The Thailand Econometrics Society, Vol. 2, No. 2 (January 2010), 174 - 192

On Prediction of Bivariate Extremes

Pál Rakonczai[a] and Nader Tajvidi[a]

[a] Department of Mathematical Statistics, Lund Institute of Technology, Sweden

ARTICLE INFO

Keywords:
Multivariate extreme value models,
prediction regions,
comperative study,
wind speed extremes.

JEL classification codes:

ABSTRACT

There are mainly two competing approaches to modeling high dimensional extremes, namely multivariate extreme value distributions and multivariate peaks over threshold models which lead to a class of distribution called multivariate generalized Pareto distributions. Although the probability theory for these models is fairly well developed the statistical properties of them are generally unknown. We compare performances of these models for prediction of extremes in different settings and apply the results to modeling of wind speed data in several cities in Germany.

When modeling such extreme events one should not leave out of consideration that observations measured in closely located stations usually show strong dependence. Thus, besides fitting univariate marginals, the knowledge of the dependence structure among the stations is also crucial. For the bivariate maxima the parametric cases are fully developed, but these structures are not always flexible enough for real applications. A promising alternative way for modeling the dependence could be obtained by non-parametric dependence functions. The most efficient known non-parametric models were introduced by Capéera et al. (1997) and Hall and Tajvidi (2000). However to obtain density estimation further refinements are needed, since these approaches do not result in dependence functions which are differentiable everywhere. To tackle this problem polynomial smoothing splines have been considered taking into account all required constraints on dependence functions. It should be noted that investigating "only" the maxima can hide the time structure within the given period, so we do not know whether the different components of the maxima occurred really simultaneously (e.g. in the same day) or not. To avoid this problem exceedances over a high threshold can be considered. We applied a new definition for describing the distribution of the exceedances proposed by Rootzén and Tajvidi (2006). The main curiosity of it is including also those observations in modeling which are above the threshold at least in one component. Both of the approaches for maxima and exceedances have been applied for bivariate datasets arising from wind time series of the last 5 decades measured in north Germany. We compute prediction regions for all fitted models, which makes the models easily comparable. Finally, as the statistical properties of the proposed exceedance model is still not fully studied, we investigate its accuracy and compare it with rather standard block maxima approach by a simulation study.

1. Introduction

Extreme value theory (EVT) is a fast growing field of research which is concerned with the analysis and modeling of "extremely" high (or low) observations. Under general assumptions on the underlying distribution of observations the theory gives the limit results for the distribution of the normalized maximum of a high number of observations (*block maxima*), or equivalently, the distribution of exceedances of observations over a high threshold (*peaks over threshold*). For modeling *block maxima*, e.g. monthly wind speed maxima, there is a natural finite-dimensional parametric family, called (generalized) extreme value distribution (EVD), having cdf as

$$G(x) = \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right\} \qquad (1)$$

where $1+\gamma\frac{x-\mu}{\sigma} > 0$. This family contains all the possible limiting distributions of the suitably normalized maxima with only 3 parameters namely $\mu \in \mathbb{R}$ which is called the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the shape parameter; see Leadbetter et al. (1983) for rigorous mathematical details of these results. When modeling *peaks over threshold*, one is concerned with not only the maxima, but all observations whose values are larger than a considerably high threshold level (e.g. 98% quantile of the observations). It can be shown that if the normalized maxima converges to an EVD then the distribution of exceedances over a high threshold can be approximated by the generalized Pareto distribution (GPD), having cdf as

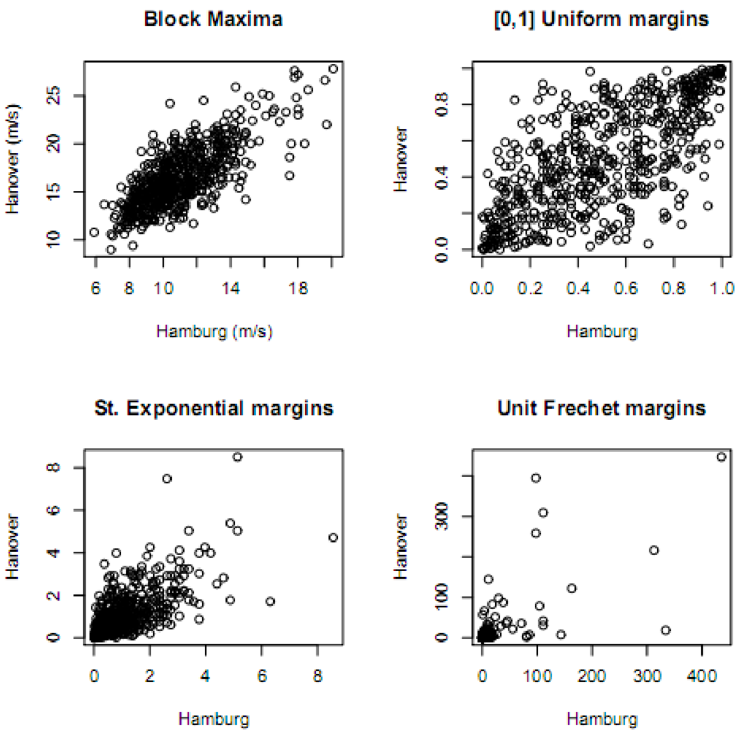
$$H(x) = 1-\left(1+\gamma\frac{x}{\sigma}\right)^{-\frac{1}{\gamma}}, \qquad (2)$$

where $1+\gamma\frac{x}{\sigma} > 0$; see e.g. Pickands (1975) and Tajvidi (1996). Moreover 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 GPD if and only if the distribution of the block maxima (as the block size tends to infinity) converges to EVD.

The multivariate extension of the univariate model is more problematic. One major complication is that the limiting distribution of the component-wise maxima of random vectors belongs to an infinite-dimensional parametric family. This is in contrast with the univariate case where the 3-parameter family in (1) provides all possible limiting distributions. The most general solution for tackling this problem is to handle the dependence structure separately from the marginals and consider certain subclasses of the multivariate extreme value distributions (MEVD) indexed by a function representing their dependence structure. There are various equivalent alternatives for characterizing the dependence structure of bivariate extremes. For instance one could work with the exponent measure, spectral density, stable tail dependence function, Pickands dependence function or copulas. It is mainly a matter of convenience which approach is used and, of course, all of them lead us to the same model.

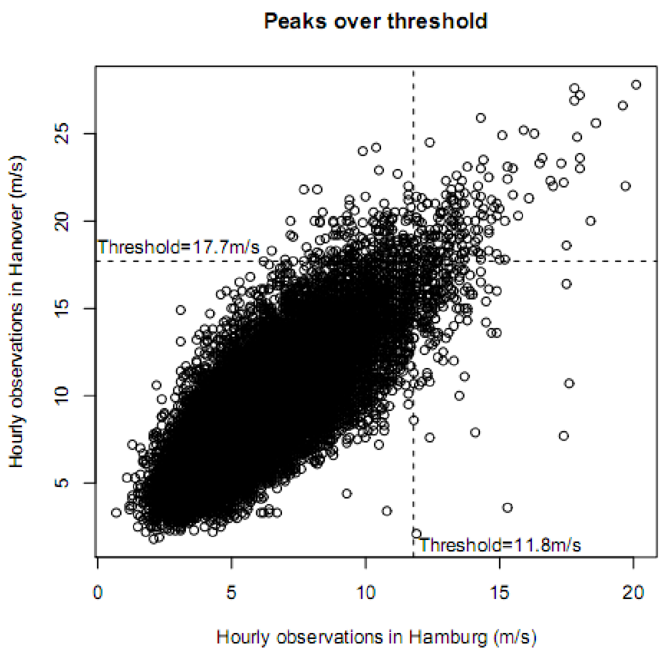
In Section 2 we define the characterization which is considered to be the most suitable to present our methods; other definitions and formulas can be found for example in Berlaint et al. (2004). One can also express the MEVD assuming EVD, uniform, exponential or Frechet margins (see Figure 1 for an illustration). This depends on which one provides the most natural and handy formula for the problem, but the choice itself is not supposed to affect the end results after transforming back to the original scale. Here we also present a definition of an alternative extension of the bivariate GPD.

Figure 1: Bivariate monthly maxima of wind speed at Hamburg and Hanover at different scales



This version has been proposed by Rootzén and Tajvidi (2006) and differs from the usual approaches in the way that it is capable to model all observations which exceed the threshold at least in one component (see Figure 2 for an illustration). Other possible extensions deal only with the observations which are over the threshold in all components.

Figure 2: The hourly observations of wind speed measurements at Hamburg and Hanover, and the 98% quantiles of the marginals as threshold levels



The main aim of the paper is to fit reasonable models both for maxima and for threshold exceedances in the bivariate setting and find prediction regions with a given high probability where the future observations are supposed to occur. In Section 3 we investigate the accuracy of prediction regions based on different modeling approaches by a simulation study. Here we generate large bivariate samples having specific marginal distributions which have EVD or GPD as a limit for their maxima or exceedances respectively, and link these marginals with various kind of dependence structures at different level of association. After choosing a suitable block size or thresholds both models have been fitted and their performances for prediction were studied.

Section 4 is devoted to the application of these methods to a wind speed dataset from Germany. We present the prediction regions for bivariate wind speed according to block maxima and threshold exceedances methods and compare the results with the observations. Similar bivariate GPD models have also been used in a recent paper by Brodin E. and Rootzén H. (2009) to study wind storm losses in Sweden.

2. Bivariate Extreme Value Models

In this section we briefly summarize the definitions for the bivariate extreme value models which are discussed in the later sections. Of course other alternative characterizations are also possible, many of them are gathered together in the book of Berlaint et al. (2004). At first we give a short overview of the bivariate block maxima methods, concerning to parametric and non-parametric cases too. These methods are extensively studied in the recent literature. In the second part of the section we present an alternative extension of the GPD for modeling bivariate threshold exceedances. The original idea of the method can be found in Tajvidi (1996) which has been further developed for more general cases in Rootzén and Tajvidi (2006). To the best of our knowledge the statistical properties of these models have not been studied completely yet.

2.1 Bivariate Block Maxima Methods

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. 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 EVD, and by its Pickands dependence function A, through (see (8.45) in Berlaint et al. (2004))

$$G(x_1,x_2)=\exp\left\{\log(G_1(x_1)G_2(x_2)) \times A\left(\frac{\log(G_2(x_2))}{\log(G_1(x_1)G_2(x_2))}\right)\right\}. \quad (1)$$

In this setting *A(t)* is responsible to capture the dependence structure between the margins. The Pickands dependence function (or shortly dependence function) A is necessarily convex and lies entirely in the triangle defined by the points (0, 1),(1, 1) and (1/2, 1/2) binding the upper left and right corners. Formally *A(t)* satisfies the following three properties, which we denote by (P)

1. *A(t)* is convex
2. $\max\{(1-t), t\} \le A(t) \le t$
3. $A(0) = A(1) = 1$.

The lower bound in the second item of (P) corresponds to the complete dependence $G(x_1, x_2) = \min\{G_1(x),G_2(x_2)\}$, whereas the upper bound corresponds to (complete) independence $G(x_1, x_2) = G_1(x_1)G_2(x_2)$. Due to the probability integral transform $U_i = G_i(X_i)$, i = 1, 2 we obtain uniformly distributed variables on the unit interval, which can easily be further transformed to any desired distribution. So for simplicity, instead of the general formula of (3) BEVD it is often formulated assuming standard exponential margins. Therefore denote $Y_i = T_i(X_i) = -\log(U_i)$, i = 1, 2 then the joint survival of the new vector (Y_1, Y_2) can be written as

$$\bar{G}_*(y_1, y_2) = P(Y_1 > y_1, Y_2 > y_2) \qquad (4)$$

$$= P(-\log(G_1(X_1) > y_1, -\log(G_2(X_2) > y_2)$$

$$= P(X_1 \le G_1^{-1}(e^{-y_1}), X_2 \le G_2^{-1}(e^{-y_2}))$$

$$= \exp\left\{-(y_1 + y_2)A\left(\frac{y_2}{y_1 + y_2}\right)\right\},$$

where "*" denotes the marginal change in the distribution. The benefit of the new simpler form (beyond the handyness in computing) will be elaborated further when we derive the non-parametric estimator of A(t). To obtain the density of BEVD one needs the first and second derivatives of the dependence function denoted by $A'(\cdot)$ and $A''(\cdot)$ respectively. The density g (on the original scale) can be expressed as

$$g(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} G(x_1, x_2)$$

$$= \frac{\partial^2}{\partial x_1 \partial x_2} \bar{G}_*(T_1(x_1), T_2(x_2)) \qquad (5)$$

$$= \bar{G}_*(T_1(x_1), T_2(x_2)) T_1'(x_1), T_2'(x_2)$$

$$\times\Big(\big(A(\xi) + (1-\xi)A'(\xi)\big)$$

$$\times\big(A(\xi) - \xi A'(\xi)\big) + \eta A''(\xi) \Big),$$

Where

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

$$T_i'(x) = -\frac{1}{\sigma_i}\left(1 + \gamma_i \frac{x - \mu_i}{\sigma_i}\right)^{-\frac{1}{\gamma_i}-1}, i = 1, 2$$

$$\xi = \frac{T_2(x_2)}{T_1(x_1) + T_2(x_2)}$$

$$\eta = \frac{T_1(x_1) T_2(x_2)}{\big(T_1(x_1) + T_2(x_2)\big)^3}.$$

So clearly in order to have a density for BEVD density, we need assume that *A(t)* is two times differentiable, which is not included in (P). The usual parametric models satisfy this property as well, but it turns out to be a more problematic issue when using non-parametric estimates for *A(t)*. As an illustration, the two most popular parametric models are sketched below but a number of other parametric dependence functions are presented on Figure 3. The symmetric and asymmetric logistic dependence functions are shown below

$$A_{\log}(t) = \big((1-t)^\alpha + t^\alpha\big)^{1/\alpha}, \qquad (6)$$

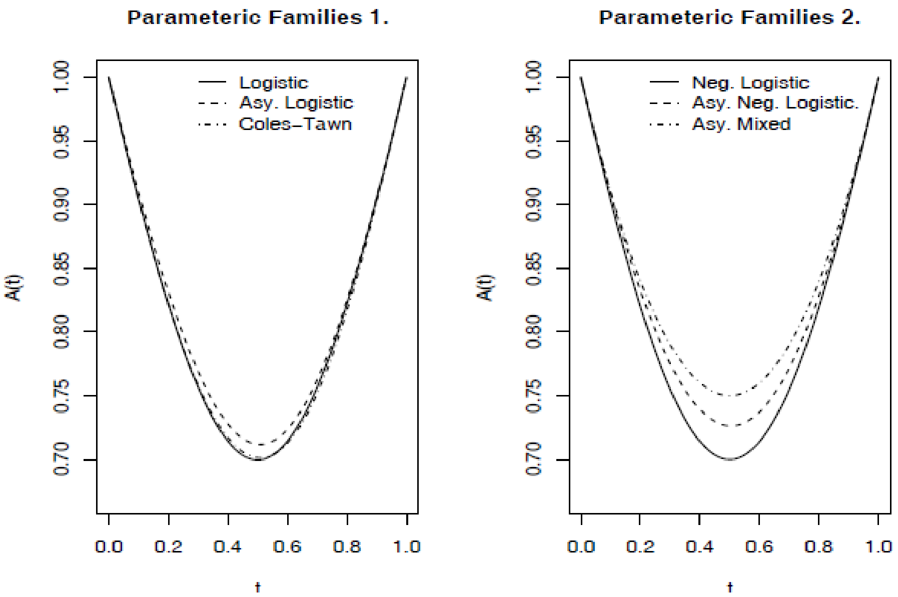
where $\alpha \ge 0$. The independence case corresponds to $\alpha = 1$.

$$A_{asy.\log}(t) = \big((\theta(1-t))^\alpha + (\phi t)^\alpha\big)^{1/\alpha} + (\theta - \phi)t + 1 - \theta, \qquad (7)$$

where $\theta \ge 0, \phi \le 1, \alpha \ge 1$ and if $\theta = \phi = 1$ the model reduces to the symmetric logistic model. Independence is obtained by $\theta = 0$ together with either $\phi = 0$ or $\alpha = 1$. This model contains some other existing models too. Details about these and other classes as well as the properties of maximum likelihood estimation can be found in Coles and Tawn (1991). For non-parametric estimation of *A(t)* we introduce the modified version of the Pickands estimator. Let the random vector (Y_1, Y_2) be as in (4), then $Z = \min\{Y_1/(1-t), Y_2/t\}$ has exponential distribution with mean *1/A(t)* for any $t \in [0,1]$, as below

$$P\left(\min\left\{\frac{Y_1}{1-t}, \frac{Y_2}{t}\right\} > x\right) = P(Y_1 > (1-t)x, Y_2 > tx) \quad (8)$$
$$= \exp\{-xA(t)\}, x \geq 0.$$

Figure 3: Different classes of parametric dependence functions fitted by {evd} package of R for observations from Hamburg and Hanover



The approximation of *1/A(t)* by the sample mean provides a very visionary estimation method. Let $(Y_{1,j}, Y_{2,j})$, $j = 1, \ldots, n$ denote a vector sample from (Y_1, Y_2). The estimator proposed by Pickands (1981,1989) can be written to the following form

$$\frac{1}{\hat{A}_n^P(t)} = \frac{1}{n}\sum_{j=1}^{n} \min\left\{\frac{Y_{1,j}}{1-t}, \frac{Y_{2,j}}{t}\right\}.$$

However the estimator has the drawback that itself actually is not dependence function according to (P). By proposing some appropriate marginal adjusment, for this define $\bar{Y}_i = n^{-1}\sum_{j=1}^{n} Y_{i,j}$, $i = 1,2$ the estimator of Hall and Tajvidi (2000)

$$\frac{1}{\hat{A}_n^{HT}(t)} = \frac{1}{n}\sum_{j=1}^{n} \min\left\{\frac{Y_{1,j}/\bar{Y}_i}{1-t}, \frac{Y_{2,j}/\bar{Y}_2}{t}\right\}, \quad (9)$$

satisfies $\hat{A}_n^{HT}(0) = 1$ as well as $\hat{A}_n^{HT}(t) \geq \max(t, 1-t)$ Although it is still not convex, by replacing it with its greatest convex minorant $\breve{A}_n^{HT}$, we obtain an estimator, which already satisfies all the necessary criteria of (P). At the expence of its flexibilty even $\breve{A}_n^{HT}$ does not verify the "extra" property of differentiability, so the BEVD density function is still not available by assuming $\breve{A}_n^{HT}$ to be A in (5). To tackle this problem there has been another modification suggested in Hall and Tajvidi (2000), namely that $\hat{A}_n^{HT}$ can be approximated by smoothing splines, constrained to satisfy (P). By choosing an appropriate fine $0 = t_0 < t_1 < \cdots < t_m = 1$ division of the interval [0, 1], and a given a smoothing parameter $\lambda > 0$, we can take $\tilde{A}_\lambda$ to be the polynomial smoothing spline of degree 3 or more which minimizes

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

subject to $\tilde{A}_\lambda(0)=\tilde{A}_\lambda(1)=1$, $\tilde{A}_\lambda'(0)\geq -1$, $\tilde{A}_\lambda'(1)\leq 1$ and $\tilde{A}_\lambda''(t)>0$ on [0,1]. By solving the non-linear optimalization problem above one van obtain a proper non-parametric estimator, on those the density estimation can be based. Remark: In reality of course we do not observe $(Y_{1,j}, Y_{2,j})$, 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 EVD (or empirical distribution function) and plug the estimator into the transformation like $\hat{Y}_{i,j}=-\log\hat{G}_i(X_{i,j})$ $i = 1, 2$ and $j = 1, ..., n$.

The comparison of the exact (symbolically calculated, see Appendix A) derivatives of the parametric dependence functions and the (numericallly approximated) derivatives of the spline smoothed non-parametric functions with different smoothing parameter is shown on Figure 4.

2.2 Bivariate Threshold Exceedances

As taking componentwise maxima can hide the time structure within the months, we do not know if the different components of the block maxima occurred simultaneously (in the same day for instance) or not. To avoid this problem, instead of taking the highest value of a given block, we can 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 block maxima method. Let (Z_1,Z_2) be the observed random variable, (u_1, u_2) a given threshold vector and $(X_1,X_2) = (Z_1 - u_1, Z_2 - u_2)$ the random vector of exceedances. For our purposes we define the bivariate generalized Pareto distribution (later BGPD) for the exceedances by its cdf as in the paper of Rootzén and Tajvidi (2006)

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

for some BEVD G with non-degenerate margins and with $0 < G(0, 0) < 1$. So practically the probability measure is *positive* in the upper three quarter planes and *zero* in the bottom left one. The main curiousity of this definition is that the BGPD distribution models those observations too, which are extremes merely in one component. (Another approach is concentrated on the upper right quarter plane, putting some probability mass onto the axes, see Tajvidi(1996)) The h density of BGPD is easily obtainable by straitforward computations (for details see Appendix B) and it can be expressed with the terms of (5) as follows

$$h(x_1,x_2)=\frac{T_1'(x_1)T_2'(x_2)}{c_0}\times\eta A''(\xi), \quad (10)$$

where

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

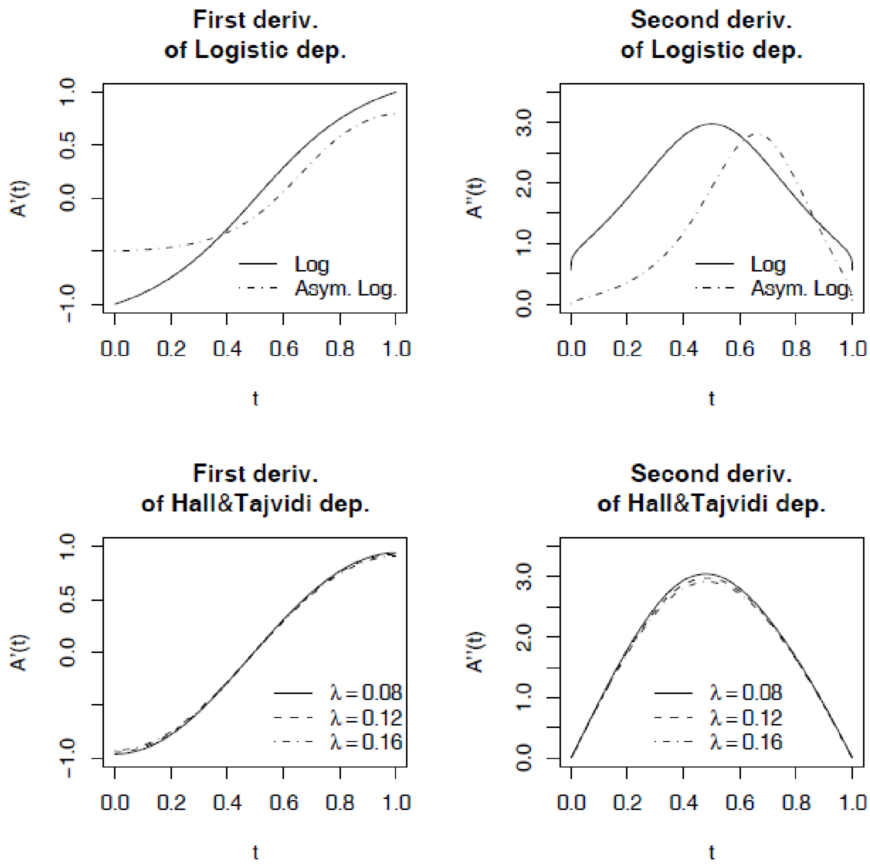
$$T_i'(x)=-\frac{1}{\sigma_i}\left(1+\gamma_i\frac{x-\mu_i}{\sigma_i}\right)^{-\frac{1}{\gamma_i}-1}, i=1,2$$

$$\xi=\frac{T_2(x_2)}{T_1(x_1)+T_2(x_2)}$$

$$\eta=\frac{T_1(x_1)T_2(x_2)}{\left(T_1(x_1)+T_2(x_2)\right)^3}.$$

$$c_0=-(T_1(0)+T_2(0))A\left(\frac{T_2(0)}{\left(T_1(0)+T_2(0)\right)}\right).$$

Figure 4: Upper block: The first and second derivatives for symmetric logistic (α = 2.06) and asymmetric logistic (α = 2.5, θ = 0.8, ϕ = 0.6) models. (The choice of α in the symmetric case is in line with the wind speed applications, in the other case the parameters are just arbitrarily chosen.) Lower block: The first and second derivatives for the smoothed Hall and Tajvidi estimators (fitted to wind speed data) with different λ smoothing parameters.



3. Simulations study

As mentioned before, the statistical properties of the proposed multivariate exceedance model is still not fully understood. In this section we present the results of a simulation study investigated its accuracy and compared it with rather standard block maxima approach by a simulation study. The general methodology we followed was that in first step we simulated bivariate samples whose marginals were in the domain of attraction of an EVD (or GPD, equivalently). Then, after computing block maxima or threshold exceedances, we fitted BEVD for the block maxima and BGPD for the threshold exceedances. At the last step we checked the accuracy of the estimated prediction region for both models. The model evaluation has been made by comparing the theoretical probability (rate) of outfalling from a region with the observed probability of outfalling, e.g. we computed how many of the simulated values turned to be fallen out of 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:

Marginals: Exponential or GPD distributions Both of these parametric families have EVD or GPD as a limit for their maxima or exceedances, respectively. To be more realistic in the choice of marginals, different parameters have been chosen, specifically $X_1 \sim Exp(2)$ and $X_2 \sim Exp(3)$ in the exponential case, and $X_1 \sim GPD(0.08, 0.13)$ and $X_2 \sim GPD(0.012, 0.09)$ in the GPD case. (For the GPD distribution the first parameter is the scale, and the second one is the shape parameter.)

Dependence structures: Logistic type Technically, we simulated from Gumbel copula for convenience, using an equivalent representation of (3). (For alternative representations of multivariate extreme value distributions see 8.6.2 in Berlaint et al. (2004)) In the last example we also investigated other type of copulas such as Clayton, Student and Gauss family; see Cherubini et al. (2004) for further examples of copulas.

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.

Sample sizes The sample sizes as $N = 5,000$, 10,000, 20,000 have been used for the simulations, illustrating how the estimations can possibly improve by the increased number of observations.

Prediction levels High γ-s. 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.

3.1 Performance of BEVD models

As it has been mentioned before in order to investigate the performance of the BGPD we need some other models as a reference to be compared with. For this purpose we have chosen 3 BEVD models, namely logistic, asymmetric logistic and the non-parametric model by Hall and Tajvidi (H&T). For the simulations a "medium" sample size $N = 10,000$ has been used. For instance if one considers the time units as days then this sample size is roughly about a 30-year long series of observations. This sample size is also generally used in this simulation study except that part, where the effect of the sample size itself has been discussed. We used 50 as block length for the models, which represents a 1-2 months long time periods in the same one observation a day analogy. For this block size we end up with 200 block maxima to be used in fitting the proposed models. We have made 100 simulations with all parameter settings and fitted the 3 proposed models. The results (expected and observed rates) for 3 prediction levels and 3 τ-values are summarized in Tables 1 and 2 for exponential and GPD margins, respectively.

Generally we found that both of the logistic and the H&T model performed similarly well. The logistic case proved to be slightly better due to the fact that the observations were originally generated from the same family. At the lowest prediction level, $\gamma = 0.75$, the non-parametric estimates have had a small bias (see the values for $\tau = 0.3$ or $\tau = 0.7$), but in all other levels both models were very close to each other and had the same variance. In the case of the asymmetric logistic model we found more bias and higher variance as well. This could be explained by the fact that the distribution we intend to model by an asymmetric model turned out to be originally symmetric and so the additional two parameters are redundant and make the estimation more uncertain and more complicated to carry out. The choice of marginals does not have any significant effect on the model performance. Table 1 and Table 2 confirm mainly the same results. One step of the simulation is

Table 1: Expected and observed rates of 3 BEVD models for simulations with exponential margins and different τ Kendall's correlations.

BEVD: Exponential margins and Gumbel copula							
Logistic		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
Hall&Tajvidi		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	0.011	0.007	0.009	0.005	0.012	0.006
95%	0.05	0.049	0.010	0.050	0.012	0.047	0.011
75%	0.25	0.261	0.020	0.254	0.018	0.221	0.021
Asym.Logistic		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	0.013	0.009	0.017	0.009	0.026	0.013
95%	0.05	0.062	0.020	0.068	0.015	0.087	0.027
75%	0.25	0.277	0.028	0.287	0.026	0.319	0.046

Table 2: Expected and observed rates of 3 BEVD models for simulations with GPD distributed margins and different τ Kendall's correlations.

BEVD: GPD margins and Gumbel copula							
Logistic		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
Hall&Tajvidi		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	0.011	0.006	0.009	0.006	0.011	0.007
95%	0.05	0.052	0.011	0.048	0.012	0.047	0.013
75%	0.25	0.256	0.022	0.248	0.021	0.225	0.020
Asym.Logistic		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	0.015	0.015	0.016	0.009	0.023	0.014
95%	0.05	0.063	0.025	0.071	0.019	0.091	0.029
75%	0.25	0.282	0.036	0.295	0.035	0.340	0.058

3.2 Performance of BGPD models

Similar to the BEVD case we carried out the predictions based on the logistic BGPD for the exceedances. In this case we also applied different sample sizes for the simulations as before, namely N = 5000, 10,000 and 20,000. As there was a need of a fast and automatic method for the simulation study we have chosen an universal, relatively high 98% quantile as threshold level for both margins in all cases. (Most probably the models could be slightly improved by a finer threshold selection method.) We considered this level high enough to use the asymptotic results to approximate the distribution of exceedances. In addition, even in the case of the smallest sample size, there have been sufficient number of exceedances (at least couple of hundreds) remained for the model fitting. As in the BEVD case, the choice of marginal distributions does not have any significant effect on the results either, so here we are limiting ourself to present the complete table only for the exponential case (many parts of the results for the GPD margins are presented in later tables for further discussion). The comparison of the expected and observed rates are shown in Table 3.

Table 3: Expected and observed rates of logistic BGPD models for simulations with exponential margins, different τ Kendall's correlations and N sample sizes.

BGPD: Exponential margins and Gumbel copula							
τ = 0.3		N = 5.000		N = 10.000		N = 20.000	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.005	0.006	0.004	0.004	0.003	0.002
95%	0.05	0.038	0.015	0.037	0.008	0.037	0.006
75%	0.25	0.280	0.024	0.281	0.015	0.281	0.011
τ = 0.5		N = 5.000		N = 10.000		N = 20.000	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.008	0.007	0.008	0.004	0.007	0.003
95%	0.05	0.046	0.015	0.046	0.009	0.043	0.007
75%	0.25	0.258	0.020	0.257	0.016	0.258	0.011
τ = 0.7		N = 5, 000		N = 10, 000		N = 20, 000	
Level	Exp.rate	Obs.rate	St.Err.	Obs.rate	St.Err.	Obs.rate	St.Err.
99%	0.01	0.010	0.008	0.010	0.004	0.011	0.004
95%	0.05	0.050	0.015	0.051	0.009	0.051	0.007
75%	0.25	0.251	0.022	0.253	0.016	0.251	0.010

We found that the BGPD model performed well when τ was in the range of medium to strong but there has been a clear bias for τ = 0.3. This bias is more serious for high prediction levels, e.g. for the 0.99% region the observed rate is only 0.005 instead of being close to 0.01 meaning that the estimated region is conservative. Roughly, there is only half of the observations outside the region than what would have been expected. The bias is presumably due to the fact that for such a low association for the original data the association level falls drastically down for the exceedances (the estimated dependence parameter α is very close to zero). The same phenomena also exist for the block maxima, but not as drastic as in this case. One should notice that for stronger association such as τ = 0.7 the model fit is almost perfect. Otherwise, by investigating the estimates based on different sample sizes, we can see that the estimates are stable and close to their expected values and they do not actually differ for different sample sizes. Moreover, there is an improvement in variances, which are decreasing by the increased sample size.

3.3 Comparison the performance of BEVD and BGPD models

As a summary of the previously presented results on the fit of different models we restructured the above tables to see the performance of BEVD and BGPD models side-by-side. In Table 4 and 5 the differences between the fit of the logistic BEVD and BGPD are presented for different margins and associations between them. The results are for sample size N =

10,000. One snapshot of the simulation is displayed by Figure 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).

Figure 5: Snapshot picture of the simulation in the case of exponential margins, linked by Gumbel copula with τ = 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 γ = 0.99 (solid line) and γ = 0.95 (dashed line). By comparing the above tables we can conclude that BEVD model provided

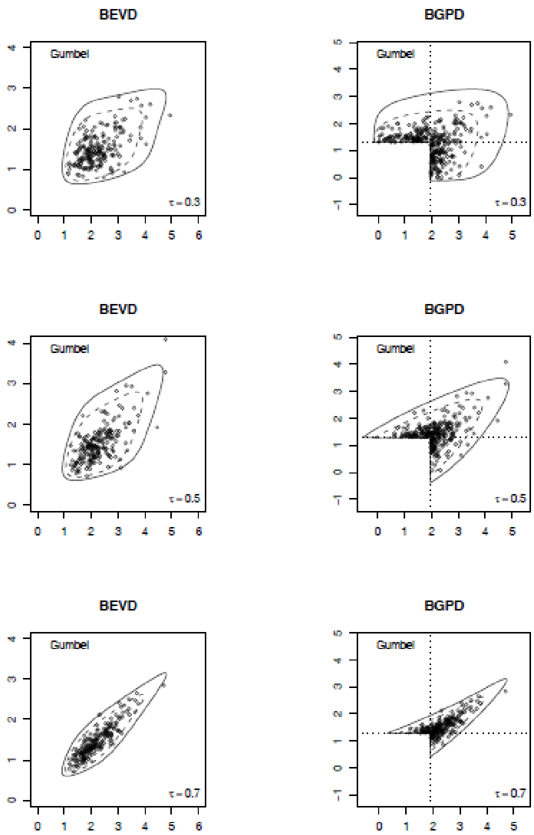


Table 4: Expected and observed rates of logistic BEVD and logistic BGPD models for simulations with exponential margins

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 5: Expected 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

remarkably good fit in all of the cases. In face with this BGPD alternated between better and weaker fit, namely for low $\tau = 0.3$ it was quite biased, but for high $\tau = 0.7$ it has been in some cases even slightly better than the BEVD model, in the sense of same rates but lower variances. Since in every cases a logistic type link has been assumed between the margins (e.g. Gumbel copulas) we also investigated what happens if another type of link function substitutes the Gumbel copula. Here we applied 3 other different parametric copula families, namely Clayton, Student-t and Gauss copulas. Their parameters have been chosen to model association with τ = 0.5 to be in line with the previous simulations. For results see Table 6.

Assuming new families of copulas reflects the same consideration as we have

Table 6: Expected and observed rates of logistic BEVD and logistic BGPD models for simulations with exponential margins linked by different copulas having the same τ = 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

already concluded before. The outcome of the performance of BGPD looks really to be dependent on the strength of association. For example if the original data have been linked by Clayton copula, then also the extremes of their marginals are supposed to have strong association. In this case the BGPD has been fairy accurate. However, the accuracy diminishes if consider the Student-t copula case. Here the extremes are not so strongly associated and finally the fit is the weakest for Gaussian copula when the dependence parameter is very close to 1. Further illustration for simulations with different copula models can be found in Figure 9 of Appendix C.

4. Application to wind speed data

Having the bivariate density in hand makes it possible to determine the probability of any specific region by integration over the region. So one can easily construct bivariate quantile regions with given γ probability level. Following the notation in Hall and Tajvidi (2004) a prediction region is defined as

$$\hat{\mathcal{R}}(u) = \left\{(x, y) : \hat{h}(x, y) \geq u\right\}$$

$$\beta(u) = \int_{\hat{\mathcal{R}}(u)} \hat{h}(x, y) dxdy$$

for $\hat{h}$ estimator of the bivariate density h as e.g in (5) or (10). 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 an γ-level *prediction region* for the future value of (X, Y).

4.1 Predictions by BEVD

For the monthly maxima different prediction regions have been estimated based on the logistic, asymmetric logistic and the spline smoothed Hall-Tajvidi estimators. In the latter case the choice of λ smoothing parameter can have considerable effect on the estimation curves. A detailed investigation on this effect of λ has been made, for the results see Appendix C, here we present only the two alternatives, which are supposed to be the most reasonable for this practical case. The curves of the prediction regions (on the original scale) are presented in Figure 6. This makes the models visually comperable, however at the first glance there is no flagrant difference among the proposed model. The more quantitative results for estimators are summarized in Table 7, where the expected number of observations falling out of a given region has been compared with the observations at different γ levels.

Figure 6: Prediction regions estimated by 2 parametric and 2 non-parametric models at high prediction levels, γ = 0.99 (solid line) and γ = 0.95 (dashed line)

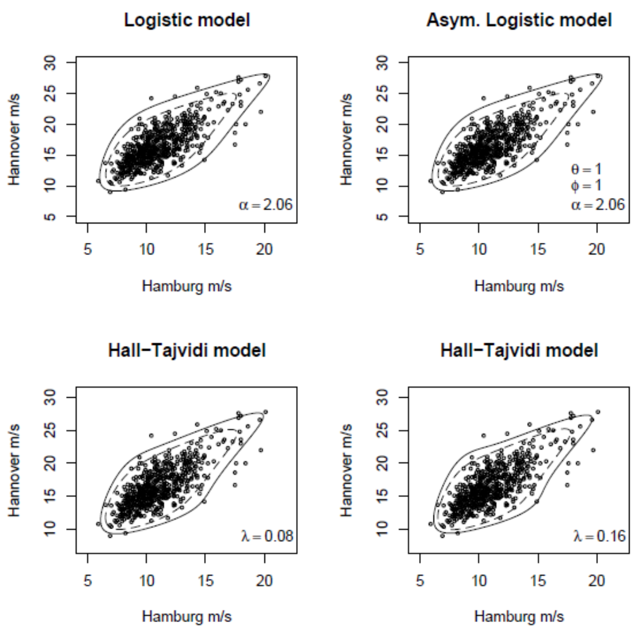


Table 7: Performance of different models at different predictive levels

BEVD Models		Logistic		Hall-Tajvidi	
Level	Expect.	Symm.	Asymm.	λ=0.08	λ=0.16
50%	296	311	311	316	309
75%	148	142	142	149	140
95%	$\approx$30	31	31	34	29
99%	$\approx$6	12	12	15	15
Log-likelihood		-2531.5	-2531.5	-2533.8	-2534.6

We see that different models can perform different ways at different γ levels so this kind of comparison is consequently not enough to decide which one is best alternative. Even if there were a model which would be very close to the observation for this given "realization" of the windspeed, we might not say it is generally the best. In order to see how these models perform we need some further investigations. Since our focus is on the ability in prediction (not only on the model fit) instead of applying formal goodness-of-fit tests a cross-validation procedure has been performed. The data has been splitted into two equal complementary parts by resampling without replacement. One set has been considered as the knowledge about the past (training set) and the other part (testing set) as "future" observations we intend to predict. The following steps have been repeated 100 times:

1. Partitioning the windspeed maxima into training and testing sets (50%-50%)
2. Model fitting on the training set by maximum likelihood estimation
3. Validation on the testing set (by the same statistics as in Table 7)

The results are presented in Table 8. We found that there is no major difference in model performance, although asymmetric logistic model looks to have slightly higher variance than the others. Since the asymmetry parameters of (7) are $\theta \approx \phi \approx 1$, there is in fact no considerable reason to assume assymetry and so get the model more complicated. Technically, by using the two additional parameter θ and ϕ in the logistic setting (6), it becomes more difficult to find proper initial values for the likelihood estimation as well. As an automatic algorithm has been needed for the cross validation method, we started the optimaliztion for the maximum likelihood from parameter vector $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\gamma}_1, \hat{\mu}_1, \hat{\sigma}_2, \hat{\gamma}_2, \alpha_0)$ given by $\hat{G}_1, \hat{G}_2$ the univariate EVD estimates for the margins and fixed $\alpha_0 = 2$. By using this setting as default the fitted asymmetric logistic model turned out to have lower likelihood value than the symmetric one, meaning that the method found only a local maximum in that case. Generally we found that the prediction regions underestimate the windspeed at most levels, except at 75% level. The most serious bias is at the highest 99% level when there are 6 observations falling out in average from where only 3 would be expected.

Table 8: Cross-validation

Cross-validation of BEVD					
BEVD Models		Logistic		Hall-Tajvidi	
Level	Expect.	Symm.	Asymm.	λ=0.08	λ=0.16
50%	148	154.2(7.3)	157.7(8.3)	156.4(6.6)	155.2(6.6)
75%	74	71.0(6.4)	74.6(8.3)	73.2(6.3)	71.2(6.0)
95%	$\approx$15	16.3(2.6)	17.4(3.8)	17.3(2.8)	16.3(2.9)
99%	$\approx$3	5.9(1.7)	6.1(2.0)	5.9(1.7)	5.8(1.7)

4.2 Predictions by BGPD

The prediction regions have been calculated also for the exceedances. At first we considered the density (10) with a logistic dependence function which has been inherited from the block maxima method and have been kept fixed $\alpha = 2.06$ during the maximum likelihood optimalization. Practically by doing this we just adjusted the 6 marginal parameters to an existing dependence function. After this step we also let the dependence parameter be free and optimized the maximum likelihood for all 7 parameters. Finally asymmetry parameters θ, ϕ have been added as well so 6 parameters have been optimized for the two margins and 3 for the dependence. Technically the first method was the most complicated. The convergence seemed to be extremely slow, and have been interrupted when there was no significant improvment in the likelihood (−2202.5) value after a long time running. Numerically the second method was excellent, started from the default initial values it converged very fast and gave back higher likelihood value (−2187.8) than the previous cumbersome method. For the assymmetric logistic model we could find better fit (−2165.6), but only at the expense of manual setting. The prediction regions are presented in Figure 7 and 8. The statistics for the regions together with the likelihood values are summerized in Table 9.

On the analogy of the block maxima method cross-validation has been carried out for the logistic BGPD model as well. In contrast to BEVD models there was no such an obvioius underestimation found, the estimated quantiles have been more appropriate comparing with the observations, see Table 10 for further notes.

Acknowledgments

The work of the first named author was mainly supported by agrant from the Marie Curie RTN Seamocs project "Applied stochastic models for ocean engineering, climate and safe transportation".

Figure 7: Prediction regions by a BGPD model with "inherited" logistic dependence function, which has been estimated for block maxima, at high prediction levels γ = 0.99 (solid line) and γ = 0.95 (dashed line)

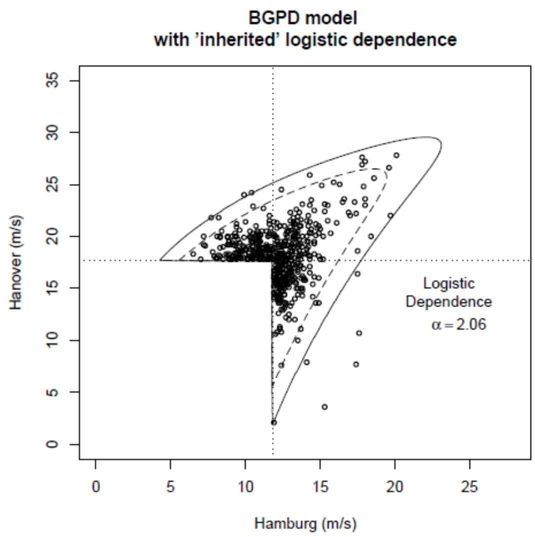


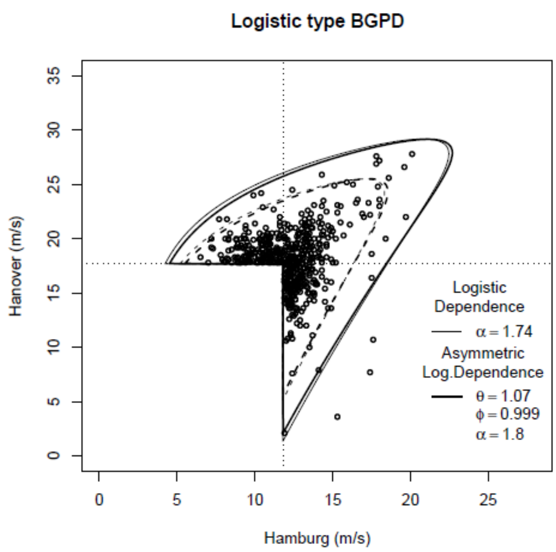
Table 9: Performance of symmetric and asymmetric logistic BGPD at different prediction levels

Logistic BGPD Models				
Level	Expect.	Fixed symm.	Symm.	Asymm.
50%	264	272	278	282
75%	132	123	119	128
95%	≈26	26	23	25
99%	≈5	8	3	4
Log-likelihood		-2202.5	-2187.8	-2165.6

Table 10: Cross-validation for logistic BGPD

Cross-validation of BGPD		
Logistic BGPD Model		
Level	Expect.	Symm.
50%	132	139.3(4.6)
75%	66	60.5(3)
95%	13.2	12.3(1.9)
99%	2.6	1.8(0.9)

Figure 8: Prediction regions estimated by symmetric and asymmetric logistic BGPD at high prediction levels, γ = 0.99 (solid line) and γ = 0.95 (dashed line)



References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J., 2004, Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistcs, 342348.

Brodin E., and Rootzén H., 2009, Univariate and Bivariate GPD Methods for Predicting Extreme Wind Storm Losses. Insurance: Mathematics and Economics 44 , 345-356

Capéraá, P., Fougéres, A.-L., and Genest, C., 1997, Nonparametric estimation procedure for bivariate extreme value copulas. Biometrika 84(3), 567-577.

Cherubini, U., Luciano, E., and Vecchiato, W., 2004, Copula methods in Finance, WileyFinance, West Sussex, England.

Coles, S., Heffernan, J., and Tawn J., 1999, Dependence measures for extreme value analysis. Extremes 2, 339-365.

Coles, S., and Tawn, J., 1991, Modelling extreme multivariate events. J. R. Statist. Soc. Ser. B 53, 377-392.

Hall, P., and Tajvidi, N., 2000, Prediction regions for bivariate extreme events. Aust. N. Z. J. Stat 46(1), 99-112.

Hall, P., and Tajvidi, N., 2004, Distribution and dependence function estimation for bivariate extreme value distributions. Bernoulli 6(5), 835-844.

Joe, H., 1997, Multivariate Models and Dependence Concepts. London: Chapman & Hall.

Kotz, S. and Nadarajah, S., 2000, Extreme Value Distributions. Imperial College Press, London.

Leadbetter M.R., Lindgren G., and Rootzén H., 1983, Extremes and related properties of stationary sequences and processes. Springer: New York

McNeil, A. J., Frey R., and Embrechts P., 2005, Quantitative Risk Management. Princeton University Press

Rootzén H., and Tajvidi N., 2006, The multivariate generalized Pareto distribution. Bernoulli 12, 917-930

Tajvidi, N., 1996, Multivariate generalized Pareto distributions. Article in PhD thesis, Department of Mathematics, Chalmers, Göteborg.

Appendices

A First and second derivatives of two parametric dependence functions

$$A'\log.(t) = \left((1-t)^{\alpha} + t^{\alpha}\right)^{\frac{1}{\alpha}-1}\left(-(1-t)^{\alpha-1} + t^{\alpha-1}\right)$$

$$A''\log.(t) = (1-\alpha)\left((1-t)^{\alpha} + t^{\alpha}\right)^{\frac{1}{\alpha}-2}\left(-(1-t)^{\alpha-1} + t^{\alpha-1}\right)^2 + (1+\alpha)\left((1-t)^{\alpha} + t^{\alpha}\right)^{\frac{1}{\alpha}-1}\left((1-t)^{\alpha-2} + t^{\alpha-2}\right)$$

$$A'a\log(t) = \left((\theta(1-t))^{\alpha} + (\theta t)^{\alpha}\right)^{\frac{1}{\alpha}-1}\left(-\frac{(\theta(1-t))^{\alpha}}{1-t} + \frac{(\theta t)^{\alpha}}{t}\right) + \theta - \phi$$

$$A''a\log(t) = (1-\alpha)\left((\theta(1-t))^{\alpha} + (\theta t)^{\alpha}\right)^{\frac{1}{\alpha}-2}\left(-\frac{(\theta(1-t))^{\alpha}}{1-t} + \frac{(\theta t)^{\alpha}}{t}\right)^2 + (1+\alpha)\left((\theta(1-t))^{\alpha} + (\theta t)^{\alpha}\right)^{\frac{1}{\alpha}-1}\left(\frac{(\theta(1-t))^{\alpha}}{(1-t)^2} + \frac{(\theta t)^{\alpha}}{t^2}\right)$$

B Density of BGPD

When calculating the density 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

$$\begin{aligned} h(x,y) &= \frac{\partial^2 H(x,y)}{\partial x \partial y} \\ &= \frac{\partial^2}{\partial x \partial y}\left(1 - \frac{\log G(x,y)}{\log G(0,0)}\right) \\ &= C_0 \frac{-T_1(x)T_2(y)}{(T_1(x)T_2(y))^3} \\ &\quad A''\left(\frac{T_2(y)}{T_1(x)+T_2(y)}\right)T_1'(x) + T_2'(y), \end{aligned}$$

Taking the usual marginal transformations we get

$$\begin{aligned} -\log G(x,y) &= -\log G_*(-\log G_1(x), -\log(G_2(y)) \\ &= -\log G_*(T_1(x), T_2(y)) \\ &= -\log G_*(t_1, t_2) \\ &= -\log\{\exp(-(t_1+t_2)A(\frac{t_1}{t_1+t_2}))\} \end{aligned}$$

$$= (t_1 + t_2) A(\frac{t_1}{t_1 + t_2}),$$

where we must get the second mixed partial derivatives of the above form

$$\frac{\partial^2}{\partial t_1 \partial t_2}\left((t_1 + t_2) A(\frac{t_2}{t_1 + t_2}) \right)$$

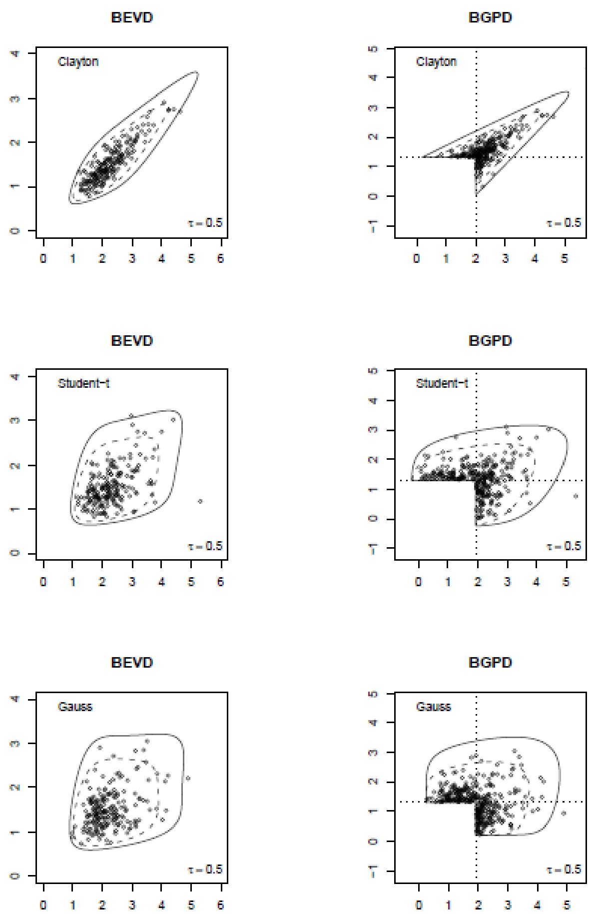
$$= \frac{\partial}{\partial t_2}(A(\mu) + (t_1 + t_2) A'(\mu) \frac{-t_2}{(t_1 + t_2)^2})$$

$$= \frac{\partial}{\partial t_2}(A(\mu) - \mu A'(\mu))$$

$$= A'(\mu)\mu'_{t_2} - \mu'_{t_2} A'(\mu) - \mu A''(\mu)\mu'_{t_2}$$

$$= \frac{t_1 t_2}{(t_1 + t_2)^3}) A''(\mu).$$

Figure 9:



C Simulation results for different copula families

Snapshot pictures of simulations assuming exponential margins, linked by Clayton, Student-t and Gauss copula with ¿ = 0.5 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 ° = 0.99 (solid line) and ° = 0.95 (dashed line).