Multivariate generalized Pareto distribution in practice: models and estimation

December 6, 2010

Pál Rakonczai
Probability Theory and Statistics Department
Eotvos University, Budapest, Hungary
&
Department of Mathematical Statistics
Lund Institute of Technology, Sweden
e-mail: pal@maths.lth.se

Contents

1 Introduction 3

2 Models 4
  2.1 Univariate extreme value models . . . . . . . . . . . . . . . . . . . 4
  2.2 Multivariate extreme value models . . . . . . . . . . . . . . . . . . . 5
  2.3 Multivariate threshold exceedance models . . . . . . . . . . . . . . 6
  2.4 Parametric families of dependence structures . . . . . . . . . . . . . 8
  2.5 Construction of new non-exchangeable models . . . . . . . . . . . . 12

3 Statistical inference on wind speed data 15
  3.1 Computational aspects of estimation and prediction . . . . . . . . . . 15
  3.2 Results for bivariate observations . . . . . . . . . . . . . . . . . . . 15
  3.3 Results for trivariate observations . . . . . . . . . . . . . . . . . . . 18

4 Discussion 18

A Appendix: "classical BGPD" models fitted by 'fbvpot' 26
Abstract

Extreme values are of substantial interest in fields as diverse as finance, environmental science and engineering, because they are associated with rare but hazardous events (such as flooding, mechanical failure or severe financial loss). There is often interest in understanding how the extremes of two different processes are related to each other, and, in particular, in estimating the probability that the processes will both simultaneously become extreme. Analyses of the extremes of two variables are typically based on fitting a specific parametric subfamily of the multivariate extreme value distribution to component-wise maxima. An alternative approach, which potentially uses the available data much more efficiently, involves fitting the multivariate generalised Pareto distribution (MGPD) to data that exceed a suitably high threshold. There are several (non-equivalent) definition of MGPD in use, one is for exceedances over a threshold in at least one of the components (Rootzén and Tajvidi, 2006) and another(s) for exceedances which are over the threshold in all components. Here we investigate the applicability of the first definition assuming different underlying dependence models and compare the performance of the two substantially different type of MGPD models. Although the dependence models are intensively used in different extreme value models, to the best of our knowledge none of these models have been discussed and applied for data in the new MGPD framework apart from the bivariate (symmetric) logistic model. As there are just a few of these families, which allow asymmetry for the different components and produce absolutely continuous models for the new MGPD case, we introduce a general transformation for creating such models from symmetric ones. We apply the proposed models to wind speed data for modeling exceedances occurring at locations in northern Germany, and outline methods for calculating prediction regions as well as evaluating goodness-of-fit.

Key words: Multivariate threshold exceedances, parametric dependence models, wind speed data.
1 Introduction

Multivariate extreme value theory involves describing the statistical properties of the extreme values of two or more processes, or, in the spatial context, of the same process at two or more sites. It is concerned, in particular, with quantifying the probability that extreme values will occur simultaneously in more than one process, or at more than one location. Such probabilities are of substantial interest in a wide range of fields, including the environmental sciences, finance and internet traffic monitoring. Even when interest ultimately lies in describing the extreme values of a single process at a single location, there will still often be value in using data from other processes or locations to improve the efficiency of estimates relating to the location and process of interest (Ribatet al., 2007). Efficiency can sometimes also be improved by utilizing data that have been collected on the same process at different temporal resolutions (e.g. daily and hourly; Nadarajah et al., 1998), and this also requires the use of a multivariate model.

A sophisticated mathematical theory to describe the characteristics of multivariate extreme events has been developed (e.g. Pickands, 1981 and Rootzén & Tajvidi, 2006) but there are substantial computational and statistical challenges in applying the resulting asymptotic models to data. Univariate extreme value methods, in contrast, are routinely used for data analysis within finance, the environmental sciences, and a range of other scientific applications (e.g. Coles, 2001). Multivariate extremes are often modeled by applying block maxima methods (BMM) to the component-wise maxima of the series being studied (e.g. Tawn, 1988). In environmental applications this commonly means focusing on the annual maximum value of each process. If the length of the block is sufficiently long, and under certain other conditions, then theoretical results suggest that the distribution of these componentwise maxima can be approximated by a multivariate extreme value distribution (MEVD). The practical application of this result typically involves selecting a particular parametric model from within the MEVD class, and then drawing statistical inferences about the parameters of this model using maximum likelihood estimation.

An important drawback of the BMM approach is the fact that it ignores information on whether or not the extremes of the different processes do actually occur simultaneously. The annual maxima of the processes may all have arisen on the same day, for example, but it is also possible that they all occurred in completely different months - the basic implementation of the MEVD model does not distinguish between these situations (although an extension of the approach does allow some information on timing to be incorporated into the modeling: Stephenson and Tawn, 2005). This problem can be avoided by modeling the sizes of all observations that exceed a given high threshold, rather than modeling the highest value within a particular block of time. This method usually uses more data than the block maxima approach, and so also leads to efficiency gains (i.e. more precise estimates for the quantities of interest).

Methods for analyzing multivariate threshold exceedances were originally developed by Tajvidi (1996), and have been further developed for more general cases in Rootzen and Tajvidi (2006). These papers have demonstrated that, under rather mild conditions and given an appropriately high threshold vector, the multivariate exceedances of a random vector over the threshold can be approximated by a multivariate general-
ized Pareto distribution (MGPD). Note that the definition of the MGPD in these papers substantially differs from the definitions which can be found in other textbooks (except the recent book of Beirlant et al. (2004), where this definition is mentioned but not discussed). The main advantage of this approach is that it includes all observations that are extreme in at least one component. Although mathematical theory for the MGPD model has been developed, the statistical properties of this model are not yet fully understood - the only paper discussing the performance of the model in practice is, to the best of our knowledge, that of Rakonczai and Tajvidi (2009). Similar BGPD models have also been used in a recent paper by Brodin and Rootzén (2009) to study wind storm losses in Sweden. Throughout this paper only stationary distributions are investigated. A non-stationary 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) is discussed by Rakonczai et al. (2010).

Standard models for univariate and multivariate extremes are presented in Section 2, as well as their role and properties in MGPD models. A new general method for constructing asymmetric versions of known symmetric models is also presented. In Section 3 we discuss the practical issues involved in estimating the parameters via maximum likelihood, and outline a procedure for construction prediction regions and model validation techniques. Additionally some useful conclusions are summarised in Section 4.

2 Models

In this section – after a short introduction to univariate extreme value models – we present the key models that are currently used for modeling maxima of multivariate observations and simultaneously discuss their applicability within models for multivariate threshold exceedances. The presented models are available in the ‘mgpd’ package of R, which has been produced for providing a tool for the statisticians to apply the proposed methods and models and for illustrating the practical applications on wind speed data.

2.1 Univariate extreme value models

Univariate extreme value theory provides the limit results for the distribution of extremes of a single process: either the maximum of observations or the distribution of exceedances of observations over a high threshold. Maxima (daily, weekly, yearly etc.) can be modeled using the generalized extreme value distribution (EVD), which has a distribution function (df.) of the form

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

These models are different from those which are available in the ‘evd’ package of R, the main differences are discussed later on.
where $1 + \frac{\xi x}{\sigma} > 0$, $\mu \in \mathbb{R}$ is called the location parameter, $\sigma > 0$ the scale parameter and $\xi \in \mathbb{R}$ the shape parameter. Similarly, threshold exceedances (e.g. over a high quantile of the observations) can be approximated by the generalized Pareto distribution (GPD), which has a cdf. of the form

$$H(x) = 1 - \left(1 + \frac{x}{\sigma}\right)^{-\frac{1}{\xi}},$$

(2)

where $1 + \frac{\xi x}{\sigma} > 0$ and $\sigma > 0$. Both of the limit distributions 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 the GPD if and only if the distribution of the maximum converges to the EVD.

### 2.2 Multivariate extreme value models

Analogously to the univariate case it can be shown if the componentwise maxima has a limit distribution, it is necessarily a multivariate extreme value distribution (MEVD). Additionally, all of its univariate margins are univariate EVD-s. There are several different (but equivalent) way of characterising the underlying dependence structure. One of these is shown by Resnick (1987) assuming unit Fréchet margins

$$G_{\text{Fréchet}}(t_1, \ldots, t_d) = \exp\left(\nu((0, t_1] \times \cdots \times [0, t_d])\right),$$

(3)

with

$$V(t_1, \ldots, t_d) = \nu((0, t_1] \times \cdots \times [0, t_d]) = \int_{S_d} \left(\frac{w_i}{t_i}\right) S(dw),$$

(4)

where $S$ is a finite measure on the $d$-dimensional simplex $S_d = \{w \in \mathbb{R}^d : |w| = 1\}$, which satisfies

$$\int_{S_d} w_i S(dw) = 1 \text{ for } i = 1, \ldots, d,$$

where $V$ and $S$ are called exponent measure and spectral measure respectively. In particular, the total mass of $S$ is always $S(S_d) = d$. If $G$ is absolutely continuous, then we can reconstruct the densities of $S$ from the derivatives of $V$. In this context it is better using "densities" instead of "density" as $S$ can have a density not only on the interior of $S_d$ but also on each of the lower-dimensional subspaces of $S_d$. E.g. if $d = 2$, the unit simplex is partitioned into two vertices and the interior of the interval or if $d = 3$ into three vertices, three edges and the interior of the triangle. Hence, even if $G$ is absolutely continuous, the spectral measure $S$ can put positive mass to the vertices; for instance, when the margins of $G$ are independent then $S(\{e_i\}) = 1$ for all $i = 1, \ldots, d$. More details about the characterization and its properties can be found in the recent textbooks as e.g. Chapter 3 in Kotz and Nadarajah (2000) or Chapter 8 in Beirlant et al. (2004).

Further relevant formulas for maximum likelihood inference are the following. Coles and Tawn (1991) found how to compute the spectral densities of all subspaces
from the partial derivatives of $V$. E.g for the interior the spectral density $s$ can be expressed as

$$
\frac{\partial V}{\partial t_1 \cdots \partial t_d} (t_1, \ldots, t_d) = -\left(\sum_{i=1}^{d} t_i \right)^{-(d+1)} s\left(\frac{t_1}{\sum_{i=1}^{d} t_i}, \ldots, \frac{t_d}{\sum_{i=1}^{d} t_i}\right).
$$

Another characterisation of MEVD, due to Pickands (1981) is possible by the so-called **dependence function**. In the bivariate setting the dependence function $A(t)$ must satisfy the following three properties, which we denote by (P).

i. $A(t)$ is convex

ii. $\max\{(1-t), t\} \leq A(t) \leq t$

iii. $A(0) = A(1) = 1$.

The lower bound in (ii.) corresponds to the complete dependence, whereas the upper bound corresponds to independence. The exponent measure and its mixed partial derivative can be written as

$$
V(t_1, t_2) = \left(\frac{1}{t_1} + \frac{1}{t_2}\right) A\left(\frac{t_1}{t_1 + t_2}\right)
$$

$$
\frac{\partial V}{\partial t_1 \partial t_2} (t_1, t_2) = \left(\frac{1}{t_1} + \frac{1}{t_2}\right) \times \left( A''(\zeta) \zeta_1' \zeta_2' + A'(\zeta) \zeta_1'' \zeta_2'' \right) - A'(\zeta) \times \left( \frac{\zeta_2'}{t_1} \zeta_1 + \frac{\zeta_1'}{t_2} \zeta_2 \right),
$$

where

$$
\zeta = \frac{t_1}{t_1 + t_2}, \zeta_1' = \frac{t_2}{(t_1 + t_2)^2}, \zeta_2' = \frac{-t_1}{(t_1 + t_2)^2}, \zeta_1'' = \frac{t_1 - t_2}{(t_1 + t_2)^3}.
$$

### 2.3 Multivariate threshold exceedance models

The multivariate extension of the GPD models has also been intensively investigated in the last decades and as a result, different definitions run up different careers. Based on its longer history, the multivariate GPD models using exactly the same construction as the MEVD model has, became more popular, we refer it as the "classical MGPD", which is also implemented in the evd package of R. It concentrates on those exceedances which are greater than the threshold in every single components. Its univariate margins are GPD distributions, which are linked by a dependence model. Using the following transformation for the GPD margins

$$
\tilde{t}_i = \hat{t}_i(x_i) = \frac{-1}{\log \frac{1}{\hat{H}_{\xi, \sigma}}(x_i)} = \frac{-1}{\log \left\{ 1 - \left( 1 + \frac{\xi}{\sigma} \frac{x_i}{\hat{t}_i} \right)^{-\frac{1}{\xi}} \right\}},
$$

where $1 + \frac{\xi}{\sigma} > 0$ and $\sigma > 0$ the MGPD is the form of

$$
\hat{H}(x_1, \ldots, x_d) = \hat{H}_{\text{Fréchet}}(\tilde{t}_1, \ldots, \tilde{t}_d) = \exp\left(-V(\tilde{t}_1, \ldots, \tilde{t}_d)\right),
$$

(5)
as regardless of the dimension of the MGPD. Further, the MGPD density can be obtained by the distribution of $H$ over if $\xi$ is interpreted as marginal parameters for the model, as the upper and lower panels of Fig. 7 where the two substantially different approaches remain in the distribution even if $1 + \xi_i(x_i - \mu_i)/\sigma_i > 0$ and $\sigma_i > 0$ as in (1) for $i = 1, ..., d$. In the following, the definition in (6) is called MGPD and (5) is called the "classical MGPD". It is important to notice that even though the dependence structure for two MGPD models in is the same, the models are substantially different. Again, the most conspicuous difference between the MGPD and "classical MGPD" is that (6) gives (parametric) probabilistic inference about the exceedances which are above the threshold at least in one component, in contrast to the another one modeling those ones, which are higher than the threshold in every components. For illustrations on real observations compare the upper and lower panels of Fig. 7 where the two substantially different approaches of constructing BGPD models are displayed assuming symmetric logistic family as dependence structure for both cases. Further difference between (6) and (5) is in parameter interpretation. Namely, that the $\xi_i, \mu_i$, and $\sigma_i$ parameters of (6) can not be any more considered as marginal parameters for the model, as the $1/\log G(0, ..., 0)$ term remains in the distribution even if $d - 1$ components of $x$ converges to infinity. Moreover if $H_1(x) = H(x, \infty)$ then $H_1$ is not a one dimensional GPD. Although it is shown in Rootzén and Tajvidi (2006) that if $X_1$ has a distribution $H_1$ then conditional distribution of $X_1 | X_1 > 0$ is GPD. Similar properties hold for all marginal distributions regardless of the dimension of the MGPD. Further, the MGPD density can be obtained as

$$h(x) = \frac{\partial H}{\partial x_1 \cdots \partial x_d}(x) = \frac{\partial}{\partial x_1 \cdots \partial x_d} \left(1 - \frac{\log G(x)}{\log G(0)}\right)$$

$$= \frac{\prod_{i=1}^d t_i'(x_i)}{V(t_1(0), \ldots, t_d(0))} \times \frac{\partial V}{\partial t_1 \cdots \partial t_d}(t_1(x_1), \ldots, t_d(x_d)),$$
Although the logarithm in (6) cancels the exponent in (3), the MGPD density $h(x)$ cannot be computed immediately from the spectral density, as the constant term $G(0)$ still contains $V$ as well. Additionally, for the MGPD models, it is reasonable to require that there will be no positive probability mass put on the boundary of the distribution, otherwise the model will not stay absolutely continuous, which is rarely realistic and causes further complications for the maximum likelihood estimation.

**Lemma 1.** Let $H$ be a MGPD represented by an absolutely continuous MEVD $G$ with spectral measure $S$. The distribution $H$ is absolutely continuous if and only if $S(S_d) = S(\text{int}(S_d)) = d$, i.e all mass is in the interior of the simplex.

**Corollary.** In such cases the MGPD decays to zero when reaching the boundaries

$$
\lim_{x_i \to x_i^+} H(x_1, \ldots, x_d) = 0 \text{ for any } x_j > 0, j \neq i,
$$

where $x_i^+ = \mu_{x_i} - \sigma_{x_i}/\xi_{x_i}$, is the finite lower endpoint of the $i$-th univariate GEV margins of the underlying MEVD model.

**Proof.** It is easy to see that the subspaces of the $S_d$ unit simplex represent the boundaries of the space containing the unit Fréchet coordinates. E.g. if $d = 2$ then the point $(w, w - 1) \in \text{int}(S_2)$ represents the $y = \frac{1-w}{w}x$ line in the unit Fréchet scale and limiting cases are $\{1, 0\}$ and $\{0, 1\}$ representing the $x$ and $y$ axes, respectively. This terminology remains the same for higher dimensions as well. Therefore when calculating the distribution function for any point of the interior on the unit Fréchet scale, the masses originated from the faces will always be cumulated, which is equivalent to causing a jump in the function and so its continuity gets broken.

In the next subsections we summarize, which specific parametric cases include such a model. As we shall see, that there are just a few such models, in section 2.5, we propose a general approach to construct asymmetric models from symmetric ones.

### 2.4 Parametric families of dependence structures

Since MGPD models are defined by the underlying MEVD model, and practically MEVD models are defined by the dependence structures (apart from margins), the most popular parametric families for MEVD models have been considered as foot-stones for the MGPD models. The most important characteristics of these models are summarized in Table 1. The list is not exclusive but covers a rather wide range of families. Although these models are intensively used in different extreme value models, to the best of our knowledge none of these models have been discussed and applied for data in the MGPD framework apart from the bivariate (symmetric) logistic model. As characterization we give both the exponent measure function and the spectral density for the models. Additionally we discuss which parameter setting allows non-exchangeability and puts all mass in the interior of $[0,1]$ providing absolutely continuous models within the BGPD framework by Lemma 1. More details about these models can be found in the respective papers indicated, giving the first appearance of the related models. Graphical illustration of the bivariate distribution function for some of the models can be found in Fig 5.
Table 1: Summary of bivariate dependence models. Asymmetric logistic, negative logistic and mixed models are considered excluding their symmetric case, because their different continuity properties. Absolutely continuous BGPD models can be obtained if all mass is in the interior of the \([0,1]\), these cases are highlighted by bold fonts. The non-exchangeable cases are the bilogistic, negative bilogistic and Dirichlet models.

Most of the bivariate cases are straightforward to extend to higher dimensions, but these extensions rarely result valid MGPD models with the full parameterisation. As an example, the multivariate logistic model provides a rich set of model parameters, but due to Lemma 1 the absolute continuity holds only in a very limited parameter setting (it coincides with the symmetric and non-exchangeable case). The multivariate negative logistic or nested logistic models behave similarly. An alternative, the multivariate extension of Coles-Tawn model (Dirichlet model) provides a flexible non-exchangeable model if such a model is needed, but it requires sophisticated numerical integration tools for its fitting, as only the spectral density is given explicitly. For the detailed description of the above models see Section 2.4.1 and 2.4.2.

2.4.1 Bivariate models

- Asymmetric logistic model (Tawn, 1988b).
  \[ V_{\text{Log}}(x,y) = (1 - \psi_1)/x + (1 - \psi_2)/y + (\psi_1/x)\alpha + (\psi_2/y)\alpha \] and \[ s_{\text{Log}}(w) = (\alpha - 1)\psi_1^\alpha \psi_2^\alpha (w(1 - w))^{\alpha - 2}((\psi_2 w)^\alpha + (\psi_1 (1 - w))^\alpha)^{1/\alpha - 1}, \]
  where \( \alpha \geq 1 \) 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, otherwise \( S(\{0\}) = 1 - \psi_2 \) and \( S(\{1\}) = 1 - \psi_1 \). (See the upper left panel of Fig 5.)

- Asymmetric negative logistic model (Joe, 1990). The negative logistic model is similar in structure to the logistic
  \[ V_{\text{Neglog}}(x,y) = 1/x + 1/y - ((\psi_1/x)\alpha + (\psi_2/y)\alpha)^{-1/\alpha} \] and \( s_{\text{Neglog}}(w) = -s_{\text{Log}}(w) \), where \( \alpha > 0 \) and \( 0 < \psi_1; \psi_2 \leq 1 \). Analogously \( S(\{0\}) = 1 - \psi_2, S(\{1\}) = 1 - \psi_1 \) and so \( \psi_1 = \psi_2 = 1 \) gives the only

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-Exchangeable</th>
<th>All mass in ( \operatorname{int}(S_d) )</th>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sym. Logistic</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Asym. Logistic</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sym. Neg. Logistic</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Asym. Neg. Logistic</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Bilogistic</td>
<td>✓</td>
<td>✓</td>
<td>not explicit</td>
</tr>
<tr>
<td>Neg. Bilogistic</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gen. Sym. Logistic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dirichlet</td>
<td>✓</td>
<td></td>
<td>only ( s(w) ) is given</td>
</tr>
<tr>
<td>Mixed</td>
<td>¬</td>
<td>Unless ( \psi = 1 )</td>
<td></td>
</tr>
<tr>
<td>Asym. Mixed</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
symmetric version with all its mass in the interior. We refer to it as **symmetric negative logistic model**. (See the upper right panel of Fig 5.)

- The bilogistic model (Smith, 1990). The bilogistic model is an asymmetric extension of the logistic model, where the spectral measure $S$ does not put positive mass on the boundary points. Its exponent measure function is

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

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

$$s_{\text{Bilog}}(w) = \frac{(1 - \psi_1)(1 - q)q^{1-\psi_1}}{(1 - w)w^2((1 - q)\psi_1 + q\psi_2)}.$$

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

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

(See the lower left panel of Fig 5.) 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 $s_{\text{Negbilog}}(w) = -s_{\text{Bilog}}(w)$ and similarly $S(\{0\}) = S(\{1\}) = 0$.

- Tajvidi’s generalized symmetric logistic model (Tajvidi, 1996).

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

where $1 \leq \alpha$ and $1 < \psi \leq 2\alpha - 2$.

- Dirichlet model (Coles and Tawn, 1991). The Dirichlet model is non-exchangeable like the two bilogistic models above, and has all probability mass of the spectral density confined to the interior.

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

and

$$s_{\text{CT}}(w) = \frac{\psi_1^\psi_1 \psi_2^\psi_2 \Gamma(\psi_1 + \psi_2 + 1)}{\Gamma(\psi_1)\Gamma(\psi_2)} \frac{w^{\psi_1 - 1}(1 - w)^{\psi_2 - 1}}{(\psi_1 w + \psi_2 (1 - w))^{1+\psi_1+\psi_2}},$$

where $q = \psi_1 y / (\psi_1 y + \psi_2 x)$, $\psi_1, \psi_2 > 0$, $\beta$ is a normalized incomplete beta function. (See the upper right panel of Fig 5.)
• Polynomial model (Klüppelberg and May, 1999).

\[ V_{\text{Pol}}(x, y) = 1/x + 1/y - \sum_{k=2}^{m} \psi_k \sum_{l=0}^{m-k} \left( \frac{m-k}{l} \right) x^{l+k-1} y^{m-k-l-1} (x + y)^{m-1} \]

and \( s_{\text{Pol}}(w) = m(m-1)\psi_m w^{m-2} + (m-1)(m-2)\psi_{m-1} w^{m-3} + \ldots + 2\psi_2 \), 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 \( S \) on the sides on the boundary points is \( S(\{0\}) = 1 - \sum_{k=2}^{m} \psi_k \) and \( S(\{1\}) = 1 - \sum_{k=2}^{m} (k-1)\psi_k \).

A special case is the asymmetric mixed model (Tawn, 1988b). \( V_{\text{AsyMix}}(x, y) = 1/x + 1/y - xy(x + y)^{-2}((\psi_1 + \psi_2)/x + (\psi_1 + 2\psi_2)/y) \)

and \( s_{\text{AsyMix}}(w) = \psi_1 w^3 + \psi_2 w^2 - (\psi_1 + \psi_2)w + 1 \), 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 \( S(\{0\}) = 1 - \psi_2 \) and \( S(\{1\}) = 1 - \psi_2 \). Consequently \( \psi_2 = 1 \) is the only symmetric case when all mass is in the interior of \([0,1]\).

### 2.4.2 Multivariate models

• Logistic models and variations (Tawn, 1990). Let \( B \) be the set of all nonempty subsets of \( \{1, \ldots, d\} \) and let \( B_1 = \{ b \in B : |b| = 1 \} \). The \( d \)-dimensional logistic model can be represented by

\[ V_{\text{Logistic}}(t_1, \ldots, t_d) = \sum_{b \in B} \left( \sum_{j \in b} \left( \frac{\psi_{j,b}}{t_j} \right)^{\alpha_b} \right)^{1/\alpha_b} \]

where the dependence parameters \( \alpha_b \)'s are in \((0,1]\) for all \( b \in B \setminus B_1 \) and the asymmetry parameters \( \theta_{j,b} \)'s are in \([0,1]\) for all \( b \in B \) and \( j \in b \). The constraints

\[ \sum_{b \in B_1} \theta_{j,b} = 1, \quad \text{for } j = 1, \ldots, d \]

for \( j = 1, \ldots, d \) ensure that the univariate margins are of the correct form, where we define \( B_{(j)} = \{ b \in B : j \in b \} \). As we pointed out previously in the bivariate case, for getting an absolutely continuous model, both asymmetry parameters must be equal to 1. Consequently, as every bivariate margins must satisfy the same criteria, all asymmetry parameters must necessarily be equal to 1, and so actually almost all terms assuming different dependence parameters for any subsets of components is cancelled from the formula apart from the one which captures all \( d \) components together. Following the above logic the only realistic logistic model within an MGPD is the form of

\[ V_{\text{Logistic}}(t_1, \ldots, t_d) = \left( \sum_{i=1}^{d} \left( \frac{\psi_i}{t_i} \right)^{\alpha} \right)^{1/\alpha} \]
having only one parameter to capture the dependence among the dimensions. Of course we should also emphasize, that this simplicity appears only in the dependence structure and the correct choice for the marginal parameters gives more flexibility to the model. Moreover we should notice that marginal parameters of the underlying MEVD model are not marginal parameters in the usual sense for the MGPD model, because of having the \( \log G(0,\ldots,0) \) constant term in its definition (6).

- **Dirichlet model** (Coles and Tawn, 1991) The spectral density of the model is

\[
s_{\text{Dirichlet}}(w) = \frac{\Gamma\left(\left(\sum_{i=1}^{d} \psi_i\right) + 1\right)}{\left(\sum_{i=1}^{d} \psi_i w_i\right)^{p+1}} \prod_{i=1}^{d} \frac{\psi_i}{\Gamma(\psi_i)} \prod_{i=1}^{d} \left(\frac{\psi_i w_i}{\sum_{j=1}^{d} \psi_j w_j}\right)^{-1},
\]

where \( \psi_i > 0 \) for \( i = 1,\ldots,d \). Although the corresponding exponent measure \( V \) is complicated, it is obtainable by using numerical integration.

### 2.5 Construction of new non-exchangeable models

From the second and third column of Table 1 we can see that there is a lack of easily computable non-exchangeable models, especially if all probability mass is required to be put on the interior of the \( S_2 \), ensuring the absolute continuity of the model. The bilogistic and negative bilogistic models are available, but without having an explicit formula for their exponent measures, or the Dirichlet model, in which case there is explicit formula only for the spectral density. In order to solve this problem, here we propose a methodology which allows to construct new dependence models with extra asymmetry parameter(s) from any valid models. As the result of this method we may obtain more flexible non-exchangeable models defining a new class for absolutely continuous MGPD-s. Because of its mathematical simplicity we illustrate the method in the bivariate setting using dependence function, but the same methodology can be extended to the higher dimensional cases as well. The algorithm is fairly simple:

1. Take a differentiable **baseline dependence model** from Table 1 (except asym. logistic or asym. negative logistic) and switch characterisation form from exponent measure to dependence function by \( A(t) = V(\frac{t}{\Psi(t)}, \frac{1}{\Psi(t)}) \);
2. Take a strictly monotonic transformation \( \Psi(x) : [0,1] \rightarrow [0,1] \), such that \( \Psi(0) = 0, \Psi(1) = 1 \);
3. Construct a **new dependence model** from the baseline model \( A_{\Psi}(t) = A(\Psi(t)) \);
4. **Check the constraints**: \( A_{\Psi}(0) = -1, A_{\Psi}'(1) = 1 \) and \( A_{\Psi}'' \geq 0 \) (convexity).

For the construction of a feasible transformation we can assume e.g. that it has a form of

\[
\Psi(t) = t + f(t)
\]

hence

\[
(A(\Psi(t)))' = A'(\Psi(t))\Psi'(t) = A'(\Psi(t)) \times [1 + f'(t)].
\]
Obviously by choosing the second term of $\Psi(t)$ such that $f'(0) = f'(1) = 0$, the $A'_{\psi}(0) = -1$ and $A'_{\psi}(1) = 1$ constrains are fulfilled for the new model if and only if $A'(0) = -1$ and $A'(1) = 1$ is true for the baseline model. The only problem which can occur, is that the new dependence function $A_{\psi}$ is not necessarily convex. In general we found that the following functional form leads to valid models

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

where $\psi_1 \in \mathbb{R}$ and and $\psi_2 \geq 2$ are asymmetry parameters. (See Fig 1 for illustration.) If $\psi_1 = 0$ we get back the baseline model. The convexity inequation below provides the valid range for the asymmetry parameter $\psi_1$, assuming fixed baseline $A(t)$ and $\psi_2$

$$(A(\Psi(t)))'' = A''(t + f(t))(1 + 2f'(t)) + A'(t + f(t))f''(t) \geq 0.$$ 

Graphical illustration of some valid model parameterisation can be found in Fig 2. Later when using these transformation we put the "\$\Psi-\$" prefix before the name of the baseline dependence model, as e.g. $\Psi$-logistic model. The difference caused by the new asymmetry parameters can be seen in Fig 3 for 3 $\Psi$-logistic BGPD models.
Figure 2: First and second derivatives of \( \Psi \)-logistic and \( \Psi \)-negative logistic dependence functions, where the transformation \( \Psi(t) = t + \psi_1 \psi_2(t) \) has fixed \( \psi_2 = 2 \) and different \( \psi_1 \) parameters.

Figure 3: BGPD density plots using \( \Psi \)-logistic dependence models with fixed \( \psi_2 = 2 \) and \( \psi_1 = -0.3/0/0.3 \) parameters.
3 Statistical inference on wind speed data

We illustrate the practical application of the proposed methods using data on wind speeds that have been collected at four locations in northern Germany over the past five decades (1957-2007): Hannover, Bremerhaven, Schleswig and Fehmarn. The entire period covers 18061 days, and 17926 complete observations are actually available during this period (the remaining 135 values have missing coordinates). The observations are considered as being stationary, but non-stationarity can also be handled e.g. by choosing time-dependent model parameters as in Rakonczai et al. (2010). After choosing a relatively high quantile (95%) as threshold level the main aim is modeling threshold exceedances occurring simultaneously at any pairs or triple of measuring stations. The bivariate data and the threshold levels are displayed in Fig 6. The wind speed data for the above cities are rather strongly correlated, due to the relatively short distance between them (with Kendall’s correlations in the range 0.45-0.55 for pairwise comparisons between the four stations), and previous studies have suggested that BGPD models can have reasonable performance for this level of association (Rakonczai and Tajvidi, 2010).

3.1 Computational aspects of estimation and prediction

The parameters can be estimated by numerical maximum likelihood method and in a Bayesian way by using MCMC simulations with conditional resampling algorithm. The effective parameterisation of the MCMC algorithm seemed to be more complicated and required longer time for computation, so we found the maximum likelihood method preferable. Although the maximum likelihood estimator is complicated to derive analytically it is fairly easy to find by numerical optimisation using the Nelder-Mead algorithm (as implemented in R using the optim function). Some complications may arise as usually the $\mu_i, \sigma_i$ parameters show high correlation within the estimated MGPD models. Our general finding is that the effect of setting one of these parameters (e.g. $\mu_1$) to be zero has a negligible effect upon the parameter estimates for $\xi_i$ and $\alpha_i$ and upon the maximum value of the log-likelihood, suggesting that the models with and without fixed $\mu_1$ are effectively equivalent. Fixing $\mu_1$ to be zero removes the strong correlations between the remaining parameters. Another problem can be that in the parameterisation of the MGPD there are no automatic constrains providing positivity in the original scale of the observations. Hence, theoretically, it can happen that the fitted model gives positive probability on negative regions even for wind speed data being non-negative by its nature. This discrepancy can be avoided by assuming the margins having finite left endpoints not less than the theoretical minimum for the given applications. Similar assumptions have been used in Dryden and Zempléni (2006).

3.2 Results for bivariate observations

The large number of model parameters (7-9), which are beyond the interpretable complexity (location and scale parameters of the underlying MEVD’s of MGPD are not any more location and scale parameters for the MGPD margins) requires some transparent illustration. Hence we used bivariate prediction regions as an alternative for visualising
the estimates. A prediction region (Hall and Tajvidi, 2004) for a given probability $\gamma$ is a region bounded by a horizontal level curve of the bivariate density over which the integral of the density equals $\gamma$. The dependence parameters for 4 pairs of locations can be found in Table 2 and the prediction regions of some models are displayed in Fig 4. In the new ($\Psi$) models the asymmetry parameters seem to be non-zero, providing some evidence of asymmetry in the data, whereas the dependence parameters are very close to those estimated for the baseline models. The effect of asymmetry parameters can be clearly seen in the top right panel of Fig 4, where the $\Psi$-NegLog model has a relevant torsion in its regions to the left. Similar torsion occurs for the regions calculated from the Coles-Tawn dependence model. It is still difficult to make difference based on visually very similar regions, hence a formal test would be very useful. The test we propose in this case can be easily performed as a simple byproduct of the prediction region method. It is known for any region that how many observations are expected to be in/out of it, as well as between two neighboring regions, so by comparing these theoretical frequencies with the realisation we can perform a $\chi^2$ test. The results for Bremerhaven and Schleswig can be found in Table 3. The critical value for the test is 11.07 ($p = 0.95, df = 5$), hence the Coles-Tawn ($T_{CT} = 9.1$) and $\Psi$-Neglog models ($T_{\Psi\text{-NegLog}} = 10.7$) may be accepted. In general the new asymmetric ($\Psi$) models are significantly better than their symmetric baseline versions, $T_{\text{Log}} = 21.5$ reduces to $T_{\Psi\text{-Log}} = 16.9$ and $T_{\text{NegLog}} = 14$ reduces to $T_{\Psi\text{-NegLog}} = 10.7$. The Tajvidi’s generalized symmetric logistic model ($T_{\text{Tajvidi’s}} = 11.5$) is the best among the symmetric ones.

Figure 4: Prediction regions at Bremerhaven and Schleswig for 4 different models (negative logistic, $\Psi$-negative logistic, Coles-Tawn and Tajvidi’s generalized symmetric)
<table>
<thead>
<tr>
<th>Models</th>
<th>Hannover and Schleswig</th>
<th>Bremerhaven and Fehmarn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>$\alpha = 1.66$</td>
<td>$\alpha = 1.985$</td>
</tr>
<tr>
<td>$\Psi$-Log</td>
<td>$\alpha = 1.662, \psi_1 = 0.207, \psi_2 = 2$</td>
<td>$\alpha = 2.049, \psi_1 = -0.423, \psi_2 = 2$</td>
</tr>
<tr>
<td>NegLog</td>
<td>$\alpha = 0.94$</td>
<td>$\alpha = 1.257$</td>
</tr>
<tr>
<td>$\Psi$-NegLog</td>
<td>$\alpha = 0.963, \psi_1 = 0.354, \psi_2 = 2$</td>
<td>$\alpha = 1.287, \psi_1 = -0.249, \psi_2 = 2$</td>
</tr>
<tr>
<td>Mix</td>
<td>$\psi_1 = 0, \psi_2 = 1$</td>
<td>$\psi_1 = 0, \psi_2 = 1$</td>
</tr>
<tr>
<td>C-T</td>
<td>$\psi_1 = 1.42, \psi_2 = 0.69$</td>
<td>$\psi_1 = 1.12, \psi_2 = 2.39$</td>
</tr>
<tr>
<td>Tajvidi’s</td>
<td>$\psi_1 = 1.84, \psi_2 = 0.37$</td>
<td>$\psi_1 = 2.24, \psi_2 = 0.44$</td>
</tr>
<tr>
<td>Bilog</td>
<td>$\psi_1 = 0.63, \psi_2 = 0.55$</td>
<td>$\psi_1 = 0.43, \psi_2 = 0.57$</td>
</tr>
<tr>
<td>Negbiog</td>
<td>$\psi_1 = 0.8, \psi_2 = 1.37$</td>
<td>$\psi_1 = 1.02, \psi_2 = 0.6$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>Hannover and Schleswig</th>
<th>Bremerhaven and Fehmarn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>$\alpha = 2.06$</td>
<td>$\alpha = 1.95$</td>
</tr>
<tr>
<td>$\Psi$-Log</td>
<td>$\alpha = 2.072, \psi_1 = 0.334, \psi_2 = 2$</td>
<td>$\alpha = 1.992, \psi_1 = 0.369, \psi_2 = 2$</td>
</tr>
<tr>
<td>NegLog</td>
<td>$\alpha = 1.34$</td>
<td>$\alpha = 1.22$</td>
</tr>
<tr>
<td>$\Psi$-NegLog</td>
<td>$\alpha = 1.303, \psi_1 = 0.231, \psi_2 = 2$</td>
<td>$\alpha = 1.226, \psi_1 = 0.315, \psi_2 = 2$</td>
</tr>
<tr>
<td>Mix</td>
<td>$\psi_1 = 0, \psi_2 = 1$</td>
<td>$\psi_1 = 0, \psi_2 = 1$</td>
</tr>
<tr>
<td>C-T</td>
<td>$\psi_1 = 2.23, \psi_2 = 1.25$</td>
<td>$\psi_1 = 2.2, \psi_2 = 1.06$</td>
</tr>
<tr>
<td>Tajvidi’s</td>
<td>$\psi_1 = 2.13, \psi_2 = 0.1$</td>
<td>$\psi_1 = 2.26, \psi_2 = 0.59$</td>
</tr>
<tr>
<td>Bilog</td>
<td>$\psi_1 = 0.54, \psi_2 = 0.42$</td>
<td>$\psi_1 = 0.56, \psi_2 = 0.45$</td>
</tr>
<tr>
<td>Negbiog</td>
<td>$\psi_1 = 0.59, \psi_2 = 0.93$</td>
<td>$\psi_1 = 0.62, \psi_2 = 1.05$</td>
</tr>
</tbody>
</table>

Table 2: Some fitted models. The dependence parameters are listed for 4 pairs of locations. It is difficult to compare the results by the parameter values, the differences can be shown by quantiles or prediction regions (see Fig 4).

<table>
<thead>
<tr>
<th>$\gamma$-range</th>
<th>0.99 $&gt;$</th>
<th>0.95-0.99</th>
<th>0.9-0.95</th>
<th>0.75-0.9</th>
<th>0.5-0.75</th>
<th>&lt;0.5</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>12.6</td>
<td>50.3</td>
<td>62.8</td>
<td>188.6</td>
<td>314.2</td>
<td>628.5</td>
<td>-</td>
</tr>
<tr>
<td>Log</td>
<td>7</td>
<td>30</td>
<td>66</td>
<td>216</td>
<td>350</td>
<td>588</td>
<td>21.5</td>
</tr>
<tr>
<td>$\Psi$-Log</td>
<td>9</td>
<td>33</td>
<td>60</td>
<td>215</td>
<td>349</td>
<td>591</td>
<td>16.9</td>
</tr>
<tr>
<td>Neglog</td>
<td>11</td>
<td>31</td>
<td>66</td>
<td>215</td>
<td>334</td>
<td>600</td>
<td>14.0</td>
</tr>
<tr>
<td>$\Psi$-NegLog</td>
<td>11</td>
<td>34</td>
<td>63</td>
<td>210</td>
<td>337</td>
<td>602</td>
<td>10.7</td>
</tr>
<tr>
<td>Mix</td>
<td>12</td>
<td>46</td>
<td>71</td>
<td>247</td>
<td>331</td>
<td>550</td>
<td>30.3</td>
</tr>
<tr>
<td>C-T</td>
<td>12</td>
<td>34</td>
<td>65</td>
<td>209</td>
<td>330</td>
<td>607</td>
<td>9.1</td>
</tr>
<tr>
<td>Tajvidi’s</td>
<td>11</td>
<td>35</td>
<td>62</td>
<td>212</td>
<td>340</td>
<td>597</td>
<td>11.5</td>
</tr>
<tr>
<td>Bilog</td>
<td>6</td>
<td>29</td>
<td>62</td>
<td>220</td>
<td>354</td>
<td>586</td>
<td>25.6</td>
</tr>
<tr>
<td>NegBilog</td>
<td>13</td>
<td>32</td>
<td>61</td>
<td>220</td>
<td>337</td>
<td>594</td>
<td>15.5</td>
</tr>
</tbody>
</table>

Table 3: Number of observations between the $\gamma = 0.99, 0.95, 0.9, 0.75, 0.5$-prediction regions (graphical illustration in Fig 4.) at Bremerhaven and Schleswig. In the last column there are the $\chi^2$ test statistics. The critical value for the test ($p = 0.95, df = 5$) is 11.07, hence the C-T and $\Psi$-Neglog models may be accepted.
3.3 Results for trivariate observations

There are 3D-logistic and negative logistic models estimated for the three triplets of stations. The shape parameters and dependence parameter are summarized in Table 4. By investigating the estimated marginal quantile curves (Table 5) we found that there is a significant overestimation of the high quantiles by the logistic models. The rates for the negative logistic are closer to their nominal level. Further improvements can be possibly obtained by allowing non-exchangeability in the dependence structure. To this end our proposed modification in Section 2.5 is the most promising, as fitting the Dirichlet model gets significantly more complicated than in the bivariate case. The main difficulty is that the spectral density must be numerically integrated over the interior of the unit simplex. The 3-dimensional modeling based on the proposed asymmetric extension for the above models is in progress.

<table>
<thead>
<tr>
<th>Models</th>
<th>ξ₁</th>
<th>ξ₂</th>
<th>ξ₃</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>-0.146</td>
<td>0.038</td>
<td>2.157</td>
<td></td>
</tr>
<tr>
<td>NegLog</td>
<td>0.093</td>
<td>0.110</td>
<td>0.770</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>ξ₁</th>
<th>ξ₂</th>
<th>ξ₃</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>-0.099</td>
<td>0.015</td>
<td>2.031</td>
<td></td>
</tr>
<tr>
<td>NegLog</td>
<td>0.088</td>
<td>0.098</td>
<td>0.817</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>ξ₁</th>
<th>ξ₂</th>
<th>ξ₃</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>-0.083</td>
<td>0.011</td>
<td>2.189</td>
<td></td>
</tr>
<tr>
<td>NegLog</td>
<td>0.075</td>
<td>0.157</td>
<td>0.927</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Fitted 3D models with logistic or negative logistic dependence structures.

4 Discussion

We have presented several models for multivariate exceedances and illustrated the practical application of these models using a real dataset on wind speed. An R library for fitting the proposed model, mgpd, has been developed, and will, in due course, be submitted to CRAN. For a visual (and empirical) model evaluation we suggested to compare the cover rate of different prediction regions. While this is a valuable tool, the development of more formal tests, like the ones known for copula modeling (Rakonczai and Zempléni, 2008) would definitely be a further improvement. Alternative methods for comparing models, assessing the statistical significance of individual terms in the model, and assessing predictive performance should also be explored.

The suggested asymmetric models are promising, as with their help one becomes able to fit non-exchangeable MGPD families without heavy computational burden. Our results have shown that they are real competitors of the previously known such models in terms of coverage probabilities.
Hannover, Bremerhaven and Fehmarn

<table>
<thead>
<tr>
<th>Quant.</th>
<th>$H(x, y, \infty)$</th>
<th>$H(x, \infty, z)$</th>
<th>$H(\infty, y, z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>0.95</td>
<td>0.004</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>0.90</td>
<td>0.012</td>
<td>0.009</td>
<td>0.018</td>
</tr>
</tbody>
</table>

Hannover, Fehmarn and Schleswig

<table>
<thead>
<tr>
<th>Quant.</th>
<th>$H(x, y, \infty)$</th>
<th>$H(x, \infty, z)$</th>
<th>$H(\infty, y, z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>0.95</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.90</td>
<td>0.011</td>
<td>0.014</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Bremerhaven, Fehmarn and Schleswig

<table>
<thead>
<tr>
<th>Quant.</th>
<th>$H(x, y, \infty)$</th>
<th>$H(x, \infty, z)$</th>
<th>$H(\infty, y, z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>0.95</td>
<td>0.007</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>0.90</td>
<td>0.024</td>
<td>0.026</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 5: GoF for 3D models by marginal quantile curves. There is a significant over-estimation of the high quantiles by the logistic models in general. The rates for the negative logistic are closer to their nominal level. Further improvement could be obtained by more complex non-exchangeable models.

There are a number of ways in which this work could be developed further. The most obvious extension would be assuming time-dependent model parameters or parameters depending on multiple explanatory variables. There are no additional conceptual difficulties involved in extending the model in this way, but the computational issues involved in maximising the likelihood function would probably be even more pronounced than for the current model. This may motivate the consideration of alternative methods of inference, such as Markov chain Monte Carlo (e.g., in the context of extreme value modeling, Fawcett & Walshaw, 2008).

The possible presence of residual dependence (caused by the clustering, which implies that a single extreme event may be associated with multiple extreme values) may also be investigated. Residual dependence occurs in the extremes of many environmental (and financial) time series, and methods that ignore this will tend to under-estimate standard errors and other measures of uncertainty. The block bootstrap, or extensions thereof, may provide a strategy for constructing confidence intervals in a way that accounts for residual dependence.

Acknowledgements

The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1./B-09/KMR-2010-0003.
References


Figure 5: Examples. BGPD distribution functions with same $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\xi_1 = \xi_2 = 0.1$ parameters assuming different dependence structures. In the logistic and negative logistic models $\alpha = 0.5$, in the bilogistic model $\psi_1 = 1.25$, $\psi_2 = 2$ and in the Dirichlet model $\psi_1 = 0.6$, $\psi_2 = 0.2$. 
Figure 6: Example. Observations of daily wind speeds over the 95% threshold level for the period 1957-2007. Those exceedances that exceed the threshold in both components are distinguished by blue colour.

Figure 7: "classical BGPD" and BGPD models fitted to exceedances above the 95% marginal quantiles. The shape and dependence parameters are \((\hat{\xi}_1, \hat{\xi}_2, \hat{\alpha}) = (-0.001, -0.06, 1.49)\) and \((\hat{\xi}_1, \hat{\xi}_2, \hat{\alpha}) = (0.18, 0.16, 1.59)\) for the left and right panel respectively.
Figure 8: Exceedances in 3D. As only one component must be over a threshold, the 2D margins can contain observations which are below both of their thresholds. In these cases the third component exceeds its threshold.
Figure 9: Examples. 2D margins of the 3D logistic and negative logistic models for Bremerhaven, Fehmarn and Schleswig. The two models are rather different, the logistic model leads to relevantly higher quantile curves than the negative logistic. The comparison with the observation shows that there is a strong overestimation in the logistic and a slight overestimation in the negative logistic model.
A Appendix: "classical BGPD" models fitted by 'fbvpot'

In Section 2.3 there are two different approaches described for modelling multivariate exceedances. Later on, the difference between the "classical BGPD" in (5) and BGPD in (6) is illustrated in Fig 7. Further details about "classical BGPD" are presented here, making the results comparable with those we got for the other definition. The parameter estimates, produced by `fbvpot` routine of the `evd` package, are summarised in Table 6. Similarly to the results in Table 2 there is a slight asymmetry present in the data. In this "classical" case there is parametric inference only on the upper right quarter of $\mathbb{R}^2$. The density plots are presented in Fig 10.

<table>
<thead>
<tr>
<th>Models</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log</td>
<td>$\alpha' = 0.635$</td>
</tr>
<tr>
<td>NegLog</td>
<td>$\alpha' = 0.858$</td>
</tr>
<tr>
<td>Bilog</td>
<td>$\psi_1 = 0.694 \psi_2 = 0.559$</td>
</tr>
<tr>
<td>Ngebilog</td>
<td>$\psi_1 = 0.872 \psi_2 = 1.517$</td>
</tr>
<tr>
<td>C-T</td>
<td>$\psi_1 = 1.123 \psi_2 = 0.562$</td>
</tr>
</tbody>
</table>

Table 6: "classical BGPD" models fitted by `fbvpot` in `evd`. Prediction regions are not available in default, although the density plot is given. (See Fig 10 below.) In the logistic and negative logistic models $\alpha = 1/\alpha'$. 

26
Figure 10: "classical BGPD" contour plots