,j})$*, but* $(X_{1,j}, X_{2,j})$ *from* (X_1, X_2)*, since the marginal distributions are unknown. Hence it is common practice to estimate them by fitting the GEV (or the empirical distribution function) and plug the estimator into the transformation like* $\hat{Z}_{i,j} = -\log \hat{G}_i(X_{i,j})\, i = 1, 2$ *and* $j = 1, ..., n$.

As illustration we may see Figure 4.1 which compares the exact (symbolically calculated) derivatives of the symmetric and asymmetric logistic dependence functions as well as the (numerically approximated) derivatives of the spline–smoothed

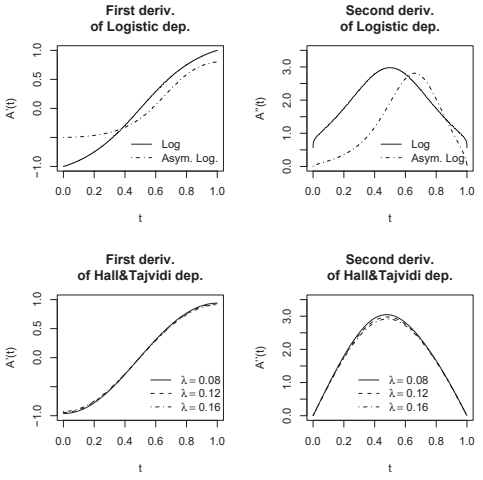

Figure 4.1: Upper block: The first and second derivatives for symmetric logistic ($\alpha = 2.06$) and asymmetric logistic ($\alpha = 2.5, \psi_1 = 0.8, \psi_2 = 0.6$) models. Lower block: The first and second derivatives for the smoothed Hall and Tajvidi estimators with different λ smoothing parameters.

nonparametric functions with different smoothing parameter. We will see a simulation study involving $\breve{A}_\lambda$ in subsection 4.5.1 and in a practical application for wind speed maxima in subsection 5.2.1.

4.3 Goodness-of-Fit Methods

A very important issue after fitting a model is how to verify the choice for the model family. In the following we show two methods for testing model hypotheses. In subsection 4.3.1 we propose to use the very basic concept of χ^2-test for testing bivariate models, with a slight modification of allocating cells for the statistics by density curves instead of by bivariate quantile curves. This approach will be often used in chapter 5 for BEVD and BGPD models. In subsection 4.3.2 we present a goodness-of-fit test for copula models, which will be very useful in section 4.6 for testing the so called autocopulas.

4.3.1 Prediction Regions and GoF Methods

The main motivation behind fitting extreme value models is to have some reasonable estimation for the tail of the distribution. In the univariate case this coincides with estimation of high quantiles, which can be interpreted as estimated values which will be exceeded with a given small probability. Analogously, for bivariate observations one might consider bivariate quantile curves (or semi-infinite prediction region in 3.1 in H all and Tajvidi, 2004).For given probability γ and bivariate distribution F this quantile curve is defined as $C_\gamma = \{(x_1, x_2) : F(x_1, x_2) = \gamma\}$. By definition, any (x_1, x_2) point of C_γ determines a semi-infinite rectangle $(-\infty, x_1] \times (-\infty, x_2]$ which is a $\gamma-$level prediction region for the future values of the variable. However having a compact prediction region in hand would be even more appealing and easier to interpret. See Figure 4.2 for graphical examples.

Prediction Regions

The construction of such a region is described in Hall and Tajvidi (2004), in the following we recall the definition, and later refer to this definition as $(\gamma-)$prediction region.

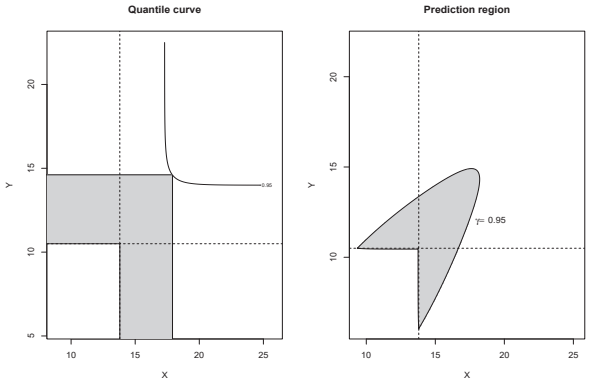

Figure 4.2: Quantile curve (on the left) and compact prediction region (on the right) for logistic BGPD.

Definition 11.

$$\begin{aligned} \hat{\mathcal{R}}(u) &= \{(x,y) : \hat{h}(x,y) \geq u\} \\ \beta(u) &= \int_{\hat{\mathcal{R}}(u)} \hat{h}(x,y)dxdy \end{aligned}$$

for any $\hat{h}$ estimator of a bivariate density h as e.g. in Equation 2.19 or in Equation 2.31. Given a prediction level γ, let $u = \hat{u}_\gamma$ denote the solution of the equation $\beta(u) = \gamma$. Then $\hat{\mathcal{R}}(\hat{u}_\gamma)$ is called a γ-level prediction region for the future values of (X, Y).

Throughout the later sections this kind of (density based) compact prediction regions will be used in order to evaluate the fitted models. A natural idea for model evaluation is comparing the theoretical probability (rate) of points being outside a given region with the observed probability of the same event. Basically, the expected number of observations falling in or out of a given prediction region can be compared with the realizations at different $\gamma-$levels. Such comparisons have been summarized e.g. in Table 4.2 and Table 4.3 on simulated data.

χ^2-Test for Prediction Regions

More quantitative evaluation can be done by using χ^2 goodness-of-fit test based on the above regions. The regions can be viewed as a partition of the plane. As the expected number of observations in each of these partitions is known, we can compare the theoretical frequencies with the realization by a χ^2 statistic. If the parameters were known, the test statistics would follow, approximately, a χ^2 distribution with $(k-1)$ degrees of freedom where k is the number of parts. Unfortunately the parameters are unknown in most of the cases and there is no hope to prove a theorem about the limit distribution of the test-statistics in case of estimated parameters. Simulations do not support the usual limit distribution either. However, the simulated values under the nullhypothesis can serve as the null-distribution of the test and thus they can provide critical values. We do not pursue this idea furter, and use the χ^2-teststatistics for quantitative comparisons only.

This procedure depends on the choice of partitions, so as we are more interested in the higher quantiles we have typically chosen high levels as e.g $\gamma = 0.75, 0.95, 0.99, \ldots$ for the statistics. Often in the practical applications we just use the above χ^2 statistic itself as a measure the distance from a given sample. Specially, we can order the models according to their distance from observed sample. Model comparisons have been made e.g. in Table 5.9 and Table 5.11 on wind speed data.

4.3.2 GoF Methods for Copulas

There has been various alternatives of copula goodness-of-fit methods published in the last decade. Instead of giving a comprehensive review of these methods here we focus on a specific test algorithm of Rakonczai and Zempléni (2007) which is modified version of a basic approach presented by Genest et al. (2006). Further, less computation intensive methods are also available in Kojadinovic et al. (2011) or Kojadinovic and Yan (2011). First, there are some univariate examples presented and then we describe the actual test procedure applied in section 4.6.

Univariate GoF Tests

There are several univariate GoF tests available in the related literature. For our purposes those members of the Cramér-von Mises family have been chosen which are proven to be sensitive to detect discrepancies near the tail of the distribution. Generally they can be formulated (not denoting the dependence on the parameters) as

$$T = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \Phi(x) dF(x),$$

where F_n is the empirical cdf., F is the cdf. which is to be fitted and $\Phi(x)$ is a weight function. In the simplest case, when $\Phi(x) = 1$ we get the Cramér-von Mises statistics:

$$T_{CvM} = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x) \tag{4.4}$$

Focusing on the tails we may set the weight function as $\Phi(x) = (F(x)(1 - F(x)))^{-1}$. Using this weighting we get the Anderson-Darling statistics:

$$T_{AD} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x). \tag{4.5}$$

In many cases when only one of the tails is important (usually maximum for environmental or insurance loss data) the following test statistic is more efficient

$$T_{uAD} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{1 - F(x)} dF(x) \tag{4.6}$$

for the case of maximum and with $\Phi(x) = F(x)$ in the place of $\Phi(x) = 1 - F(x)$ for minimum. The advantage of (Equation 4.6) in comparison to (Equation 4.5) is that its sensitivity is concentrated to discrepancies at the relevant tail of the distribution; see Zempléni's test (Kotz and Nadarajah 2000, p.77). The computation of these statistics is straightforward and for arbitrary weight function $\Phi(x)$ can be numerically approximated.

Copula GoF Test

Even in the multivariate case it is possible to use the univariate techniques described above, after performing an appropriate dimension reduction procedure on

the copula distribution. For this purpose we suggest the use of the so-called Kendall's transform as follows

$$\mathcal{K}(t) = P(H(X,Y) \leq t) = P(C(F(X), G(Y))) \leq t) = P(C(U,V)) \leq t). \quad (4.7)$$

The empirical version of $\mathcal{K}$ can be computed as

$$\mathcal{K}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(E_{in} \leq t),\ t \in [0,1],$$

where

$$E_{in} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(U_j \leq U_i, V_j \leq V_i).$$

Although a closed formula for Equation 4.7 is only available for some specific copula families, $\mathcal{K}(t)$ can be easily approximated by simulation from any given model with the desired accuracy. For illustrations see Figure 4.3. In the later sections we refer to the approximated version of $\mathcal{K}(t)$ simply as K-function.

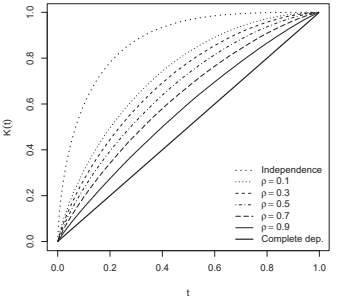

Figure 4.3: Simulated $\mathcal{K}(t)$ for copulas with different Pearson's correlation (Gauss family)

Known tests for checking the match of the theoretical $\mathcal{K}(t)$ and its empirical version use continuous functionals of Kendall's process

$$\kappa_n(t) = \sqrt{n}(\mathcal{K}(\theta_n, t) - \mathcal{K}_n(t)),$$

having favorable asymptotic properties. There are two different kind of approaches investigated in (Genest et al. 2006), Cramér-von Mises type and Kolmogorov-Smirnov type statistics as

$$\begin{aligned} S_n &= \int_0^1 (\kappa_n(t))^2 dt \\ T_n &= \sup_{0\le t\le 1} |\kappa_n(t)|. \end{aligned} \tag{4.8}$$

Since the second statistic T_n is proved to be generally less powerful in detecting discrepancies near the tails, we suggest to base the inference on the test statistics according to S_n, analogously to Equation 4.4, Equation 4.5 and Equation 4.6. The proposed test statistics are summarized in Table 4.1, where $(t_i)_{i=1}^m$ is an appropriately fine division of the interval $(0, 1)$. These statistics have been applied e.g. in Table 4.8 and Table 4.7 later.

Table 4.1: Numerically approximated Cramér-von Mises type test statistics on the Kendall's process

Focused Regions	**Test Statistics**
Global	$S_1 = \frac{1}{m}\sum_{t_i\in[0+\varepsilon,1-\varepsilon]}(\mathcal{K}(\theta_n,t_i) - \mathcal{K}_n(t_i))^2$
Upper Tail	$S_2 = \frac{1}{m}\sum_{t_i\in[0+\varepsilon,1-\varepsilon]} \frac{(\mathcal{K}(\theta_n,t_i)-\mathcal{K}_n(t_i))^2}{1-\mathcal{K}(\theta_n,t_i)}$
Lower Tail	$S_3 = \frac{1}{m}\sum_{t_i\in[0+\varepsilon,1-\varepsilon]} \frac{(\mathcal{K}(\theta_n,t_i)-\mathcal{K}_n(t_i))^2}{\mathcal{K}(\theta_n,t_i)}$
Lower and Upper Tail	$S_4 = \frac{1}{m}\sum_{t_i\in[0+\varepsilon,1-\varepsilon]} \frac{(\mathcal{K}(\theta_n,t_i)-\mathcal{K}_n(t_i))^2}{\mathcal{K}(\theta_n,t_i)(1-\mathcal{K}(\theta_n,t_i))}$

4.3.3 Goodness-of-Fit Tests for Autocopulas

In time series analysis, after estimating a given time series model one must be able to check the fit of the autocopulas of the model at different lags. Formally it is equivalent to check the null-hypotheses for any $l \geq 1$ lag

$$\mathcal{H}_0 : C_{Y_t,l} \in \mathcal{C}_{0,l} = \{C_{\theta,l} : \theta \in \Theta\}, \tag{4.9}$$

e.g. the dependence structure of the investigated autocopula arises from $\mathcal{C}_{0,l}$, which is a copula family defined by a specific time series model. One very crucial question

before starting off is what exactly should be considered as an appropriate sample for the later inference. As the usual inference theory for copulas works with the assumption of i.i.d. observations from copula models one should not [1] use all single pairs of observations from the autocopula

$$\Big\{\Big(Y_i, Y_{i-l}\Big) : i \in \{l+1, ..., n\}\Big\},$$

which are in most cases not independent. To avoid this, we can possibly use a thinned subset of the observations (thinned sample) instead of the original one, e.g. one can take only every m-th pair, where m is presumably large enough (the observations are far enough in time from each other). For simplicity let $\mathcal{T} = \{l+1, l+m+1, l+2m+1, ..., l+rm+1 \leq n\}$ denote the new thinned set of time points and $|\mathcal{T}| = r+1$ the new thinned sample size. Due to the thinning procedure, the dependence among the pairs in

$$\Big\{\Big(Y_i, Y_{i-l}\Big) : i \in \mathcal{T}\Big\}$$

is supposed to disappear and only the dependence between the dimensions remains. Of course, the choice of a proper m for thinning is highly dependent on the given application, and needs elaborate investigations. Hereinafter, following the usual methodology in this context, we consider the marginal distributions as nuisance parameters and base the GoF tests only on rank statistics. Therefore, after carrying out the appropriate thinning we perform the probability integral transformation (PIT) for both margins, viz. we map the margins into the unit interval by their empirical distribution function. The PIT is defined by

$$\Big\{\Big(\underbrace{\frac{\sum_{j\in T}\mathbf{1}(Y_j \leq Y_i)}{|\mathcal{T}|+1}}_{=U_i}, \underbrace{\frac{\sum_{j\in T}\mathbf{1}(Y_{j-l} \leq Y_{i-l})}{|\mathcal{T}|+1}}_{=V_i}\Big) : i \in \mathcal{T}\Big\}.$$

Therefore the $\{(U_i, V_i) : i \in \mathcal{T}\}$ can be interpreted as an i.i.d. sample from the underlying autocopula $C_{Y_t,l}$. Later in section 4.6 we apply goodness-of-fit tests for this kind of thinned copula samples gained from simulated time series. As an example, in section 4.6, the effect of the choice for m is investigated in a certain practical situation.

[1] Or if yes, somehow the dependence should be taken into account.

4.4 Approximate Simulation from BGPD

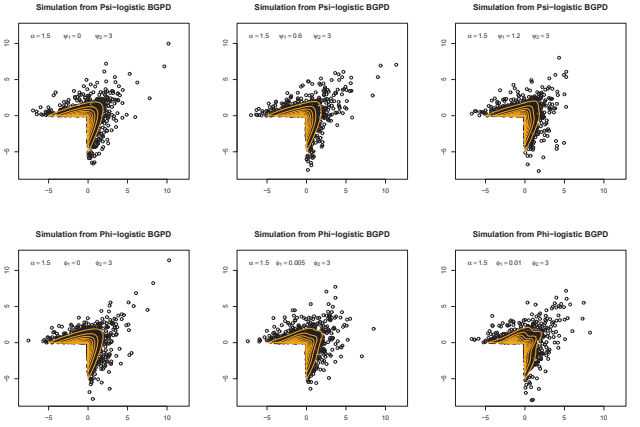

Figure 4.4: Simulated points and density level curves of $\Psi-$logistic and $\Phi-$logistic BGPD models.

There are several approaches known for simulating from BEVD. Shi et al. (1993) present a scheme for simulating from the symmetric logistic model. A more generally applicable scheme is described in Ghoudi et al. (1998). It works for all parametric forms having two times differentiable Pickands dependence function. A third method due to Nadarajah (1999), is different from these, uses limiting point process results as an approximation of the bivariate extremes. Beyond these, further (more general) multivariate simulation methods based on copula models are also available. Although according to our best knowledge the extension of these methods to the proposed BGPD case is not known yet. To resolve this problem here we suggest to use an approximative procedure which can be applied for simulating samples from any d-dimensional random vector $\mathbf{X}$ having distribution function H and density function h.

1. Truncation. Determine a rectangle $R \subset \mathbb{R}^d$ which is large enough for the

desired precision, e.g. $P(\mathbf{X} \in R) = \int_R h(\mathbf{x})d\mathbf{x} = 1 - \epsilon$, for a given $\epsilon > 0$ small real number.

2. Discretization. Divide R to a fine grid equidistantly with step size m and compute the probability of every internal rectangles $P(\mathbf{X} \in R_\gamma)$ for $\gamma \in \Gamma$, where Γ is the index set of the internal rectangles.

3. Adjustment. Cumulate the probabilities from the lower left corner of R to the upper right one and adjust all values by dividing by $1 - \epsilon$.

4. Selection. Generate a sample vector U_i, for $i = 1, \ldots, n$ from the uniform distribution on $[0, 1]$ and select the lower left vertices of the grid rectangles $\mathbf{G}_i = (G_{i,1}, \ldots, G_{i,d})$ allocated by U_i for $i = 1, \ldots, n$.

5. Scattering. Randomize every component of $\mathbf{G}_i$ by adding a scattering component $\sigma_{i,j}$ for $j = 1, \ldots, d$ following uniform distribution on $[0, m]$.

Finally

$$\mathbf{X}_i = (G_{i,1} + \sigma_{i,1}, \ldots, G_{i,d} + \sigma_{i,d})$$

for $i = 1, \ldots, n$ will be the sample having approximately H as distribution function as $\epsilon \to 0$ and $m \to 0$.

Remark 23. *The discretization step can be accelerated by approximating $P(\mathbf{X} \in R_\gamma)$ with the help of the density h. For any R_γ we can choose (randomly) a point $\mathbf{x}_\gamma \in R_\gamma$ and use $h(\mathbf{x}_\gamma) \times m^d$ for the approximation. In the later applications $\mathbf{x}_\gamma$ is typically chosen to be lower left vertex of R_γ.*

Although this method can be very computer intensive, especially if we intend to reach high precision, its bivariate and trivariate versions worked well for our practical applications. As an example there are some simulated samples presented in Figure 4.4. Finally the theorem below provides that the simulation method gives a plausible approximation.

Theorem 17 (Rakonczai and Turkman, 2012)**.** *The simulation algorithm gives an asymptotically unbiased approximation of a continuous distribution H.*

We should notice that the term "asymptotically" does not only mean that the step size tends to zero, but also we assume the boundaries of the large rectangle R tend to infinity. The sketch of the proof for the bivariate case is the following. Let us denote (X, Y) the random vector having distribution function H and density function h. Considering a grid G_m on the whole $\mathbb{R}^2$ with step size m the density function $h_m(x, y)$ of simulated random vector is step function, which consists of horizontal rectangles of size $m \times m$. Moreover the densities are equal $h(g_x, g_y) = h_m(g_x, g_y)$ for any points $(g_x, g_y) \in G_m$. In such a case $\sum_{G_m} h_m(g_x, g_y)m^2$ is actually a Riemann sum, an so

$$\begin{aligned} H_m(x, y) &= \int_{(-\infty,x)\times(-\infty,y)} h_m(x, y)dxdy \approx \sum_{g_x<x, g_y<y} h_m(g_x, g_y)m^2 \quad (4.10) \\ &\approx \int_{(-\infty,x)\times(-\infty,y)} h(x, y)dxdy = H(x, y). \end{aligned}$$

Therefore the distribution function H_m converges to H pointwise. However the grid is in our case on R on which the integral of h is $1 - \epsilon$, so according to 3. in order to have a valid density we must use an adjusting factor $(1 - \epsilon)^{-1}$. Even though this adjusting factor causes bias in the simulated distribution, by the expansion of R the adjusting factor tends to 1 and so the adjusted version of the sum in Equation 4.10 converges to its real value.

4.5 Simulation Study

In the first part of this section we fit and compare BEVD and BGPD models on a wide range of simulated samples in order to reveal, which model gives the more accurate prediction region in a given situation. In the second part of the section we continue with investigating the standard error of the estimates of the newly introduced models by a simulation study. Similar results on real wind data are shown later in subsection 5.4.2 based on bootstrap simulations.

4.5.1 Comparison of BEVD and BGPD Models

The statistical properties of the BGPD (Equation 2.30) are not fully understood yet. In this section we present the results of a simulation study which investigates accuracy of the BGPD estimates and also compares it with the rather standard block maxima approach. The general methodology we followed was that in first step we simulated a bivariate sample from a distribution whose margins were in the domain of attraction of GEV (or GPD, equivalently) and whose dependence structure was symmetric logistic (except in the very last case, when purposely other structures have been chosen). Then, after selecting block maxima or threshold exceedances from the simulated sample, we fitted BEVD for the block maxima and BGPD for the threshold exceedances assuming the logistic family as dependence structure. At the last step we computed the prediction regions (as in definition 11) and checked their accuracy for both models. The model evaluation has been made by comparing the theoretical probability (rate) of points being outside a given region with the observed probability of of the same event, e.g. we computed how many of the simulated values were outside the investigated region. In order to have a comprehensive overview of the accuracy of the proposed exceedance model, wide range of parameters has been used for the simulations. The description of the parameter settings is summarized below:

1. **Margins: exponential or GPD distributions.** Both of these parametric families have GEV or GPD as a limit for their maxima or exceedances, respectively. To be more realistic in the choice of margins, different (ad hoc) parameters have been chosen, specifically $X_1 \sim \text{Exp}(2)$ and $X_2 \sim \text{Exp}(3)$ in the exponential case, and $X_1 \sim \text{GPD}(0.08, 0.13)$ and $X_2 \sim \text{GPD}(0.012, 0.09)$ in the GPD case. (For the GPD distribution the first parameter is the scale, the second is the shape parameter.)

2. **Dependence structures.** Technically, first we simulated from Gumbel copula (Equation 3.15), which is equivalent with the symmetric logistic dependence model in section 3.1.1. Then, in the last example we switched to another type of copula and simulated from Clayton (Equation 3.14), Student (Equation 3.13)

and Gauss family (Equation 3.12).

3. **Association levels.** We used the Kendall's correlation τ as a measure of dependence, and throughout the simulations 3 levels of association have been chosen, as $\tau = 0.3$, $\tau = 0.5$ and $\tau = 0.7$ representing a relatively weak, a medium and a relatively strong association.

4. **Sample sizes.** The sample sizes were chosen as $N = 10\,000$. See Rakonczai and Tajvidi (2010) for $N = 5\,000$ and $20\,000$.

5. **Prediction levels.** As the usual interest in modeling extremes is in the upper tail of distribution (high quantiles) the model performances have been compared for high prediction levels $\gamma = 0.75, 0.95$ and 0.99.

In Table 4.2 and Table 4.3 a side-by-side comparison of BEVD and BGPD models is given assuming different margins and strength of association between them. These simulations show clearly that for logistic dependence at medium and high level of association the BGPD estimates performs acceptably well (similar to BEVD) and there is a strong bias (both in estimates and regions) for low level of association. Further investigations prove that for higher threshold levels the bias disappears. As the increased threshold requires larger sample sizes, the BEVD is more efficient for low correlation. One snapshot of the simulation is displayed by Figure 4.5, where the left block shows the estimated prediction regions for block maxima and the right block for the threshold exceedances (shifted back to the original scale).

Another type of dependence structures different from the Gumbel copula has also been investigated. Their parameters have been chosen to have a certain association of $\tau = 0.5$ being in line with the previous simulations. For results see Table 4.4. Assuming new families of copulas reflects that actually the asymptotic tail dependence has a crucial role in efficiency of BGPD. For example if the original data have been linked by Clayton copula (turned upside down), then also the extremes of their margins are supposed to have strong association. In this case the BGPD has been fairly accurate. However the accuracy diminishes for the Student-t copula, as in that case the extremes are not so strongly associated. Finally the fit is the weakest for Gaussian copula when the dependence parameter is very close to 1. We can also conclude

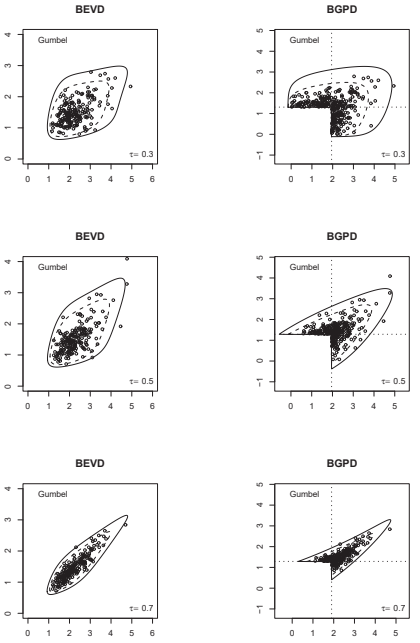

Figure 4.5: Snapshot picture of the simulation in the case of exponential margins, linked by Gumbel copula with $\tau = 0.3, 0.5$ and 0.7 Kendall's correlation. The panels of the left block show the block maxima and their estimated prediction regions by BEVD. The panels of the right block show threshold exceedances and their estimated prediction regions by BGPD. The prediction levels for the regions are $\gamma = 0.99$ (solid line) and $\gamma = 0.95$ (dashed line).

that low tail dependence implies slower convergence in thresholds and so requires more observations for the satisfactory estimation, whereas from medium level of tail dependence the BGPD is similar to BEVD. Further illustration for simulations with

Table 4.2: Expected (by the fitted model) and observed rates of logistic BEVD and logistic BGPD models for simulations with exponential margins.

BEVD vs. BGPD: Exponential margins and Gumbel copula							
BEVD		τ=0.3		τ=0.5		τ=0.7	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.010	0.006	0.011	0.005	0.010	0.006
95%	0.05	0.052	0.010	0.052	0.013	0.052	0.011
75%	0.25	0.256	0.015	0.253	0.017	0.252	0.018
BGPD		τ=0.3		τ=0.5		τ=0.7	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.004	0.004	0.008	0.004	0.010	0.004
95%	0.05	0.037	0.008	0.046	0.009	0.051	0.009
75%	0.25	0.281	0.015	0.257	0.016	0.253	0.016

Table 4.3: Expected (by the fitted model) and observed rates of logistic BEVD and logistic BGPD models for simulations with GPD distributed margins.

BEVD vs. BGPD: GPD margins and Gumbel copula							
BEVD		τ=0.3		τ=0.5		τ=0.7	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.010	0.006	0.011	0.006	0.009	0.006
95%	0.05	0.052	0.012	0.052	0.011	0.049	0.011
75%	0.25	0.253	0.019	0.250	0.018	0.253	0.015
BGPD		τ=0.3		τ=0.5		τ=0.7	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.005	0.004	0.009	0.005	0.011	0.006
95%	0.05	0.040	0.010	0.044	0.010	0.053	0.009
75%	0.25	0.276	0.017	0.252	0.017	0.250	0.015

Table 4.4: Expected and observed rates of logistic BEVD and logistic BGPD models for simulations with exponential margins linked by different copulas having the same $\tau = 0.5$ Kendall's correlation.

BEVD vs. BGPD: Exponential margins and different copulas					
Gumbel		BEVD		BGPD	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.011	0.005	0.008	0.004
95%	0.05	0.052	0.013	0.046	0.009
75%	0.25	0.253	0.017	0.257	0.016
Clayton		BEVD		BGPD	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.010	0.005	0.009	0.005
95%	0.05	0.049	0.010	0.047	0.010
75%	0.25	0.252	0.018	0.251	0.016
Student		BEVD		BGPD	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.009	0.006	0.006	0.004
95%	0.05	0.048	0.010	0.039	0.009
75%	0.25	0.254	0.020	0.269	0.016
Gaussian		BEVD		BGPD	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.008	0.005	0.003	0.002
95%	0.05	0.046	0.011	0.031	0.007
75%	0.25	0.254	0.018	0.268	0.013

different copula models can be found in Figure 4.6.

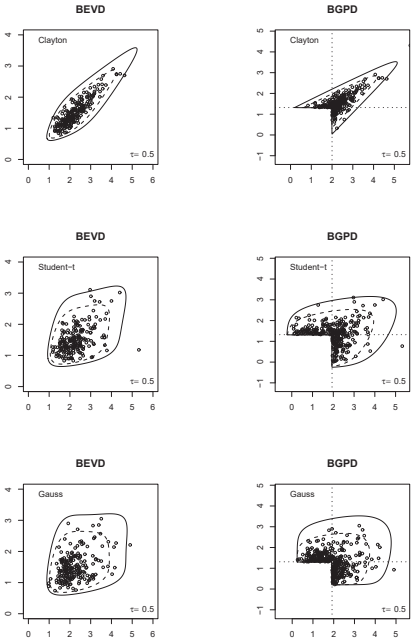

Figure 4.6: Prediction regions for simulations from exponential margins linked by Clayton, Student-t and Gauss copula with $\tau = 0.5$ Kendall's correlation. The prediction levels for the regions are $\gamma = 0.99$ (solid line) and $\gamma = 0.95$ (dashed line).

4.5.2 Standard Error of Asymmetric BGPD Estimates

As statistical properties of the BGPD estimates are not known yet, there has been a simulation study carried out investigating the standard error of the dependence parameters in the new models. Here we focus on model construction presented in subsection 3.1.3 and subsection 3.1.4, and simulate the standard error of their estimates assuming logistic and negative logistic baseline models. Although the simulation procedure (presented in section 4.4) brings an additional uncertainty into the procedure through the discretization error, we believe that the computations are delicately performed and so approximate the real values appropriately.

Study I.

First we have chosen $\Psi-$negative logistic dependence within the BGPD model with $\alpha = 1.3$, 1.4 and 1.5, $\psi_1 = 0$, 0.1, 0.2 and 0.3 and fixed $\psi_2 = 2$ (also assumed as known) parameters. For every set of model parameters (satisfying convexity) 200 samples with a relatively large sample size $n = 2000$ have been generated. The parameters have been re-estimated by maximum likelihood estimation. In this case there have been 6 marginal parameters and 2 dependence parameters to estimate. The mean values and the standard errors are summarized in Table 4.5. The small deviations from the known parameter values and the relatively small standard errors (in brackets) verify the capability of the model construction. (This also means that the simulator procedure we used had been accurate enough.)

Study II.

In another setting we have chosen logistic dependence within the BGPD model having $\alpha = 1.5$ as logistic dependence parameter and various values for the additional asymmetry parameters. In contrast with the previous case (Table 4.5) the second asymmetry parameters ψ_2, ϕ_2 are assumed to be unknown as well. In this case there have been 6 marginal parameters and 3 dependence parameters to estimate. The results can be found in Table 4.6 where both $\Psi-$ and $\Phi-$ models look to be well estimable in the asymmetric cases (if ψ_1 or ϕ_1 are non-zero), e.g. the devia-

Table 4.5: Parameter estimation and standard errors for simulated samples (n=2000) and number of repetitions: 200. If convexity assumption was not fulfilled for the given set of parameters there is "n.a" written.

Simulation study for Ψ-negative logistic BGPD						
	$\alpha = 1.3$		$\alpha = 1.4$		$\alpha = 1.5$	
	$\hat{\alpha}$	$\hat{\psi_1}$	$\hat{\alpha}$	$\hat{\psi_1}$	$\hat{\alpha}$	$\hat{\psi_1}$
$\psi_1 = 0.0$	1.30 (0.04)	0.01 (0.05)	1.40 (0.04)	0.00 (0.05)	1.50 (0.04)	0.00 (0.04)
$\psi_1 = 0.1$	1.30 (0.04)	0.11 (0.05)	1.41 (0.04)	0.12 (0.04)	1.51 (0.04)	0.12 (0.03)
$\psi_1 = 0.2$	1.32 (0.04)	0.23 (0.05)	1.41 (0.05)	0.20 (0.04)	n.a	n.a
$\psi_1 = 0.3$	1.31 (0.05)	0.30 (0.07)	n.a	n.a	n.a	n.a

tions from the known parameter values and the standard errors (in brackets) are still relatively small. Although in the symmetric case, when ψ_1 or ϕ_1 equal to zero, the standard errors are significantly increased. This is due to the fact that in this case ψ_2 or ϕ_2 are irrelevant parameters, and being so their estimates are hardly getting farther from the initial values during the maximum likelihood optimization. The second derivatives of the re-estimated dependence models (spectral density curves) can be seen in Figure 4.7.

Table 4.6: Simulation and re-estimation for $\Psi-$ and $\Phi-$ models made from logistic dependence baseline ($\alpha = 1.5$).

Estimates and standard errors							
Ψ-models				Φ-models			
ψ_1	$\hat{\psi_1}$	ψ_2	$\hat{\psi_2}$	ϕ_1	$\hat{\phi_1}$	ϕ_2	$\hat{\phi_2}$
0.0	0.063(1.173)	2	3.005(1.005)	0.000	−0.001(0.002)	3	2.222(0.613)
0.6	0.643(0.182)	2	2.024(0.154)	0.005	0.005(0.003)	3	2.810(0.474)
1.2	1.261(0.189)	2	2.021(0.095)	0.010	0.010(0.002)	3	3.007(0.164)
1.8	1.810(0.195)	2	2.004(0.083)	0.015	0.015(0.003)	3	2.990(0.141)

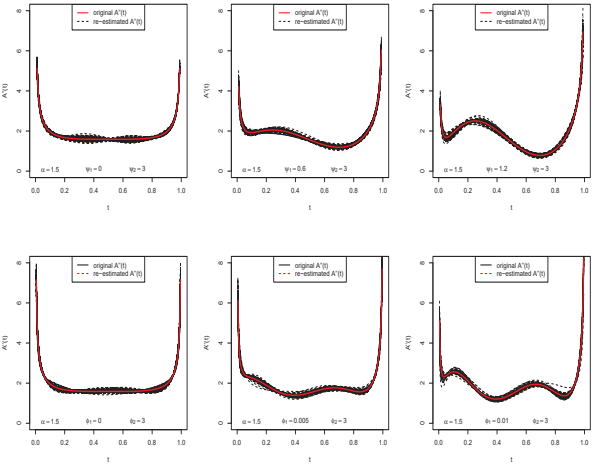

Figure 4.7: Spectral densities of the original and re-estimated models (50 repetitions): $\Psi-$logistic and $\Phi-$logistic models are in the upper and lower panels respectively.

4.6 Testing for Heteroscedasticity in AR Models

Here we illustrate the use of the test procedure suggested in section 4.3.2 and subsection 4.3.3 for autocopulas of stationary time series (see) through a simple but important practical example. More details and a hydrological application can be found in Rakonczai et al. (2011) or in Petrickova (2012). Time series having the same weak AR representation, i.e. complying the same AR model driven by *uncorrelated* innovations do have identical autocovariance structure: the one of the "classical" AR series, generated from that particular AR model by *i.i.d.* innovations. (In the sequel for AR or AR(p) series we always suppose that their innovations are i.i.d.) As a consequence, no test based on autocovariances can really make a distinction between an ARCH- (or eventually GARCH) innovation driven AR (hereinafter AR-ARCH or AR-GARCH) and an i.i.d. one driven AR series. The identification of the autocopula may serve this end, and this is what we intend to show here. So, we use the concept of autocopula for model selection purposes just as the ACF/ACVF is most frequently used. One may argue that the rejection of the AR hypothesis can be achieved e.g. by linearity tests, too, but then the possible alternative is not restricted to one particular type of models, while a well-identified autocopula may point to that within the bounds of reliability inherent in the sample.

In this example we are interested in whether the nonlinearity can be inferred from the sample when the series $Y(t)$ in question satisfies an AR-ARCH(1,1) model. We suppose that we sample from time 1 to N a time series $Y(t)$ satisfying the following equation

$$\begin{aligned} Y(t) &= \phi \cdot Y(t-1) + \varepsilon(t) \\ \varepsilon(t) &= \sigma(t) \cdot Z(t) \\ \sigma^2(t) &= \omega + \alpha \cdot \varepsilon^2(t-1), \end{aligned}$$

where $|\phi| < 1$, $\omega > 0$, $\alpha > 0$, $\alpha + \omega < 1$ and $Z(t)$ is an i.i.d. standard normal series. These conditions imposed on the parameters guarantee the existence of a stationary solution with finite variance (though the study would have been possible under more

relaxed assumptions, including infinite variance, as well). As for the terminology we call $\varepsilon(t)$ to be the innovation and $Z(t)$ the generating noise. We test, and try to reject the null-hypothesis that $Y(t)$ is an AR(1) series with i.i.d. innovations, having the same marginal distribution as the stationary distribution of the ARCH innovation $\varepsilon(t)$.

Since the simulations serve only the illustration of the proposed method, we are not aimed at a full scale investigation of the behavior of various AR-ARCH samples. So, we only consider one fixed setup of the parameters, in particular, we choose the autoregressive coefficient in the equation as $\phi = 0.5$ and the two parameters for the heteroscedastic innovation as $\omega = 0.1$ and $\alpha = 0.85$. This choice results in a definite but not too strong autocorrelation, a significant ARCH effect, and a moderate variance. We simulate relatively large time series samples with size $N = 50000$ from the AR-ARCH series Y(t). In doing this, we first generate the stationary ARCH innovations from standard Gaussian white noise, letting a sufficiently long (50000) step burn in period before we store the actual 50000 innovation values. We then create the AR-ARCH sample from these innovations by R's "arimasim" function. We also generate AR series, by first taking a random reordering of the ARCH sample, to destroy interdependence. This way we obtain an innovation, that various tests accept for an i.i.d. sample. Among them is the BDS-test (Brock-Dechert-Scheinkman) testing serial independence. For all the simulated ARCH samples the BDS-test is highly significant, i.e. rejects the null of serial independence by practically all-zero p-values, and the test is insignificant for the resampled innovations, giving typical p-values in the range of 0.4 and 0.6. This is not surprising, as the sample size is pretty high. The same can be inferred from the autocopula of the resampled series. As we noted above the autocopula of a linear process is dependent on the marginal of the innovation, so, it is of utmost importance that the newly created i.i.d. innovations have *the same marginal distribution* as the stationary ARCH series, and from this resampled innovation we create the required AR series. Re-generating the ARCH sample, and repeating the latter procedure to obtain independent innovations, we get 500 of such AR series and compute from each of them the l-lag autocopulas, and their K-functions. Averaging the K-functions out we obtain a good estimation

of the theoretical $\mathcal{K}$ in (Equation 4.7). This kind of empirical approach is necessary since it seems hopeless to compute an exact closed formula of the true theoretical $\mathcal{K}$. In order to explore the difference we compare the autocopulas of the AR and the AR-ARCH series at the same set of lags, basically by checking the distance of the empirical K-function of the AR-ARCH series from the averaged K-function of the AR series by the help of the test statistics presented in Table 4.1. For the formal test the 95% quantiles of the different test statistics have been computed as critical values. Again, as the distributions of the test statistics are not known the 95% quantiles under the null hypothesis were determined from the simulated AR samples. Practically these critical values represent the maximal "distance" which is still acceptable for a given simulation to be considered as an AR according to the null-hypothesis at $\gamma = 0.05$ significance level. The autocovariance functions of the AR and AR-ARCH series, displayed in the left panels of Figure 4.8, are very similar as expected, and their differences in the first few lags are insignificant, according to the limit distribution of the estimator.

In contrast to the almost full coincidence of the autocovariance functions, the autocopulas differ substantially. For instance the discrepancies between the 1-lag autocopulas can even be detected visually, as is shown in the right panels of Figure 4.8. We have taken a sample from both the AR-ARCH and the AR series created by the above described simulations, and thinned it for every considered lag by choosing every 10th l-lag-apart pairs of values (this means e.g. consecutive pairs in the case of $l = 1$), creating a sample of $(n = 500)$ pairs from every process.

(Note that in our notation N refers to the sample size of the original time series and n refers to the sample size of the thinned copula sample.) In view of the fading away interdependence within the time series the l-lag-apart pairs in the thinned sample for small lags hardly (insignificantly) differ from an *independent* sample of pairs of variables, with the same pairwise interdependence. This is the reason why we use those thinned samples to estimate the autocopula of the given lag. A hint for the choice of the thinning can be obtained from the autocovariance function which is practically zero for $l = 10$, so the elements of autocopula sample do not suffer from the interdependence effect of the time series. (The effect of the choice of

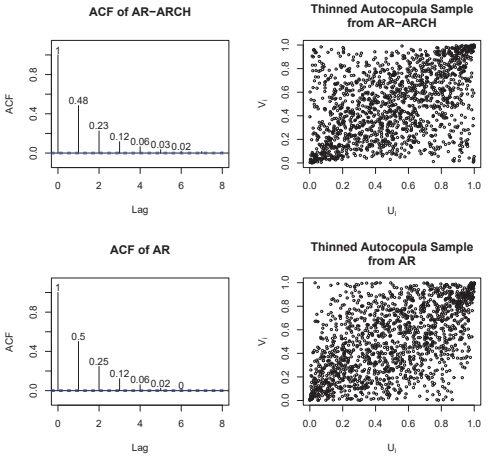

Figure 4.8: Differences between AR and AR-ARCH: The left block shows the estimated ACF-s, whereas the right block shows thinned 1-lag autocopula samples for both time series models, with thinning parameter $s = 10$ and sample size $n = 1500$.

thinning is discussed later at the end of this section.) The difference between higher lag autocopulas is not so clear visually, see Figure 4.9, so there is a definite need for more quantitative investigation, based on formal tests.

GoF tests checking the match of the autocopulas have been performed for $l = 1, ..., 7$ lags based on the 4 statistics presented in Table 4.1. The null-hypothesis stated, as mentioned before, that the autocopula of the sample arises from the AR model. The algorithm we followed at a given l-lag:

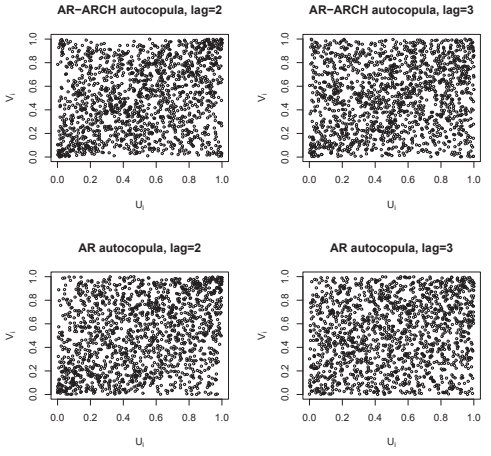

Figure 4.9: Thinned autocopula samples for AR and AR-ARCH models: $l = 2, 3$, thinning parameter $s = 10$ and sample size $n = 1500$.

1. Simulate AR time series $n = 500$ times as described above
2. Obtain their l-lag autocopula sample from the thinned by $s = 10$ series
3. Calculate the test statistics $S_{i,j},\ i = 1, ..., 4$ and $j = 1, ..., 500$ and choose the 0.95 quantiles as critical values $Q_{i,0.95},\ i = 1, ..., 4$
4. Simulate AR-ARCH time series $n = 500$ times
5. Obtain now their l-lag autocopula sample from the thinned by $s = 10$ series
6. Calculate the test statistics[1] $S_{i,j},\ i = 1, ..., 4$ and $j = 1, ..., 500$ and in every case reject H_0 when $S_{i,j} > Q_{i,0.95}$

The results are summarized in Table 4.7.

Although all 4 tests gave very similar results for $l = 1$, we found that the tail-sensitive test statistics performed definitely better, especially for higher lags. Namely

Table 4.7: Rejection percentages of the hypothesis of AR autocopula, if the true model is an AR-ARCH, at significance level $\gamma = 0.05$ and sample size $n = 500$ (out of 500 simulation)

AR-ARCH vs AR at $\gamma = 0.05$, $n = 500$				
lag	S_1	S_2	S_3	S_4
$l = 1$	72.0%	65.7%	68.4%	70.9%
$l = 2$	16.3%	28.0%	17.8%	27.8%
$l = 3$	9.4%	23.7%	10.6%	22.4%
$l = 4$	7.9%	19.5%	9.4%	17.8%
$l = 5$	5.4%	12.0%	6.6%	11.2%
$l = 6$	5.9%	9.8%	6.5%	9.3%
$l = 7$	5.3%	7.4%	5.5%	6.9%

S_2 (with upper weights) and S_4 (both upper and lower weights) turned out to be the most effective. The efficiency was not too high at this sampling level, even so the tests were able to separate almost 70% of the AR-ARCH models from the AR ones by the 1-lag autocopula, and this separation level can be increased by taking into account more lags simultaneously. Of course, one could improve the separation rate by increasing the significance level but by doing so one would reject more AR models incorrectly. In order to reach real and significant improvement in separation one should increase the sample size for the autocopulas, when possible, creating a finer resolution image, thus better approximating the real copula. Our large simulation sample size enabled us to perform the same algorithm with larger autocopula samples. Still applying the same $s = 10$ thinning parameter the tests have been recomputed with new sample sizes as $n = 1000, 1500$. The rejection rates for S_2 and S_4 are shown in Table 4.8.

We conclude that the sample size has a very significant effect on the results. Although the computations assuming larger autocopula samples are more exhaustive, the larger copula samples yield definitely stronger tests. For example the S_4

Table 4.8: Rejection rates computed from samples with sample size $n = 500, 1000$ and 1500 (out of 500 simulations)

AR-ARCH vs. AR by S_2 and S_4 for larger samples						
$s = 10$	S_2 statistics			S_4 statistics		
lag	$n = 500$	$n = 1000$	$n = 1500$	$n = 500$	$n = 1000$	$n = 1500$
$l = 1$	65.7%	93.4%	**99.3%**	70.9%	96.0%	**99.6%**
$l = 2$	28.0%	49.7%	**68.6%**	27.8%	50.4%	**69.9%**
$l = 3$	23.7%	39.6%	**51.6%**	22.4%	37.7%	**50.3%**
$l = 4$	19.5%	28.2%	**34.8%**	17.8%	25.5%	**32.2%**
$l = 5$	12.0%	15.9%	**20.4%**	11.2%	14.1%	**18.4%**
$l = 6$	9.8%	11.4%	**11.5%**	9.3%	10.3%	**10.3%**
$l = 7$	7.4%	8.0%	**8.8%**	6.9%	7.8%	**7.8%**

based test improves by 30% for $l = 1$, resulting in, that almost all of the ARCH time series has been separated from AR successfully. For higher order autocopulas ($l = 2, 3, 4, 5$) the rejection rates increase roughly $1.5 - 2$ times, as well. For greater lags ($l \geq 5$) there is no such improvement any more, but the explanation behind this phenomenon is simply the negligible association in the autocopulas. As with the increasing lags the strengths of association fades out totally, there remains practically no measurable dependence to model and compare. As it has already been mentioned, a theoretical assumption must be fulfilled when using standard results about copulas, namely that the multivariate observations are supposed to be independent. As usual in time series models the assumption of independence between close pairs: (X_t, X_{t-l}) and (X_{t+s}, X_{t+s-l}), s being small, is not realistic. (In this notation $t \geq 1$ and $l \geq 1$ denotes the time and the lag as before, moreover $s \geq 1$ is what we have earlier referred to as the thinning parameter.) Of course the larger s is the better, but in practical applications there is always a limit defined by the given sample size. So one needs to find the optimal s which is large enough for acceptable independence and small enough for having sufficiently large sample for the proper

inference. The decision can be based on the ACF by choosing such a large s for which the ACF is essentially zero.

In the given situation there is no limitation on the sample size as we can generate time series with arbitrary length. In order to investigate the effect of the thinning parameter on the test performance $n = 500$ has been chosen constantly as sample size, avoiding differences appearing in the comparison just because of the different sample sizes. The results for the relevant lags are shown in Table 4.9. By analyzing it, we can find a remarkable relapse in the test performance when the criteria of independence of sample elements is presumably violated, i.e. the observation pairs have been chosen too close to each other, as in the case of $s = 1$, or to a lesser extent for $s = 3$.

Table 4.9: Rejection rates by decreasing $s = 10, 5, 3$ and 1 thinning parameter (out of 500 simulations, with sample size $n = 500$)

AR-ARCH vs. AR by S_2 and S_4 for different thinning parameters								
	S_2 statistics				S_4 statistics			
lag	$s = 10$	$s = 5$	$s = 3$	$s = 1$	$s = 10$	$s = 5$	$s = 3$	$s = 1$
$l = 1$	65.7%	63.7%	63.9%	57.2%	70.9%	69.4%	71.1%	66.6%
$l = 2$	28.0%	26.1%	24.3%	21.4%	27.8%	25.0%	22.9%	21.8%
$l = 3$	23.7%	24.8%	24.1%	17.6%	22.4%	23.3%	22.0%	17.8%
$l = 4$	19.5%	18.4%	16.0%	11.2%	17.8%	18.3%	15.5%	10.2%
$l = 5$	12.0%	12.9%	11.1%	9.4%	11.2%	11.4%	10.0%	8.4%

Chapter 5

Applications to Wind Speed Data

As an illustration of the methods presented in the previous chapters we have investigated a wind speed dataset from north of Germany. Beyond fitting standard and new kind of extreme value models, we have also calculated some prediction regions in order to interpret the model estimates. We have calculated the standard error of the estimates by bootstrap simulations, and compared the differences in model performances by the help of prediction regions.

5.1 Wind Speed Time Series

For our study we have investigated a wind speed time series of 5 sites in Germany, namely Hamburg, Hannover, Bremerhaven, Fehmarn and Schleswig. These observations have been measured hourly for the recent 50 years observed (from 1958 till 2007). In order to reduce the serial correlation within the series, we calculated the maxima of the original observations for each day, and considered them as "daily observations". More precisely, the entire period covers 18061 days, and 17926 observations are actually available during this period (the remaining 135 values are missing). Some segments with length of 100 days are illustrated in Figure 5.1.

Before the joint modeling a preliminary univariate extreme value analysis have been performed as well. (Similar univariate analysis on flood data can be found in Bozsó el al., 2005.) There have been GEV distribution (Equation 2.2) fitted to

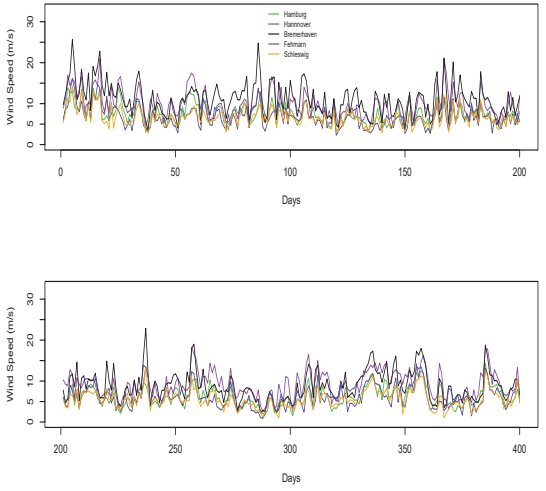

Figure 5.1: Segments of the wind speed data series for Hamburg, Hannover, Bremerhaven, Fehmarn and Schleswig.

the monthly wind speed maxima, the parameter estimates are summarized in the upper block of Table 5.1 whereas the fitted density functions are in the left panel of Figure 5.2. Analogously, for GPD (Equation 2.4) has been fitted to the exceedances over the 99% quantiles, see the lower block of Table 5.1 and the right panel in Figure 5.2. These estimates can also be used e.g. as initial values when estimating joint distributions.

Due to the relatively short distance between these locations the observations are rather strongly correlated. The Kendall's correlation is in the range of $0.43 - 0.62$ (see the legend in Figure 5.3), which can be rated to somewhere between the medium ($\tau = 0.5$) and strong ($\tau = 0.7$) association level according to the simulation study cases. As discussed in section 4.5 prediction regions based on both BEVD and BGPD

Table 5.1: Univariate Analysis: GEV and GPD parameter estimates and standard errors.

GEV parameters					
Param.	Hamburg	Hannover	Bremerhaven	Fehmarn	Schleswig
μ	9.99 (0.08)	15.14 (0.12)	13.43 (0.12)	10.10 (0.09)	9.99 (0.08)
σ	1.84 (0.06)	2.63 (0.08)	2.75 (0.09)	1.88 (0.06)	1.65 (0.05)
γ	-0.03 (0.03)	-0.08 (0.03)	-0.12 (0.02)	-0.02 (0.03)	0.01 (0.03)
GPD parameters					
Param.	Hamburg	Hannover	Bremerhaven	Fehmarn	Schleswig
σ	1.23 (0.15)	1.99 (0.21)	1.85 (0.20)	1.40 (0.15)	1.53 (0.17)
γ	0.12 (0.09)	-0.06 (0.08)	-0.02 (0.08)	0.03 (0.08)	0.03 (0.08)

model hold their nominal level for this degree of association. Further measures as

$$\chi(q) = 2 - \log(P(F_X(X) < q, F_Y(Y) < q))/\log(q) \tag{5.1}$$

and

$$\bar{\chi}(q) = 2\log(1-q)/\log(P(F_X(X) > q, F_Y(Y) > q)) - 1$$

for F_X and F_Y marginal distribution functions and $q \in (0,1)$ can be interpreted as quantile dependent measures of dependence, and so the asymptotic dependence can be measured by their limits at one. For more formulas and exact interpretation see Section 8.4 of Coles, (2001). In Figure 5.3 we can see that the limit of $\chi(q)$ is significantly above zero, and it appears that $\bar{\chi}(q)$ tends towards one as q approaches one, so that the wind speeds at these sites may be considered to be asymptotically dependent.

As the study is based on pairs and triplets of sites, there are many different combinations we may choose from. For the sake of transparency we always picked up those sites which give the best illustration to the proposed methods within a given application. Therefore the chosen sites might be different in different parts of this chapter depending on the given purpose.

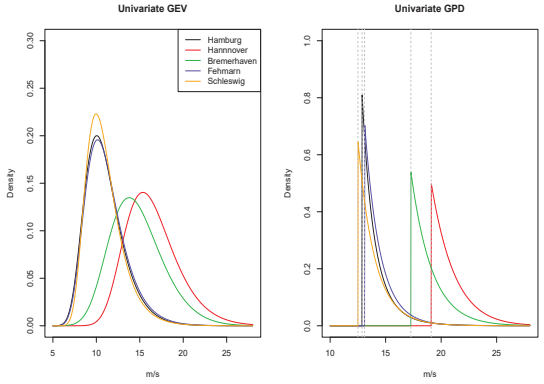

Figure 5.2: Univariate Analysis: GEV density functions fitted to monthly maxima (on the left) and GPD density function fitted to exceedances over the 99% quantiles (on the right).

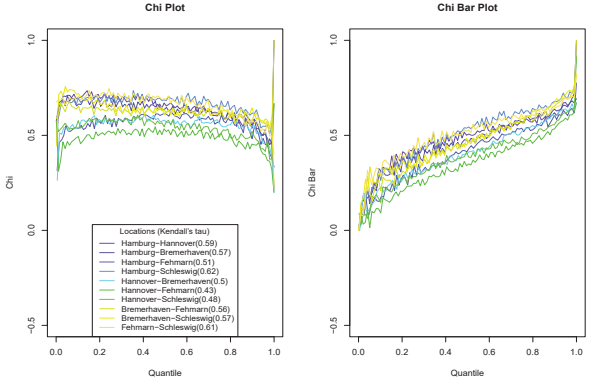

Figure 5.3: χ-plots (on the left) and $\bar{\chi}$-plots (on the right) for the pairs of sites.

5.2 Applications of BEVD and BGPD models

In this section - being in line with subsection 4.5.1 - we compare the BEVD and BGPD methods assuming standard dependence structures. More advanced structures are demonstrated later in section 5.3 for nonstationary models, and in subsection 5.4.2 for asymmetric models. For the analysis daily observations from Hamburg and Hannover have been used which are displayed on Figure 5.4. We fitted BEVD and BGPD models and checked whether the corresponding prediction regions fits the data or not.

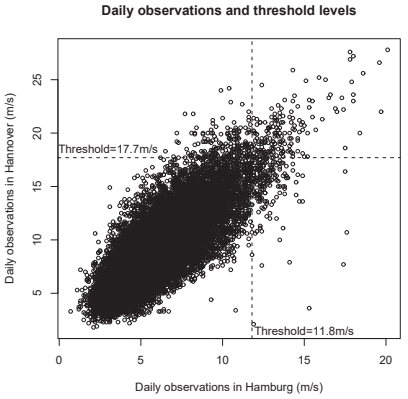

Figure 5.4: The daily observations of wind speed measurements in Hamburg and Hannover with the 98% quantiles of the margins as threshold levels

5.2.1 BEVD Prediction Regions

For modeling the monthly maxima a parametric (logistic) and nonparametric (spline smoothed Hall-Tajvidi) estimators have been used. As the choice of the smoothing parameter λ can have a considerable effect on the estimated curves here

we present two reasonable alternatives of λ. The curves of the prediction regions (on the original scale) are presented in the upper panels of Figure 5.5. Although these regions make the models visually comparable, at the first glance there is no flagrant difference among the applied models. The quantitative results for the estimators are summarized in Table 5.2, where the expected number of observations falling out of a given region has been compared with the observations at different γ levels.

As expected, the results depend on the value of the prediction level, so this kind of comparison is consequently not enough to decide which one is best alternative. Even if a fitted model was very close to the observation for this given "realization" of the wind speeds, it might not be generally the best choice from prediction point of view. As our focus is often on the accuracy of prediction regions (not only on the model fit) beyond the usual goodness-of-fit tests a cross-validation procedure is also needed. Following a standard method we can possibly split the data into two equal complementary parts and then consider one of these sets is as a training set and the other part as a testing set of observations we intend to predict. Technically, as we needed an automatic algorithm for the cross validation method, we started the likelihood maximization from the parameter vector $(\hat{\xi}_1, \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_2, \hat{\mu}_2, \hat{\sigma}_2, \alpha)$ given by the margins. The results are presented in Table 5.3. Generally the prediction regions perform acceptably well at $\gamma = 0.5, 0.75$ and 0.95 prediction levels, perhaps with slightly under-estimation of the quantiles. At the highest $\gamma = 0.99$ level the bias seems to be more serious, on average there are 2 times more observations outside this region than expected.

5.2.2 BGPD Prediction Regions

Analogously to subsection 5.2.1 prediction regions have been calculated for the exceedances as well. First we considered the density given in (Equation 2.31) with a logistic dependence function which has been inherited from the block maxima method and $\alpha = 2.06$ has been kept fixed during the maximum likelihood optimization. Practically by doing this we just adjusted the 6 marginal parameters to $\alpha = 2.06$, a given dependence parameter. After this step we also let the dependence parameter be free and optimized the maximum likelihood in all 7 parameters. The

Table 5.2: Performance of different models at different predictive levels (Hamburg and Hannover).

Level	Expect.	Log BEVD	H&T BEVD	Expect.	Fix Log BGPD	Log BGPD
50%	296	311	316	264	272	278
75%	148	142	149	132	123	119
95%	30	31	34	26	26	23
99%	6	12	15	5	8	3
Log-lik	-	-2531.5	-2533.8	-	-2202.5	-2187.8

second method returned back a higher likelihood value: -2187.8 compared to the first one's -2202.5, suggesting that the fixed dependence parameter $\alpha = 2.06$ might not be accurate enough. The prediction regions are presented in the lower panels of Figure 5.5. The statistics for the regions together with the likelihood values are summarized in Table 5.2.

Similarly to the block maxima method cross-validation has been carried out for the logistic BGPD model as well. In contrast to BEVD model no obvious underestimation was found and the estimated quantiles have been more appropriate compared with the observations, see Table 5.3 for further details.

Table 5.3: Cross-validation: Results on the test data (Hamburg and Hannover).

Level	Expect.	Log BEVD	HT BEVD	Expect.	Log BGPD
50%	148	154.2(7.3)	156.4(6.6)	132	139.3(4.6)
75%	74	71.0(6.4)	73.2(6.3)	66	60.5(3)
95%	15	16.3(2.6)	17.3(2.8)	13.2	12.3(1.9)
99%	3	5.9(1.7)	5.9(1.7)	2.6	1.8(0.9)

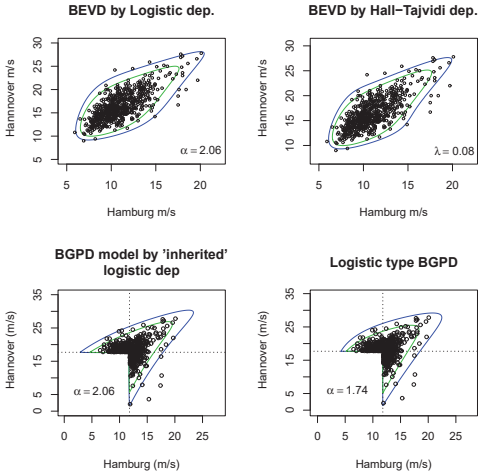

Figure 5.5: Prediction regions of BEVD and BGPD models at high prediction levels, $\gamma = 0.99$ (blue line) and $\gamma = 0.95$ (green line) for Hamburg and Hannover.

5.3 Nonstationary BGPD Models

Here we investigate a nonstationary extension of the BGPD model which allows for the possibility that the characteristics of extreme events are changing over time, or depend upon the value of some other covariate. More details can be found in Rakonczai et al. , (2010). We illustrate the practical application on the wind speed data from the previous section. We focus on modeling threshold exceedances occurring for four pairs of stations - Bremerhaven-Hamburg, Bremerhaven-Hannover, Fehmarn-Hannover and Fehmarn-Bremerhaven (see Figure 5.6) , and, for more detailed analyses, focus upon the data for Fehmarn-Hannover. We use numerical maximum likelihood to fit the stationary BGPD model to bivariate data for each of the four pairs of stations.

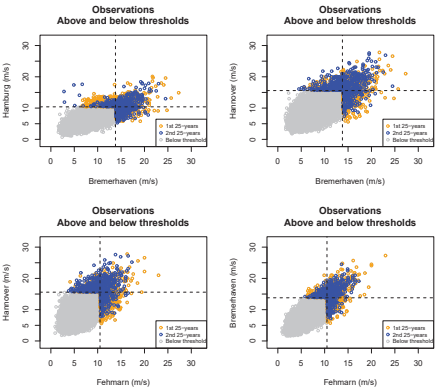

Figure 5.6: Exceedances are distinguished by different colors in order to compare values between the first and second half of the 50 year time period (1957-2007).

5.3.1 Changes over Consecutive Time Periods

In order to explore whether the parameters of the BGPD model are changing over time we have divided the time series into two sections (into two 25 year periods, see Figure 5.6), fitted the stationary BGPD model separately to data for each of these two periods, and compared the resulting parameter estimates. The results of these comparisons, for the ξ_1, ξ_2 and α parameters, are summarized in Table 5.4 together with the estimates of the full 50 years, and appear to show systematic changes in the values of the parameters between the earlier and later periods. These changes could, however, be within the (unknown) variation associated with each estimate. We do not present equivalent results for the μ_1, μ_2 and σ_1, σ_2 parameters, which have no natural interpretation within the context of the BGPD model, but instead compute prediction regions, for four different probability levels ($\gamma = 0.5, 0.75, 0.95$ and 0.99). The joint effects of changes in the seven parameters over time can be explored by looking at changes over time in the profiles of the prediction regions. Prediction regions for four probability levels are also shown in Figure 5.7 for each

Table 5.4: Parameter estimates by fitting the stationary BGPD model to data of (a) the whole period of 50 years, (b) the first 25 years and (c) the last 25 years.

Locations	**Bremerhaven & Hamburg**			**Bremerhaven & Hannover**		
Parameters	ξ_1	ξ_2	α	ξ_1	ξ_2	α
50 years	**0.092**	**0.160**	**1.834**	**0.126**	**0.108**	**1.753**
1st 25 years	0.143	0.135	1.800	0.143	0.131	1.717
2nd 25 years	0.034	0.912	1.897	0.092	−0.042	1.822
Locations	**Fehmarn & Hannover**			**Fehmarn & Bremerhaven**		
Parameters	ξ_1	ξ_2	α	ξ_1	ξ_2	α
50 years	**0.182**	**0.155**	**1.595**	**0.176**	**0.059**	**1.985**
1st 25 years	0.199	0.148	1.571	0.173	0.092	1.960
2nd 25 years	0.146	0.132	1.700	0.172	0.002	2.047

of the two 25 year periods. These figures reveal some evidence for time dependence,

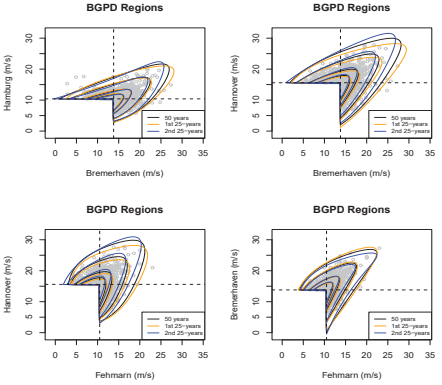

Figure 5.7: Prediction regions with $\gamma = 0.50, 0.75, 0.95$ and 0.99 for the models of Table 5.4.

since the prediction regions for the two periods are rather different. It can be seen that extreme exceedances in Bremerhaven tend to be lower in the second period than in the first period, whilst there is an apparent increase in extremes at Hannover. The strength of association may possibly also be changing in some cases.

5.3.2 Linear Trends in the BGPD Parameters

The exploratory results suggest that the properties of the extremes could be changing over time, and that it is therefore worth considering models that include temporal trends. Several nonstationary BGPD models have been fitted to the data for each of the four pairs of stations: the first of these models (which we call BGPD-1) has only one time-dependent parameter $\alpha(t)$), whilst the final model (BGDP-7) model allows all seven parameters to be linear function of time. The inclusion of more time-dependent parameters into the BGPD model allows it to be more flexible in describing changes over time in the distribution of exceedances, but may lead to overfitting. This trade-off can be examined by studying the magnitude of the increase in log-likelihood values as we move from simpler to more complicated models. In Table 5.5 the logarithms of the maximum likelihood values for the nonstationary models are compared with the corresponding values for the stationary BGPD model (which we call BGDP-0). We can see that the increase in log-likelihood values in moving from the BGPD-0 model to the BGPD-7 model is largest for Fehmarn and Hannover (126), relatively large for Bremerhaven-Hannover (84), and relatively small for the remaining two pairs of sites (24 for Bremerhaven-Hamburg and 28 for Fehmarn-Bremerhaven). These results suggest that the strongest evidence for nonstationarity is for the Fehmarn-Hannover data, but residual temporal dependence at extreme levels (clustering) makes it difficult to formally evaluate the statistical significance of any of these changes. Prediction regions for Fehmarn-Hannover are displayed in Figure 5.8 - the regions are shown for the midpoint of each of ten disjoint five year periods for which data are available. The profiles of the prediction regions for different consecutive time periods can, visually, be of assistance in detecting nonstationary, and determining the nature of this nonstationarity.

Table 5.5: Maximum log-likelihood values associated with fitting the stationary model (BGPD-0) and a range of nonstationary models to the full 50 year dataset.

Maximum Log-likelihood values for nonstationary models				
Models	BGPD-0	BGPD-1	BGPD-2	BGPD-3
Time-dep. in	-	α	ξ_1, ξ_2	ξ_1, ξ_2, α
Bremerhaven-Hamburg	-5412	-5411	-5408	-5398
Bremerhaven-Hannover	-6120	-6112	-6083	-6084
Fehmarn-Hannover	$\mathbf{-6041}$	$\mathbf{-6037}$	$\mathbf{-5961}$	$\mathbf{-5935}$
Fehmarn-Bremerhaven	-5137	-5133	-5117	-5113
Models	BGPD-2	BGPD-3	BGPD-3	BGPD-7
Time-dep. in	σ_1, σ_2	$\sigma_1, \sigma_2, \alpha$	μ_1, μ_2, α	all
Bremerhaven-Hamburg	-5407	-5397	-5400	-5388
Bremerhaven-Hannover	-6078	-6073	-6088	-6036
Fehmarn-Hannover	$\mathbf{-5952}$	$\mathbf{-5933}$	$\mathbf{-5958}$	$\mathbf{-5915}$
Fehmarn-Bremerhaven	-5116	-5113	-5116	-5109

5.3.3 Quantifying Uncertainty by Block Bootstrap

The non-independence of extreme wind speeds, as a result of clustering, invalidates the use of standard approaches for quantifying uncertainty. To investigate the degree of uncertainty associated with estimation a block bootstrap procedure has been performed assuming half year long data blocks from the data. The first part of the bootstrap study used 200 block bootstrap samples to investigate the variation of estimates within the stationary BGPD model. The results for Fehmarn and Hannover are, for illustration, presented in left columns of Table 5.6. Nonstationary BGPD models were also fitted to the block bootstrap samples. Table 5.6 shows the estimated trends over time within a BGPD-3 model that includes time-dependence in ξ_1, ξ_2 and α. The trend parameters are scaled so that they refer to the overall change between the start and end of the time series: they should be divided by 18061 in order to get the daily rate of change and by 49.5 in order to get the annual rate

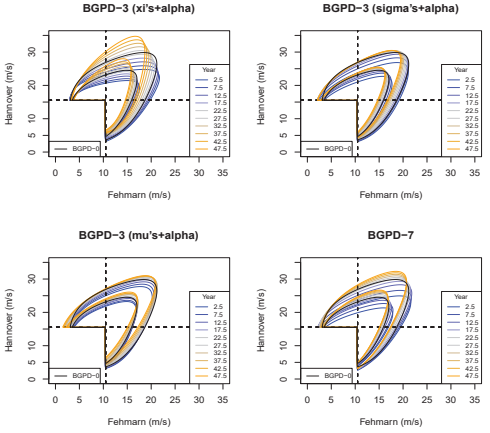

Figure 5.8: Time-dependent prediction regions with $\gamma = 0.95$ and $\gamma = 0.99$, as derived from nonstationary and stationary BGPD models fitted to the data for Fehmarn and Hannover. Prediction regions are shown for the midpoint of each five year period (of which there are ten) and the colours change from blue to orange as time moves forward.

of change.

5.3.4 Goodness-of-Fit and Prediction of Future Distribution

Including linearly varying time-dependent parameters rather than time-constant ones looks to be very promising if there is significant time-dependence to capture. However, the selection of the most appropriate model(s) is a challenging problem, since the range of possible nonstationary BGPD models is rather wide (depending how many, and which, parameters are assumed to be time-dependent).

The goodness-of-fit for a given model can be checked based on the prediction regions that have been already been computed, although in this case the regions are time-dependent as well. Therefore the prediction regions should strictly be re-

Table 5.6: Bootstrap estimates and standard errors for the parameters of a BGPD-0 and BGPD-3 model. "Fitted": estimates for the original data set. "Bootstrapped": mean parameter estimates from the 200 block-bootstrap samples, together with standard errors (in brackets).

Bootstrap results	**BGPD-0**		**BGPD-3**	
Parameters	Fitted	Bootstrapped	Fitted	Bootstrapped
Intercept for ξ_1	0.18	0.18(0.01)	0.18	0.15(0.02)
Trend for ξ_1	–	–	0.00	0.02(0.04)
Intercept for ξ_2	0.15	0.15(0.02)	0.11	0.11(0.03)
Trend for ξ_2	–	–	0.13	0.07(0.05)
Intercept for α	1.59	1.60(0.04)	1.47	1.50(0.06)
Trend for α	–	–	0.25	0.31(0.12)

calculated for every single unique time point, leading to a highly computationally intensive method. In order to avoid this the prediction regions have been recalculated yearly: this division is sufficiently fine to give reasonable results, because differences in the profiles of the regions between consecutive years are very slight. The results for four pairs of stations are summarized in Table 5.7, and clearly show that all models have similar performance for all four levels of α. In each case the very high quantiles seem to be slightly overestimated (for most pairs of stations there are less points outside the $\gamma = 0.99$ and $\gamma = 0.95$ prediction regions than expected), but this phenomena disappears for $\gamma = 0.75$ and there actually seems to be underestimation for $\gamma = 0.5$.

Table 5.7: Results of applying the goodness-of-fit procedure to BGPD-0, BGPD-3 and BGPD-7 models for four pairs of stations. Four different prediction levels are used ($\gamma = 0.99, 0.95, 0.75$ and 0.5) and the time-dependent prediction regions have been recalculated yearly (in the midpoints of the years).

Rate of falling outside prediction region					
Brem. & Hamb.	BGPD-0	BGPD-3	BGPD-3	BGPD-3	BGPD-7
Expected $(1-\gamma)$	-	ξ's+α	σ's+α	μ's+α	all
0.010	0.011	0.008	0.011	0.011	0.009
0.050	0.042	0.039	0.044	0.043	0.040
0.250	0.259	0.261	0.255	0.259	0.261
0.500	0.527	0.533	0.539	0.536	0.538
Brem. & Hann.	BGPD-0	BGPD-3	BGPD-3	BGPD-3	BGPD-7
Expected $(1-\gamma)$	-	ξ's+α	σ's+α	μ's+α	all
0.010	0.004	0.005	0.005	0.005	0.003
0.050	0.030	0.031	0.034	0.033	0.035
0.250	0.263	0.265	0.266	0.262	0.262
0.500	0.538	0.540	0.541	0.536	0.549
Fehm. & Hann.	BGPD-0	BGPD-3	BGPD-3	BGPD-3	BGPD-7
Expected $(1-\gamma)$	-	ξ's+α	σ's+α	μ's+α	all
0.010	0.001	0.002	0.002	0.002	0.004
0.050	0.031	0.032	0.033	0.031	0.029
0.250	0.262	0.278	0.266	0.266	0.258
0.500	0.552	0.562	0.550	0.543	0.542
Fehm. & Brem.	BGPD-0	BGPD-3	BGPD-3	BGPD-3	BGPD-7
Expected $(1-\gamma)$	-	ξ's+α	σ's+α	μ's+α	all
0.010	0.006	0.002	0.004	0.006	0.002
0.050	0.029	0.031	0.033	0.032	0.030
0.250	0.258	0.263	0.255	0.251	0.251
0.500	0.535	0.553	0.543	0.539	0.540

Although the goodness-of-fit results for the stationary and nonstationary models are very similar, the accuracy of predicted values for future time points could still be very different. To check whether this is the case, and whether there is actually value in using a nonstationary rather than a stationary model, we propose a cross-validation procedure. The data have been divided - according to the usual $70\%-30\%$ division - into a training set (the first 35 years of data) and a test set (the final 15 years of data). The models are fitted to the training set of 35 years, and these models are then used to generate predictions for the next 15 years; these predictions can then be compared against the test data. The results are summarized in Table 5.8, where the observed quantiles are compared with theoretical quantiles calculated from different BGPD models. It can clearly be seen that the substantial overestimation of high quantiles within the stationary model disappears when using nonstationary models instead. In general we see that the nonstationary models (BGPD-3 and BGPD-7) perform better than the stationary model (BGPD-0) in modeling high quantiles (which are the most important in the context of wind speeds), but do worse in modeling lower quantiles. Specifically, it appears that the nonstationary models do better than the stationary models for $\gamma = 0.99$ and 0.95, that the results are mixed for $\gamma = 0.75$ and that the stationary model does best for $\gamma = 0.50$.

5.4 Asymmetric MGPD models

In this section the stations Bremerhaven, Fehmarn and Schleswig have been chosen for illustration. The bivariate observations, which exceed the marginal 95% quantiles in at least one coordinate are plotted in Figure 5.9. Throughout this section the observations are considered as being stationary, but nonstationarity can also be handled e.g. by choosing time-dependent model parameters as in section 5.3. We have applied a wide range of dependence models: standard ones (some of those presented in subsection 3.1.1 and summarized in Table 3.1 without $*$ mark) and some of their extension (marked with $*$ in the table and described in subsection 3.1.2).

Remark 24. *For the sake of replicability, there are also some* `R` *codes appended later. Most of these codes are included as examples in the reference manual of* `'mgpd'`

Table 5.8: Comparison of empirical and predicted marginal quantiles (50%, 75%, 95% and 99%) for the final 15 years of data. The predicted values are based on fitting BGPD-0, BGPD-3 and BGPD-7 models to the 35 years of training data.

Cross-validation results on the test data								
Quantiles (m/s)	**50%**		**75%**		**95%**		**99%**	
Brem. & Hamb.	x	y	x	y	x	y	x	y
Observed	14.9	10.7	16.6	11.9	19.1	14.4	20.8	17.4
BGPD-0	14.3	10.9	15.8	12.0	19.8	15.0	24.8	18.8
BGPD-3	14.3	10.8	15.5	11.8	18.9	14.6	23.1	18.1
BGPD-7	14.4	10.6	15.5	11.5	18.2	14.1	21.1	17.5
Brem. & Hann.	x	y	x	y	x	y	x	y
Observed	14.6	16.4	16.2	18.1	18.9	21.9	20.5	25.1
BGPD-0	14.4	16.1	15.9	17.6	19.9	21.5	24.9	25.9
BGPD-3	13.7	16.8	15.2	18.5	18.9	22.7	23.2	27.5
BGPD-7	14.0	16.7	15.4	18.2	18.8	21.5	22.3	24.5
Fehm. & Hann.	x	y	x	y	x	y	x	y
Observed	10.9	16.5	11.9	18.2	14.3	22.0	16.4	25.1
BGPD-0	10.9	16.0	12.0	17.5	15.2	21.5	19.5	26.6
BGPD-3	9.7	16.8	11.1	18.4	13.8	22.8	17.3	28.1
BGPD-7	9.6	16.7	11.0	18.2	13.4	22.1	16.4	27.0
Fehm. & Brem.	x	y	x	y	x	y	x	y
Observed	11.0	14.9	12.1	16.6	14.3	19.1	16.5	20.8
BGPD-0	11.0	14.5	12.1	16.0	15.3	20.0	19.6	24.6
BGPD-3	10.6	14.7	11.5	16.0	14.0	19.4	17.2	23.3
BGPD-7	10.6	14.8	11.6	16.1	14.4	18.8	18.1	21.4

`package`. *Although in order to make the package upload and installation faster, some parts of the examples are commented by "*`##Not run:`*", removing these comments from the code makes all parts run.*

Because of the large number of model parameters (7-9), which are difficult to

interpret, we used bivariate prediction regions as we have already seen for another pairs of stations as e.g. in Figure 5.5 or Figure 5.7.

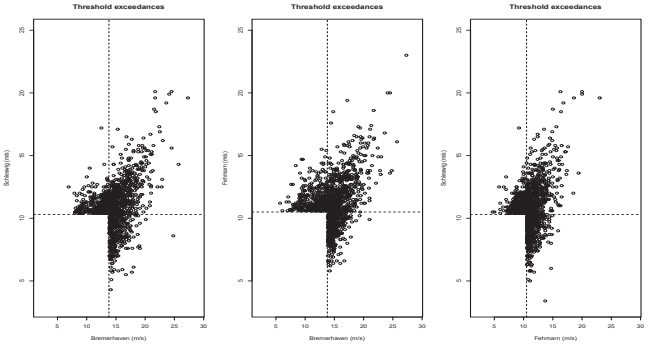

Figure 5.9: Bivariate wind speed observations exceeding the marginal 95% quantiles in at least one coordinate at three pairs of sites.

5.4.1 Baseline BGPD models for wind data

For the three pairs of the above stations there have been 6 different dependence model assumed in the BGPD model, namely logistic, negative logistic, Coles-Tawn (Dirichlet), bilogistic, negative bilogistic and Tajvidi (generalized symmetric logistic). After fitting these models to the data, prediction regions have been computed. In order to compare their fit we apply the test procedure from subsection 4.3.1, which can be easily performed as a simple byproduct of the prediction region method. Some prediction regions at Fehmarn-Schleswig can be seen in Figure 5.10 and in the first column of Figure 5.12. By applying the methods shown in section 4.3.1 we compare the theoretical frequencies with the realization through χ^2 statistics. The estimates of the dependence parameters and the χ^2 statistics for the standard models are summarized in Table 5.9. It seems that in every case an asymmetric model performs the best. Especially, we can see that at Bremerhaven-Schleswig the bilogistic ($\chi^2_{\text{BiLog}} = 15.9$) and Tajvidi ($\chi^2_{\text{Tajvidi}} = 13.6$), at Bremerhaven-Fehmarn the

Table 5.9: Estimates of dependence parameters and χ^2 statistics for the fitted BGPD models.

Performance of standard BGPD models						
Brem.-Schleswig	Log	NegLog	C-T	BiLog	NegBilog	Tajvidi
Par.I	2.06	1.34	2.22	0.54	0.59	2.12
Par.II	-	-	1.25	0.42	0.93	0.09
χ^2	17.77	18.82	18.98	15.91	16.21	13.63
Brem.-Fehmarn	Log	NegLog	C-T	BiLog	NegBilog	Tajvidi
Dep.I	1.98	1.26	1.12	0.43	1.02	2.24
Dep.II	-	-	2.38	0.57	0.60	0.43
χ^2	44.89	26.08	13.09	39.01	22.89	27.77
Fehm.-Schleswig	Log	NegLog	C-T	BiLog	NegBiLog	Tajvidi
Dep.I	1.95	1.22	2.19	0.55	0.61	2.26
Dep.II	-	-	1.06	0.45	1.06	0.59
χ^2	42.82	30.08	24.19	44.36	30.76	27.76

Coles-Tawn ($\chi^2_{\text{C-T}} = 13.1$) and negative bilogistic ($\chi^2_{\text{NegBilog}} = 22.9$), and at Fehmarn-Schleswig the Coles-Tawn ($\chi^2_{\text{C-T}} = 24.2$) and Tajvidi ($\chi^2_{\text{Tajvidi}} = 27.8$) models gave the best fit. In general, the logistic and negative logistic models perform poorly, so in the followings we check how their extensions work on these data.

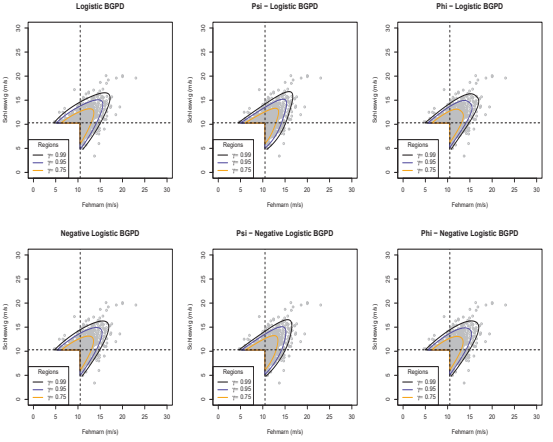

Figure 5.10: Coles-Tawn (Dirichlet), bilogistic, negative bilogistic and Tajvidi BGPD prediction regions BGPD prediction regions for Fehmarn and Schleswig.

5.4.2 New BGPD models for wind data

After applying the $\Psi-$ or $\Phi-$transformations (as in subsection 3.1.3 and subsection 3.1.4) for the logistic and negative logistic models we carried out the same procedure as for the standard BGPD models. An example of an R code (see 24), for fitting BGPD models by the `mgpd package`, is inserted below.

```
> library(mgpd)
> data(WindData)
> nms <- c( "Bremerhaven", "Fehmarn")
```

```
>
> ## Threshold selection and initial values for optimization
> demodata  <- WindData[ , nms]
> thr       <- apply( demodata, 2, quantile, prob=0.95)
> potdata   <- mgpd_data( demodata, thr=thr)
> init      <- mgpd_init( potdata)
>
> ## Model fitting
> est.log   <- fbgpd( c(init,2), dat=potdata[ , 1:2], model="log"
+ , fixed=FALSE)
> est.log$value

[1] 5137.405

> ##Not run: est.psilog <- fbgpd( c(est.log$par[1:7], 0, 1.8)
> ##Not run: , dat=potdata[,1:2], model="psilog", fixed=FALSE)
> ##Not run: est.philog <- fbgpd( c(est.log$par[1:7], -0.001, 2)
> ##Not run: , dat=potdata[ , 1:2], model="philog", fixed=FALSE)
```

The effect of newly introduced asymmetry parameters on the spectral densities is shown in Figure 5.11. The prediction regions of new BGPD models compared with the baseline models are displayed in Figure 5.12 for Fehmarn-Schleswig. Here we can see how the new dependence models change the BGPD density curves. An example of an R code (see 24), for drawing prediction regions by the `mgpd package`, is the following.

```
>
> ## Computing prediction regions
> x <- seq(-12, 15,  0.05  )
> y <- seq( -6, 15,  0.05  )
> z <- outer(  x,  y,  dbgpd,  model="log"
+ , mar1 = est.log$par[1:3], mar2 = est.log$par[4:6]
+ , dep = est.log$par[7] )
```

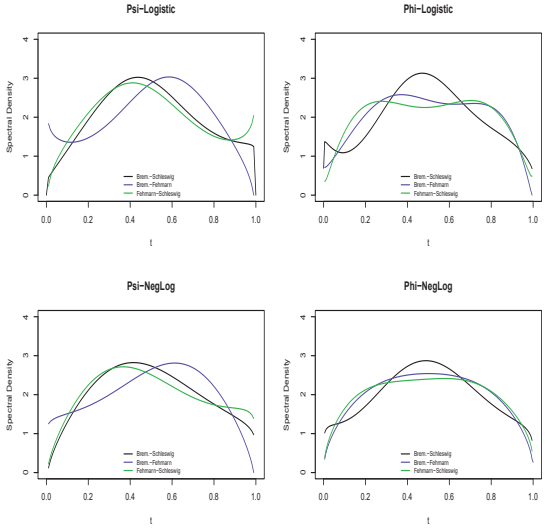

Figure 5.11: Fitted spectral densities for the 3 pairs of sites.

```
> reg <- dbgpd_region( x,  y,  z )
>
> ## Plot data VS Prediction regions
> plot(potdata[,1]+thr[1], potdata[,2]+thr[2], cex=0.7, col="gray"
+ , main="Logistic BGPD Regions",
+ , xlab=paste(nms[1]," (m/s)",sep="")
+ , ylab=paste(nms[2]," (m/s)",sep="")
+ , xlim=c(0,30), ylim=c(0,30)  )
> contour(  reg$x+thr[1],  reg$y+thr[2], reg$z, levels=reg$q,
+ drawlabels=FALSE, add=TRUE, col=c(1,4,"orange"))
> abline( h=thr[2],  v=thr[1],  lty=2 )
> legend( "bottomleft", c(expression(gamma==0.95)
```

```
+ , expression(gamma==0.9), expression(gamma==0.75))
+ , lty=1, col=c(1,4,"orange"), title="Regions")
```

The expected values and number of observations in between the neighboring regions can be seen in Table 5.10. The performance of the different models shows mixed results, but in some cases the improvement is obvious. E.g. at Bremerhaven-Schleswig the $\Psi-$logistic model is consequently closer to the expectation than its baseline (logistic), as well the $\Psi-$negative logistic model at Bremerhaven-Fehmarn.

In order to have a more comprehensive picture about the fit, the estimates and the χ^2 statistics have been summarized in Table 5.11. As we expected, the logistic improved a lot due to the $\Psi-$transformation at Bremerhaven-Schleswig, from $\chi^2_{\text{Log}} = 17.8$ to $\chi^2_{\Psi-\text{Log}} = 11.9$ (so performed much better than any of the standard ones). At Bremerhaven-Fehmarn the $\Psi-$negative logistic BGPD model shows the most significant improvement, as $\chi^2_{\Psi-\text{Log}} = 13.7$, but here also the Φ-logistic BGPD improved substantially, from $\chi^2_{\text{Log}} = 44.9$ to $\chi^2_{\Phi-\text{Log}} = 21.6$. Finally at Fehmarn-Schleswig the most substantial reduction in the test-statistics is observed when using Φ-transformation for the logistic model whereas the Ψ-transformation worked better for the negative logistic model, which is actually the overall best with having the smallest $\chi^2_{\Psi-\text{NegLog}} = 23.6$ statistic. As a conclusion, we can say that the proposed transformations (Ψ or Φ) improve the models (logistic or negative logistic) in general, and beyond that the most improved models turn out to be one of the overall best (including the standard ones as well) for any pairs of stations. Further more technical advantage is that the new models are significantly easier (and so much faster) to compute than e.g. the bilogistic and negative bilogistic models.

In order to estimate the uncertainty of the above parameters a bootstrap simulation study has been carried out. The wind data has been bootstrapped 100 times and for all bootstrap samples the various BGPD models have been re-estimated. A typical example for the results is shown in Table 5.12, which has been calculated for the pair Fehmarn-Schleswig. Generally, we found all the standard errors of the estimates to be reasonable comparing with those we got from the simulation study in Table 4.6.

Table 5.10: Number of observations between the $\gamma = 0.99, 0.95, 0.9, 0.75, 0.5, 0.25$ prediction regions (see also Figure 5.12 for Fehmarn-Schleswig)

Goodness-of-fit tables for χ^2 statistics						
Range	1-0.95	0.99-0.95	0.95-0.75	0.75-0.5	0.5-0.25	0.25-0
Bremerhaven and Schleswig						
Expected	12.2	49	245	306	306	306
Log	14	40	265	353	289	263
$\Psi-$L	12	43	263	343	296	267
$\Phi-$L	9	43	257	359	314	242
NegLog	22	38	254	344	286	280
$\Psi-$NegL	20	39	256	344	288	277
$\Phi-$NegL	18	35	265	366	289	251
Bremerhaven and Fehmarn						
Expected	12.6	50.3	251.4	314.3	314.3	314.3
Log	10	28	292	343	351	233
$\Psi-$L	10	39	286	344	332	246
$\Phi-$L	11	49	311	319	297	270
NegLog	11	38	281	335	346	246
$\Psi-$NegL	11	38	268	338	333	269
$\Phi-$NegL	12	47	291	321	341	245
Fehmarn and Schleswig						
Expected	12.5	49.8	249.2	311.5	311.5	311.5
Log	10	34	259	379	333	231
$\Psi-$L	12	37	257	365	349	226
$\Phi-$L	12	40	272	365	311	246
NegLog	11	37	261	370	323	244
$\Psi-$NegL	13	40	255	356	335	247
$\Phi-$NegL	16	37	260	363	319	251

Table 5.11: Estimates of dependence models, log-likelihood values and χ^2 statistics for the fitted BGPD models.

Performance of the model extensions						
Brem.-Schl.	Log	Ψ−Log	Φ−Log	NegLog	Ψ−NegL	Φ−NegL
Dep.	2.06	2.04	2.07	1.34	1.32	1.30
Asy.I	-	0.21	-0.003	-	0.20	-0.004
Asy.II	-	1.74	2.63	-	1.95	2.13
-Log.lik.	4872.1	4858.9	4871.1	4870.8	4861.3	4869.5
χ^2	17.77	11.85	24.97	18.82	15.98	30.95
Brem.-Fehm.	Log	Ψ−Log	Φ−Log	NegLog	Ψ−NegL	Φ−NegL
Dep.	1.98	1.98	1.96	1.26	1.29	1.26
Asy.I	-	-0.15	0.01	-	-0.26	0.002
Asy.II	-	1.22	1.80	-	1.91	2.00
-Log.lik.	5137.4	5115.8	5122.5	5124.9	5113.0	5122.9
χ^2	44.89	26.46	21.61	26.08	13.72	24.16
Fehm.-Schl.	Log	Ψ−Log	Φ−Log	NegLog	Ψ−NegL	Φ−NegL
Dep.	1.95	1.97	1.91	1.22	1.23	1.21
Asy.I	-	0.20	0.01	-	0.32	0.003
Asy.II	-	1.66	2.04	-	2.00	2.18
-Log.lik.	4636.1	4624.4	4617.6	4622.8	4612.7	4618.8
χ^2	42.82	40.74	27.01	30.08	23.59	25.23

Table 5.12: Fehmarn-Schleswing: Bootstrap estimates and standard errors for the fitted models.

Fehmarn-Schleswig: Bootstrap estimates and standard errors						
Model	Log	Ψ−Log	Φ−Log	NegLog	Ψ−NegL	Φ−NegL
Dep.	1.95(0.04)	2.07(0.13)	1.926(0.045)	1.22(0.04)	1.23(0.05)	1.218(0.008)
Asy.I	–	0.23(0.13)	0.009(0.002)	–	0.33(0.18)	0.002(0.001)
Asy.II	–	2.01(0.27)	2.130(0.182)	–	2.1(0.92)	2.635(0.026)

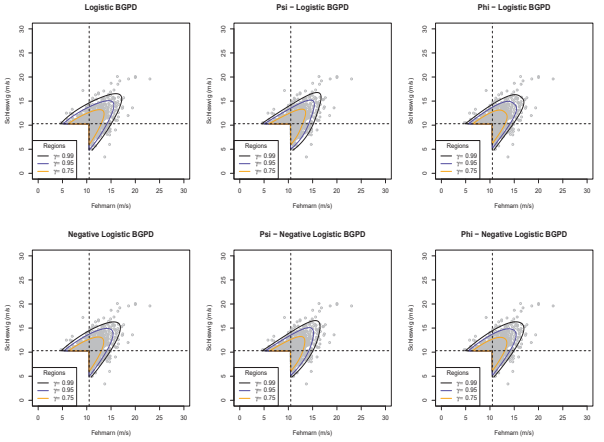

Figure 5.12: BGPD prediction regions for Fehmarn and Schleswig.

5.4.3 New TGPD models for wind speed

Based on the results of the previous section, it is obvious that there is relevant improvement achieved by using an appropriately transformed asymmetric model. It is a further interesting question whether similar considerations can be successfully applied in the trivariate case as well. Another reason is that we have seen in section 3.2 that there are not too many possibilities available to choose the dependence structure from if $d > 2$, specially if asymmetry is required. Here we attempt to apply the extension of the $\Psi-$transformation from subsection 3.1.3 for modeling 3 dimensional wind data. So now, instead of considering pairs, we focus on triplets of observations from the same three sites. The data, trivariate exceedances at Bremerhaven-Schleswig-Fehmarn, are plotted in Figure 5.13. We should note that this time the plots of the bivariate margins contain observations in the lower left quarter plane as well, in contrast with the previous section. The observations of this quarter are the ones which are actually below the two corresponding marginal

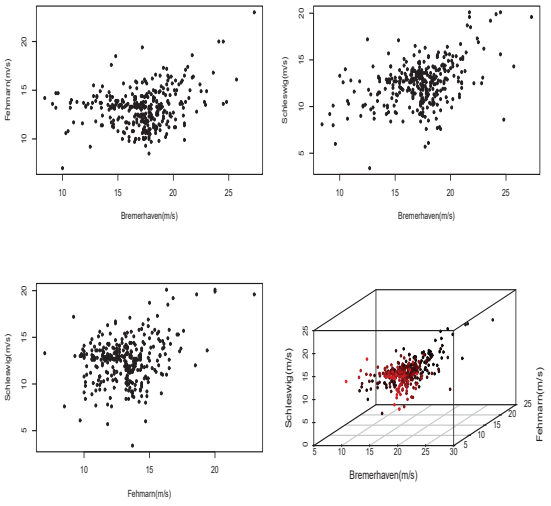

Figure 5.13: Threshold exceedances for triplets of observations at Bremerhaven-Schleswig-Fehmarn.

thresholds, but over the third one. In fact, the three thresholds divide the space into 8 partitions and only one of these partitions contains observations under the threshold in all components. We intend to estimate TGPD which describes the behavior of the exceedances in the remaining 7 partitions of the 3D space.

Technical issues

The multivariate symmetric logistic and negative logistic models are clearly not flexible enough having one single parameter only, not even for BGPD as we have already seen. Dirichlet model (in Equation 3.7), pairwise beta (in Equation 3.9) and their further extensions by Ballani and Schlather (2011)are more flexible options, but unfortunately in their case full-likelihood considerations are very difficult to carry

out as there is no explicit formula known for their exponent measures (only for their spectral densities). We recall that in the MGPD density there is the constant factor $\frac{-1}{\log G(0,\dots,0)} = \frac{1}{V(0,\dots,0)}$ present, so basically we need to have the exponent measure and the spectral density at the same time for the maximum likelihood estimation.

Remark 25. *Estimates of the Dirichlet parameters are not unattainable though, as using bivariate BGPD densities for pairwise composite likelihood method (see subsection 4.1.2) instead of the full-likelihood estimation might work in this case. An attempt on the pairwise likelihood for $d = 5$ is shown later in Table 6.1.*

Our suggestion is using one of the model extensions described in subsection 3.2.2. By doing this there is no need of numerical approximation of the exponent measure as both the exponent measure and the spectral density is given explicitly. Basically, the only nontrivial criteria one must take care of is convexity. Checking the second order condition in Equation 3.10 might be done relatively fast by Cholesky decomposition of the Hessian matrix. In our typical example in subsection 3.2.2 the set of asymmetry parameters is 4 dimensional as there are $\psi_{1,1}, \psi_{1,2}, \psi_{2,1}, \psi_{2,2}$ extra parameters included. The difficulty due to the unknown set of valid asymmetry parameters verifying convexity constraints can be tackled by allowing the entire $\mathbb{R}^4$ for $\psi_{1,1}, \psi_{1,2}, \psi_{2,1}, \psi_{2,2}$ during the maximum likelihood optimization and rejecting the invalid parameters during the optimization process by penalizing when necessary. Technically, in the `mgpd` package the likelihood is set to be infinite for the above invalid cases. Another important technical problem is how to choose an appropriate initial value for the maximization. Our general finding was that an arbitrary initial value for the Ψ-logistic model rarely ended up with higher likelihood value than likelihood of the baseline. In order to avoid the algorithm to get stuck at local maxima we first estimated the baseline model without the extra parameters and then used the results as initial values for the extended model. However often finding the proper initial values for the baseline model optimization can be difficult.

Remark 26. *In the* `mgpd` *there is an option for using least square estimation (LSQ) by minimizing the distance between the empirical and fitted distribution functions in the observation points. The LSQ estimates might also been used as initial values.*

The fitted models

In this section we apply the new parametric families of trivariate generalized Pareto distribution (TGPD) to the same three sites, investigated in the bivariate case. An example of an R code (see 24), for fitting TGPD models by the mgpd package, is inserted below.

```
> nms             <- c("Bremerhaven", "Fehmarn", "Schleswig")
> demodata        <- WindData[ , nms ]
> thr             <- apply( demodata, 2, quantile, prob=0.99 )
> potdata         <- mgpd_data( demodata, thr=thr )
> init            <- mgpd_init( potdata[,1:3])
> est.log         <- optim( c(init, 1.5), ml3_log
+ , dat=potdata[,1:3], silent=TRUE)
> est.log$value

[1] 2074.082

> est.psilog <- optim( c(est.log$par[1:10], 0, 0, 1, 1)
+ , ml3_psilog, dat=potdata[,1:3], checkconv=F, silent=TRUE)
```

Maximum likelihood estimates of the dependence models are summarized in Table 5.13, where the last row shows the (minus) loglikelihood values. According to likelihood ratio test both Ψ-models are significantly better then their baselines, e.g. for logistic $\chi^2_{\text{Log.}} = 2(74.1 - 68) = 12.2$ and for negative logistic $\chi^2_{\text{NegLog.}} = 2(47.9 - 39) = 17.8$, so the p-values are 0.016 and 0.001 with $14 - 10 = 4$ degrees of freedom, respectively. There has also been a bootstrap simulation performed for computing the standard error of the estimates. The 3D wind data has been bootstrapped 100 times and for all bootstrap samples the TGPD models have been re-estimated. A typical example for the results is shown in Table 5.14, which has been calculated for $\Psi-$logistic model. Generally, we found the standard errors of the estimates to be small, showing that the extra parameters are reasonable choice for modeling. Finally, the effect of the asymmetry parameters are very conspicuous, see Figure 5.14. The differences between the density functions of the baseline and

Table 5.13: Estimates and loglikelihood values of TGPD models fitted to Bremerhaven-Fehmarn-Schleswig.

TGPD model at Bremerhaven-Fehmarn-Schleswig				
Model	Log	$\Psi-$Log	NegLog	$\Psi-$NegLog
α	1.991	2.012	0.768	0.766
$\psi_{1,1}$	0.000	-0.023	0.000	-0.001
$\psi_{1,2}$	0.000	0.007	0.000	-0.006
$\psi_{2,1}$	0.000	0.899	0.000	1.052
$\psi_{2,2}$	0.000	1.268	0.000	1.025
-Log.lik	2074.1	2068.0	2147.9	2139.0

Table 5.14: $\Psi-$logistic TGPD model at Bremerhaven-Fehmarn-Schleswig: Estimates and standard errors by bootstrap.

Bootstrap for $\Psi-$logistic TGPD model				
α	1.991	1.983 (0.033)	2.012	2.004 (0.038)
$\psi_{1,1}$	0.000	0.000 (0.000)	-0.023	-0.023 (0.002)
$\psi_{1,2}$	0.000	0.000 (0.000)	0.007	0.003 (0.003)
$\psi_{2,1}$	0.000	0.000 (0.000)	0.899	0.952 (0.041)
$\psi_{2,2}$	0.000	0.000 (0.000)	1.268	0.988 (0.045)

the Ψ-model, we got for the wind speed data, can be seen in the right panel. An example of an R code (see 24), for comparing TGPD density functions by the mgpd package, is inserted below.

```
> dens.logfixz<-function( x, y,
+ param=c( 0, 1, 0, 0, 1, 0, 0, 1, 1, 2),
+ fixz=1, ...)
+ {
+ fixz <- rep(fixz, length(x))
+ logmod <- dtgpd_psilog(x, y, fixz
+ , mar1 = param[1:3], mar2= param[4:6]
+ , mar3 = param[7:9], dep = param[10]
+ , checkconv=FALSE)
+ logmod
+ }
> dens.psilogfixz <- function( x, y
+ , param=c(0, 1, 0, 0, 1, 0, 0, 1, 1, 2, 0, 0, 2, 2)
+ , fixz=1,...)
+ {
+ fixz <- rep( fixz, length(x))
+ psilogmod <- dtgpd_psilog(x, y, fixz
+ ,mar1 = param[1:3], mar2= param[4:6]
+ ,mar3 = param[7:9], dep = param[10]
+ ,A1   = param[11] , A2  = param[12]
+ ,B1   = param[13] , B2  = param[14]
+ , checkconv=FALSE)
+ psilogmod
+ }
> xx  <- yy <- seq( -6, 10, .1)
> zz0 <- outer(xx, yy, dens.logfixz
+ , param=est.log$par, fixz=-0.5)
> zz  <- outer(xx, yy, dens.psilogfixz
```

```
+ , param=c(est.log$par,-0.02,0.01,1,1.3), fixz=-0.5)
> image.plot(xx+thr[1],yy+thr[2],zz
+ ,main=paste("Psi-Log TGPD density, ",nms[3],": ",-0.5+thr[3]," m/s",sep="")
+ ,xlab=paste(nms[1], "m/s",sep=""),ylab=paste(nms[2], "m/s",sep=""))
> abline(h=thr[2],lty=2)
> abline(v=thr[1],lty=2)
> image.plot(xx+thr[1],yy+thr[2],zz0-zz
+ ,main=paste("Differece between Log and Psi-Log TGPD")
+ ,xlab=paste(nms[1], "m/s",sep=""),ylab=paste(nms[2], "m/s",sep=""))
> abline(h=thr[2],lty=2)
> abline(v=thr[1],lty=2)
```

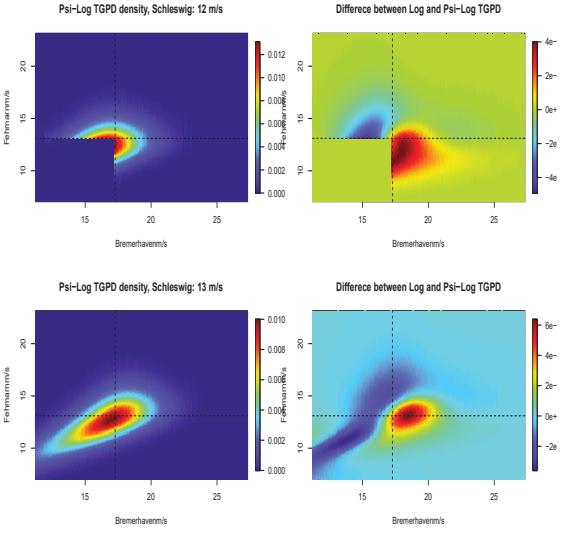

Figure 5.14: On the left: The conditional bivariate density function of logistic TGPD (the value of third variable is fixed), Right panel: differences between TGPD densities with logistic and $\Psi-$logistic dependence structures fitted to wind data.

Chapter 6

Future Objectives

Here I sketch some ideas for such further research objectives, in which I have already some preliminary results, but they are not yet completely elaborated. Significant increase in the number of sites is still a very challenging problem. In the followings I discuss some initial attempts to solve it, at least for our case.

6.1 Fitting MGPD in dimension 5

Finally here we attempt to fit a model including all sites. As full likelihood estimation would be very complicated for most known models here we propose the use of the pairwise composite likelihood method (see subsection 4.1.2). The most suitable parametric dependence model for fitting MGPD this way is the Dirichlet model in Equation 3.7 because the d-dimensional joint distribution can be built up using the 2 dimensional margins due to the property presented in remark 18. For likelihood optimization we have used the mean of the BGPD estimates computed separately, see the first column of Table 6.1, and for the composite likelihood estimates see the second column of the table. The mean and standard error of bootstrap estimates computed by 100 repetitions can be seen in the last column. The original and bootstrap estimates are reasonably close to each other, and the standard errors are rather small. Further inference on the fitted model as well as application of other similar models as e.g. in Equation 3.9 are among the future research goals. An ex-

Table 6.1: Composite likelihood estimates for Hamburg-Hannover-Bremerhaven-Fehmarn-Schleswig.

Dirichlet MGPD parameters			
Loc.	Init.	Estim.	Bootstrap
Hamburg	1.16	1.18	1.21 (0.12)
Hannover	1.05	0.81	0.85 (0.07)
Bremerhaven	1.55	1.56	1.61 (0.12)
Fehmarn	1.78	1.32	1.38 (0.14)
Schleswig	1.12	1.37	1.40 (0.11)

ample of an R code (see 24), for fitting Dirichlet model in 5D by `mgpd package`, is inserted below.

```
> nms             <- names( WindData[, 2:6])
> demodata        <- WindData[, 2:6]
> thr             <- apply( demodata, 2, quantile, prob=0.95)
>
> ## pairwise composite loglikelihood
> cml_ct  <- function(param,mlmax=1e+15,...)
+ {
+ ndim    <- length(param)/4
+ param   <- matrix(param,nrow=ndim,byrow=T)
+ complik <- rep(mlmax,choose(ndim,2))
+ counter <- 0
+
+ for(i in 1:ndim){for(j in 1:ndim){if(i<j){
+ lik      <- NULL
+ counter  <- counter+1
+ xdat     <- sample_ij[[paste("sample", i, j, sep="")]]
+ lik      <- try(dbgpd(xdat[,1], xdat[,2], model = "ct"
+ , mar1 = param[i,1:3], mar2 = param[j,1:3]
```

```
+ , a    = param[i,4]   , b    = param[j,4]))
+     if(!is.null(lik))
+     {
+     complik[counter] <- -sum( log( lik))
+     if( min( 1+param[i, 3]*( xdat[, 1]-param[i, 1])
+     /param[i, 2])<0) complik[counter] <- mlmax
+     if( min( 1+param[j, 3]*( xdat[, 2]-param[j, 1])
+     /param[j, 2])<0) complik[counter] <- mlmax
+     }
+ }}}
+ sum(complik)
+ }
>
> ## optimization from random initial value
> est.cml <- optim(rep(c(0,1.5,0.1,1.4),5),cml_ct)
> est.cml <- as.data.frame(matrix(est.cml$par
> , nrow= length(est.cml$par)/4,byrow=T))
> names(est.cml) <- c("loc","scale","shape","dep")
> row.names(est.cml) <- nms
```

6.2 MGPD on a grid

A useful approach for modeling exceedances on a grid can be the use of the so called generalized Pareto process (see Ferreira and de Haan, 2012). This can be basically considered as the infinite dimensional extension of GPD in Equation 2.4. Indeed, the relationship between GEV and GPD remains valid for infinite dimensional versions as well, meaning that any Pareto process is in the domain of attraction of a max-stable process with the same spectral measure and vice versa (see Corollary 4.1,4.2 and 4.3 in Ferreira and de Haan, 2012). The following method that of Davison et al. (2011) for spatial exceedances look very promising. For this instead of fitting pairwise BEVD for maxima we may attempt to fit BGPD for bivariate ex-

ceedances. With the help of `SpatialExtremes` `R`-package (Ribatet, 2012) it is easy to simulate samples from most of well-known parametric max-stable processes. As due to Theorem 13 the Schlather process does not lead to absolutely continuous BGPD margins for the exceedances, for applications the Smith process is more recommendable (see Theorem 12). Therefore I am about to develop a new option of the `mgpd` `R`-package which is also capable to fit BGPD models having Smith type dependence. As a further step for the future I plan to use this routine to reveal the dependence structure of the process using randomly chosen bivariate samples.

Bibliography

[1] Ballani, F. and Schlather M. (2011) A construction principle for multivariate extreme value distributions. *Biometrika*, **98**(3), p.633-645.

[2] Balkema, A. A. and de Haan, L. (1974) Residual lifetime at great age. *Ann. Probab.*, **2**, p.792-804.

[3] Balkema, A. A. and Resnick, S. I. (1977) Max-infinite divisibility. *Journal of Applied Probability*, **14**, p.309319.

[4] Barnett, V. (1976) The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society*, Series A, **139**, p.318354.

[5] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*, Wiley Series in Probability and Statistcs

[6] Cooley, D., Davis R. A. and Naveau, P. (2010) The pairwise beta distribution: a flxibile parametric multivariate model for extremes. *Journal of Multivariate Anal.*, **101**, p.2103-17.

[7] Davison, A. C., Padoan, S. A. and Ribatet, M. (2011) Statistical Modelling of Spatial Extremes. (http://homes.stat.unipd.it/padoan/?file=spatrev12.pdf)

[8] Cherubini, U. Luciano, E. and Vecchiato, W. (2004) *Copula methods in Finance*, WileyFinance, West Sussex, England.

[9] Coles, S. and Tawn, J. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society*, Series B, **53**, p.377-392.

[10] Coles, S. and Tawn, J. (1994) Statistical Methods for Multivariate Extremes: an Application to Structural Design. *Appl. Statistics*, **43** p.1-48.

[11] Coles, S. (2001) *An Introduction to Statistical Modelling of Extreme Values*, Springer-Verlag.

[12] Cox, D. R. and Reid, N. (2004) Miscellanea A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**(3), p.729737.

[13] de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction*, Springer-Verlag

[14] de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. Probab.*, **12**, p.1194-1204

[15] de Haan, L. and Pickands, J. (1986) Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, **72**, p.477492.

[16] de Haan, L. and Resnick, S. I. (1987) On regular variation of probability densities. *Stochastic Processes and Their Applications*, **25**, p.83-93

[17] Eastoe, E. F. and Tawn, J. A. (2009) Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society*, Series C, **58**, p.22-45.

[18] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal events for Insurance and Finance*, Springer-Verlag

[19] Falk, M. and Reiss, R. D. (2005) On Pickands coordinates in arbitrary dimensions. *Journal of Multivariate Anal.*, **92**, p.426-453.

[20] Finkenstadt, B. and Rootzén, H. (2004)*Extreme Values in Finance, Telecommunications and the Environment*, Chapman & Hall

[21] Fisher, R. A. and Tippet, L. H. C. (1928)Limiting forms of the frequence distribution of the largest or smallest member of a sample. *Proc. Cambridge Phil. Soc.*, **24**, p.180-190.

[22] Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics*, Krieger

[23] Ghoudi, K., Khoudrai, A. and Rivest, L. P. (1998) Statistical properties of couples of bivariate extreme-value copulas. *Canad. J. Statist.*, **26**, p.187-197.

[24] Genest, C. Quessy, J.-F. and Rémillard, B. (2006) Goodnes-of-fit Procedures for Copula Models Based on the Integral Probability Transformation. *Scandinavian J. of Statistics*, **33**, p.337-366.

[25] Gnedenko, B. V. (1943) Sur la distribution limite du terme maximum d'une serie aleatoire, *Annals of Mathematics*, **44**, p.423-453.

[26] Godambe, V. P. (1960) An optimum property of regular maximum likelihood equation. *Ann. Math. Statist.*, **31**, p.1208-1211.

[27] Gumbel (1958) *Statistics of Extremes*, Columbia University Press

[28] Hall, P. and Tajvidi, N. (2000) Distribution and dependence function estimation for bivariate extreme value distributions, Bernoulli 6(5), p.835-844.

[29] Hall, P. and Tajvidi, N. (2004) Prediction regions for bivariate extreme events. *Aust. N. Z. J. Stat.* **46**(1), p.99-112.

[30] Joe, H. (1990) Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Letters*, **9**, p.75-81.

[31] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, London: Chapman & Hall.

[32] Klüppelberg, C. and May, A. (1999) The dependence function for bivariate extereme value distributions - systematic approach.

[33] Kojadinovic, I., Yan, J. and Holmes, M. (2011) Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica*, **21**, p.841871.

[34] Kojadinovic, I. and Yan, J. (2011) A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing* **21**, p.1730.

[35] Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions*, Imperial College Press, London.

[36] Krusper (2011) Többdimenziós extrém érték eloszlások alkalmazása biztosítási és pénzügyi adatokra. *MSc thesis*, Eötvös Loránd University, Budapest (in Hungarian)
(http://www.cs.elte.hu/blobs/diplomamunkak/mat/2011/krusper_marta.pdf)

[37] Leadbetter M. R., Lindgren G. and Rootzén H. (1983) *Extremes and related properties of stationary sequences and processes*, Springer, New York

[38] Lindsay, B. G. (1988) Composite likelihood methods. *Contemporary Mathematics*, **80**, p.220241.

[39] McNeil, A. J., Frey R. and Embrechts P. (2005) *Quantitative Risk Management*, Princeton University Press

[40] Marshall and Olkin (1983) Domains of attraction of multivariate extreme value distributions. *Ann. Probab.*, **11**, p.168-177.

[41] Michel, R. (2006) Simulation and Estimation in Multivariate Generalized Pareto Models. *PhD thesis*, University of Wuerzburg, Germany

[42] Nadarajah, S. (1999) Simulation of multivariate extreme values. *J. Statist. Comput. Simulation*, **62**, p.395-410.

[43] Nelsen, R. B. (2006) *An Introduction to Copulas*, 2nd ed. John Wiley & Sons.

[44] Oakes, D. and Manatunga, A. K. (1992) Fisher information for bivariat extreme value distribution. *Biometrika*, **79**, p.827-32.

[45] Pickands, J. (1975) Statistical inference using extreme order statistics. *Anals of Statistics*, **3**, p.119-131.

[46] Pickands, J. (1981) Multivariate extreme value distributions. *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session*: p.859-878.

[47] Pickands, J. (1989) Multivariate negative exponential and extreme value distributions, *In Extreme Value Theory: Proceedings*, Oberwolfach, p.262-274.

[48] Petrickova, A (2012) Modeling of the residuals of ARMA models using autocopulas *Forum Statisticum Slovacum, ISSN 1336-7420* **1**, p.49-55.

[49] Rakonczai, P. and Turkman, F. (2012) Applications of generalized Pareto processes. (Technical report under progress, OTKA outgoing mobility grant, Lisbon, Portugal)

[50] Rakonczai, P. (2012) Asymmetric dependence models for bivariate threshold exceedance models. *Forum Statisticum Slovacum, ISSN 1336-7420* **1**, p.25-32.

[51] Rakonczai, P. and Zempléni, A. (2012) Bivariate generalized Pareto distribution in practice: models and estimation. *Environmetrics*, John Wiley & Sons, **23**, p.219-227.

[52] Rakonczai, P., Márkus, L. and Zempléni, A. (2011) Autocopulas: investigating the interdependence structure of stationary time series. *Methodology and Computing in Applied Probability*, **14**, p.149-167.

[53] Rakonczai, P. (2011) `Package 'mgpd' manual`.
see http://cran.r-project.org/web/packages/mgpd/mgpd.pdf

[54] Rakonczai, P. and Tajvidi, N. (2010) On prediction of bivariate extremes. *International Journal of Intelligent Technologies and Applied Statistics*, **3**(2), p.115-139.

[55] Rakonczai, P., Butler, A. and Zempléni, A. (2010) Modeling temporal trend within bivariate generalized Pareto models of logistic type.
(Technical report, HPC-Europa2 Project, Edinburgh, UK, available at http://www.math.elte.hu/∼paulo/pdf/)

[56] Rakonczai, P. (2009) On Modelling and Prediction of Multivariate Extremes, with applications to environmental data. Centrum Scientiarum Mathematicarum, Licentiate Theses in Mathematical Sciences 2009:05

[57] Rakonczai, P., Márkus L. and Zempléni, A. (2008a) Goodness of Fit for Auto-Copulas: Testing the Adequacy of Time Series Models, *Proceedings of the 4th International Workshop in Applied Probability* CD-ROM, paper No.73., 6 pages, Compiegne, France

[58] akonczai, P., Márkus L. and Zempléni, A. (2008b) Adequacy of Time Series Models, Tested by Goodness of Fit for Auto-Copulas, *Proceedings of the COMPSTAT2008 conference*, Porto, Portugal

[59] Rakonczai, P. and Zempléni, A. (2007) Copulas and goodness of fit tests. *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific, Hackensack, NJ, p.198-206.

[60] Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag

[61] Reiss, R. D. and Thomas, M. (2007) *Statistical Analysis of Extreme Values*, Birkhauser

[62] Ribatet, M. (2012) `Package 'SpatialExtremes' manual`, see http://cran.r-project.org/web/packages/SpatialExtremes/SpatialExtremes.pdf

[63] Rootzén, H. and Tajvidi, N. (2006) The multivariate generalized Pareto distribution. *Bernoulli* **12**, p.917-930.

[64] Shi, D., Smith, R. L. and Coles, S. G. (1993) Joint versus marginal estimation for bivariate extremes. *Technical Report*

[65] Shi, D. (1995) Fisher Information for a Multivariate Extreme Value Distribution. *Biometrika*, **82**(3), p.644-649.

[66] Sibuya, M. (1960) Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, p.195210.

[67] Sklar, A. (1959) Fonctions de répartition á n dimensions et leurs marges. *Publications de lInstitut de Statistique de lUniversite de Paris'*, **8**, p.229231.

[68] Smith, R. L. (1985) Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72**, p.67-90.

[69] Smith, R. L. (1990) Extreme value theory. *In Handbook of Applicable Mathematics* (ed. W. Ledermann), **7**, p.437-471.

[70] Smith, R. L. (1994) Multivariate Threshold Methods. *In Extreme Value Theory and Applications* (ed.J. Galambos), Kluwer, p.225-248.

[71] Stephenson, A. G. (2002) `evd`*: Extreme Value Distributions.* `R News`, **2**(2), p.31-32.

[72] Tajvidi, N. (1996) Multivariate generalized Pareto distributions. Article in PhD thesis, Department of Mathematics, Chalmers, Göteborg.

[73] Tawn, J. A. (1988) Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**(3), p.397-415.

[74] Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, p.245253.

[75] Tiego de Oliveira (1962/63) The asymptotic independence of the sample mean and extremes, *Rev. Fac. Sci. Lisboa A*, **8**, p.299-309.

[76] Vatan, P. (1985) Max-infinite divisibility and max-stability in infinite dimensions. *Probability in Banach Spaces V.*, ed. A. Beck et al. Lecture Notes in Mathematics 1153, Springer, Berlin, p.400-425.

[77] Varin, C. (2008) On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**, p.128.

[78] Yee, T. W. and Stephenson, A. G. (2007) Vector generalized linear and additive extreme value models. *Extremes*, **10**, p.1-19.

[79] Zempléni, A. and Rakonczai, P. (2011) New bivariate threshold models with hydrological applications. *Conference on Environmental Risk and Extreme Events*, Ascona, July 10-15