ON MODELING AND PREDICTION OF MULTIVARIATE EXTREMES
WITH APPLICATIONS TO ENVIRONMENTAL DATA

PÁL RAKONCZAI

LUND UNIVERSITY
Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics
Acknowledgements

First I would like to thank Georg Lindgren for encouraging me to join to the SEAMOCS Project and also for motivating me to write this thesis. I also thank Nader Tajvidi for his excellent supervision and for expending so much time on correcting my mistakes even at weekends and late evenings. Further thanks to András Zempléni for his continuous and indispensable assistance even from Hungary and to László Márkus for his very valuable contribution.

Among others I would like to thank Mona Forsler for taking so much care of my matters from the first seconds of my arrival and also James Hakim, the computer guru, for his professional technical support. Thank you Jörg Wegener and Johan Sandberg for having many great lunchtime conversations in Tegners Matsalar. I am also grateful to all people in the department and around for their kindness and helpfulness during my time here.

Finally I would like to express my sincere thanks to my parents and friends for their love and support, without them I hardly could make it.

Lund, November 2009

Pál Rakonczai
# Contents

4.1 Prediction by BEVD ........................................ 41  
4.2 Prediction by BGPD ........................................ 43  

## B Copula-modeling and goodness-of-fit for 3-dimensional wind speed maxima 55  
1 Introduction .................................................. 56  
2 Copula Concepts ............................................. 57  
2.1 Elliptical Copulas .......................................... 57  
2.2 Archimedean Copulas ...................................... 58  
3 Goodness-of-Fit Tests in 3D .................................. 59  
3.1 Univariate GoF Tests ...................................... 61  
3.2 Empirical Copula Process ................................ 62  
3.3 Rosenblatt's Transform ................................... 63  
3.4 Kendall's Transform ....................................... 65  
3.5 Testing Algorithm ........................................ 67  
4 Application to Wind Speed Maxima ......................... 67  
4.1 G-Test ...................................................... 70  
4.2 B-Test ...................................................... 71  
4.3 K-Test ...................................................... 76  
5 Conclusions .................................................. 78  

## C Autocopulas: investigating the interdependence structure of stationary time series 83  
1 Introduction .................................................. 84  
2 From Copulas to Autocopulas ................................ 87  
3 Goodness-of-Fit Tests for Autocopulas ..................... 88  
3.1 Univariate GoF Tests ...................................... 89  
3.2 Copula GoF Tests .......................................... 90  
4 Example: Testing for Heteroscedasticity in the Innovation of an AR Model ........................................ 92  
5 An Application in Hydrology ................................. 99  
6 Conclusions .................................................. 105
List of papers

This thesis is based on the following papers:


Introduction

The main topic of the thesis is strongly connected to the extreme value theory (EVT), which is concerned with modeling of “extremely” high (or low) observations (extremes). In the first part of the Introduction, Section 1-3, we present a short overview of the research field. The presentation is based on Chapter 3-4 of Coles (2001) and Chapter 8-9 of Berlaint et al. (2004). In addition the article of Rootzén and Tajvidi (2006) and the materials discussed in the recent workshop series on “Risk, Rare Events and Extremes” at Bernoulli Centre (2009) have been used in these sections.

Section 1 summarizes the main results of EVT for univariate i.i.d. observations, giving limit results on the distribution of the maximum of a high number of observations and on the distribution of exceedances above a high threshold. In Section 2 the multivariate extension of the theory has been investigated. Here we present different approaches for characterizing dependence structures of multivariate extremes. In Section 3 we focus on a particular dependence concept, called copula approach, which is applicable under very general conditions. After the theoretical overview in Section 4 we give a short summary of the presented papers and finally in Section 5 we mention some possible directions for future research.


1 Univariate Extreme Value Theory

First we outline main probabilistic results providing the basis of parametric estimation techniques for extremes of univariate i.i.d. observations and then turn our attention to the multivariate questions. Let $X_1, ..., X_n$ be a sequence of inde-
pendent random variables with common distribution function \( F \) and denote
\[
M_n = \max (X_1, X_2, \ldots, X_n).
\]
In applications, \( X_i \) often represent values of an hourly or daily measured process and so \( M_n \) represents the maximum of the process over \( n \) time units. Although the distribution of \( M_n \) can be computed in a very elementary way as
\[
P(M_n \leq z) = P(X_1 \leq z, \ldots, X_n \leq z) = \prod_{i=1}^{n} P(X_i \leq z) = \{F(z)\}^n,
\]
this is not very useful in practice since \( F \) is unknown. Of course, one can suggest to estimate \( F \) from the measurements and plug the estimate into (1). But unfortunately by doing this, even small errors in the estimate of \( F \) will be multiplied up, leading to large error in the final estimate of \( F^n \). An alternative approach is proposed by EVT, suggesting to look for approximate distribution families for \( F^n \) directly, based on the extreme measurements only. Therefore, basically what we are looking for is equivalent of a central limit theory for extreme values.

1.1 Limit Distribution for Maximum

Let \( z_+ = \sup \{z : F(z) < 1\} \) denote the upper end point of the support of the distribution \( F(x) \). Then it is clear that \( M_n \to z_+ \) a.s. as \( n \to \infty \). Thus, in order to get non-degenerate limit for \( M_n \), we consider normalized maxima
\[
M_n^* = \frac{M_n - a_n}{b_n},
\]
for some sequences of constants \( \{a_n\} \) and \( \{b_n > 0\} \). The Gnedenko-Fisher-Tippet theorem states that the limit distribution, if exists, is in the class of the so-called extreme value (EV) distributions.

**Definition 1.1.** The extreme value distribution with shape parameter \( \xi \) has the following distribution function.
If \( \xi \neq 0 \),
\[
G_\xi (x) = \exp \left[-(1 + \xi x)^{-1/\xi}\right]
\]
for \( 1 + \xi x > 0 \) (otherwise 0 if \( \xi > 0 \) and 1 if \( \xi < 0 \)).
If \( \xi = 0 \),
\[
G_\xi (x) = \exp \left[-e^{-x}\right].
\]
1. Univariate Extreme Value Theory

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

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

**Definition 1.2.** The generalized extreme value (GEV) distribution is defined as

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

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

**Theorem 1.1.** [Fisher and Tippett (1928), Gnedenko (1943)] If there exist $\{a_n\}$ and $\{b_n\} > 0$ sequences such that

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

where $G$ is a non-degenerate distribution function, then $G$ necessarily belongs to the GEV family, defined in (2). In this case we say that the distribution of $X_i$ belongs to the max-domain of attraction of a GEV distribution.

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

$$P(M_n \leq z) \to G\left(\frac{z - a_n}{b_n}\right) = G'(z).$$

1.2 Limit Distribution for Threshold Exceedances

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

**Theorem 1.2.** Let $X_1, \ldots, X_n$ be a sequence independent random variables with common distribution function $F$ and let

$$M_n = \max (X_1, X_2, \ldots, X_n).$$

Suppose that $F$ belongs to the max-domain of attraction of GEV for some $\xi, \mu$ and $\sigma > 0$. Then for large enough $u$

$$P(X_i - u \leq z | X_i > u) \rightarrow H(z) = 1 - \left(1 + \frac{\xi z}{\sigma}\right)^{-\frac{1}{\xi}}$$

as $u \rightarrow z_+$, \hspace{1cm} (4)

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

The family defined in (2) is called generalized Pareto distribution (GPD).

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

Theorems 1.1 and 1.2 can be used to construct estimation methods for the distribution of maxima or exceedances. According to Smith (1985) we know that the maximum likelihood estimator is consistent and asymptotically normal if $\xi > -1/2$.

1.3 How Universal Is The Theory?

The EVT-based statistical procedures implicitly assume that most distributions of practical interest lie within the max-domain of attraction of a GEV (and hence the conditional distribution of exceedances converges to GPD). Therefore, a natural question arises: how general is the class of distributions for which the limit results in Theorems 1.1 and 1.2 hold? Although it is not difficult to find counterexamples (e.g. among discrete distributions), the most well-known continuous distributions lie within the domain of attraction of GEV and GPD. When the shape parameter $\xi \neq 0$ there exist relatively easily verifiable conditions.

**Fréchet-type limit** A distribution function $F$ belongs to the domain of attraction of GEV with $\xi > 0$ if and only if the survival function $F$ is regularly varying.
2. Multivariate Extreme Value Theory

**Definition 1.3.** We say that a measurable function $R$ is regularly varying (in Karamata’s sense) at $\infty$ if there exists a real number $\rho$ such that for every $x > 0$

$$
\lim_{t \to \infty} \frac{R(tx)}{R(x)} = t^\rho.
$$

If $\rho = 0$, the function $R$ is said to be slowly varying.

These conditions are for example satisfied by the Pareto, the Cauchy and the stable (for $\alpha < 2$) distributions.

**Weibull-type limit** In contrast to the heavy-tailed distributions, the case of $\xi < 0$ contains distributions which have a finite right endpoint $z_+ < \infty$, including e.g. the uniform and the beta distributions.

**Gumbel-type limit** The $\xi = 0$ case is much more complicated. Although there exist necessary and sufficient conditions here as well, they are hardly used in practice. It can be shown that within the domain of attraction with ($\xi = 0$) there are heavy-tailed distributions whose all moments are finite (e.g. the lognormal distribution), but also light-tailed distributions (e.g. the normal, the exponential or the gamma distribution, and even some distributions whose support is bounded to the right).

**2 Multivariate Extreme Value Theory**

The multivariate extension of the univariate theory leads to various non-trivial problems. The first fundamental question one is confronted with is how to define the multivariate extreme events, as there exists no unique ordering exist for multivariate observations. Barnett (1976) considers four different categories of order relations. The most useful one in multivariate extreme value context is called **marginal ordering** (or M-ordering): For $d$-dimensional vectors $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ the relation $x \leq y$ is defined as $x_j \leq y_j$ for all $j = 1, \ldots, d$. In this case the maximum is defined by taking the component-wise maxima. Then, for a sample of $d$-dimensional observations $X_i = (X_{i,1}, \ldots, X_{i,d})$ for $i = 1, \ldots, n$ the maximum, $M_n$ is defined as

$$
M_n = (M_{n,1}, \ldots, M_{n,d}) = (\bigvee_{i=1}^n X_{i,1}, \ldots, \bigvee_{i=1}^n X_{i,d}).
$$
2.1 Limit Distribution for Component-wise Maxima

In the subsequent sections, unless mentioned otherwise, all operations and order relations on vectors are understood to be component-wise. Analogously to the univariate case we assume that $X$ has a distribution function $F$ and there exist $a_n$ and $b_n > 0$ sequences of normalizing vectors, such that

$$P\left( \frac{M_n - a_n}{b_n} \leq z \right) = F^n(b_n z + a_n) \to G(z),$$

where the $G_i$ margins of the limit distribution $G$ are non-degenerate distributions. If (5) holds then $F$ is said to be in the domain of attraction of $G$ and $G$ itself is said to be a multivariate extreme value distribution (MEVD).

**Remark 3.** Observe that from (3) we see that for each margin

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

for any $j = 1, \ldots, d$. Since $G_j$ is non-degenerate by assumption, the margins are necessarily GEV distributions.

As the previous remark shows, the more problematic part of the extension is how to handle the dependence in the multivariate case. It can be shown that MEVD cannot be characterized as a parametric family indexed by finite dimensional parameter vector (as in the GEV case). Instead, the family of MEVD is usually indexed by the class of the underlying dependence structure. Several authors, among them Resnick (1987) and Pickands (1981), have proposed equivalent characterizations of MEVD, assuming different margins.

**Remark 4.** When studying dependence structure, it is often convenient to use standardized margins. The benefit of doing this is that some properties or characterizations are more naturally seen for a given choice. However, of course, the choice of the marginal distribution itself does not make any difference in the end result after transforming the margins back to the original GEV scale.

Below we present the two equivalent characterizations of MEVD.

Let $Y = (Y_1, \ldots, Y_d)$ denote a random vector with a distribution function $G$ as in (5) and let $G_F$ be the distribution function of

$$(-1/ \log G_1(Y_1), \ldots, -1/ \log G_d(Y_d)).$$

Due to the transformation, the new vector has unit Fréchet distribution with cdf $\phi(z) = e^{-1/z}$, $z > 0$, as marginal distribution instead of GEV

$$P(-1/ \log G_j(Y_j) \leq z) = e^{-1/z}, \text{ for } z \in \mathbb{R}_+.$$
The two distribution functions $G$ and $G_F$ can be written easily from one form to the other as

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

$$G_F(z) = G(G^{-1}(-1/z_1), ..., G_d^{-1}(-1/z_d)), \ z \in \mathbb{R}_+^d$$

Other transformations are also possible for the reduction of $G$ to have simpler margins. (Popular choices are the standard exponential, Weibull, Gumbel or uniform distributions.) Having this in mind we present a possible characterization of MEVD by Resnick (1987), assuming unit Fréchet margins.

**Definition 2.1.** A $d$-variate distribution function $G$ is called to be max-stable if for every positive integer $k$ there exist $\alpha_k > 0$ and $\beta_k$ such that

$$G^k(\alpha_k \mathbf{x} + \beta_k) = G(\mathbf{x}), \ x \in \mathbb{R}^d. \quad (7)$$

It is not difficult to show that the classes of extreme value and max-stable distribution actually coincide. Resnick (1987) shows, that all max-stable distributions with unit Fréchet marginal distribution can be written as

$$G_F(x) = \exp\{-\mu_F[-\mathbf{0}, \mathbf{x}]\}, \ x \geq \mathbf{0}, \quad (8)$$

with

$$\mu_F[-\mathbf{0}, \mathbf{x}] = \int_{\mathcal{A}} A \bigvee_{i=1}^d \left(\frac{d a_i}{x_i}\right) S(d \mathbf{a})$$

Here $S$ is a finite measure on $\mathcal{A} = \{\mathbf{a} \in \mathbb{R}^d : \|\mathbf{a}\| = 1\}$, which satisfies

$$\int_{\mathcal{A}} a_i S(d \mathbf{a}) = 1 \text{ for } i = 1, ..., d,$$

where $\|\cdot\|$ is an arbitrary norm in $\mathbb{R}^d$ and $\mu_F$ is called the exponent measure.

Another (equivalent) characterization, according to Pickands (1981), assumes standard exponential margins. In the bivariate case the joint survivor function $\bar{G}_E$ is given by

$$\bar{G}_E(y_1, y_2) = P(Y_1 > y_1, Y_2 > y_2) = \exp\left\{-(y_1 + y_2)\bar{A}\left(\frac{y_2}{y_1 + y_2}\right)\right\}, \quad (9)$$
where $A(t)$, called Pickands dependence function (or shortly dependence function), is responsible to capture the dependence structure between the margins. It can be shown that the dependence function 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.

**Remark 5.** As we mentioned before, both characterizations are equivalent. E.g. the characterization in (8) by exponent measure assuming unit Fréchet margins can be transformed to the form of characterization in (9) by dependence function as follows

$$- \log G_F(x_1, x_2) = \mu_F[-\mathbf{0}, (x_1, x_2)]' = \left( \frac{1}{x_1} + \frac{1}{x_2} \right) A \left( \frac{x_1}{x_1 + x_2} \right).$$

For higher dimensions (9) could be generalized as

$$\bar{G}_E(y) = \exp \left\{ - \left( \sum_{i=1}^{d} y_i \right) A \left( \frac{y_1}{\sum_{i=1}^{d} y_i}, \ldots, \frac{y_{d-1}}{\sum_{i=1}^{d} y_i} \right) \right\},$$

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

### 2.2 Limit Distribution for Multivariate Threshold Exceedances

As taking component-wise maxima can hide the time structure within the given period, we do not know if the different components of the maxima occurred simultaneously (in the same day for instance) or not. To avoid this problem, instead of taking only the highest value of a time period, 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 other method using only the maxima. In this case the first fundamental question is which observations exactly we should consider as extremes. Is it enough to consider only the observations being over the threshold in all dimensions, or should we include those ones too, which exceed the threshold even in one single component. Here we give the following definition by Rootzén and Tajvidi (2006) which follows the latter.
3. Multivariate Copula Models

**Definition 2.2.** A distribution function $H$ is a multivariate generalized Pareto distribution if

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

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

**Remark 6.** Note, that according to this definition the lower dimensional margins of MGPD are not GPD’s. However, if $Z$ is distributed as $H$ then the conditional distribution of $Z_i|Z_i > 0$ is GPD. This property holds for all marginal distributions in any dimension less than $d$.

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

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

**Theorem 2.1** (Rootzén and Tajvidi (2006)). Suppose, that $G$ is a $d$-dimensional MEVD with $0 < G(0) < 1$. If $F$ is in the domain of attraction of $G$ then there exist an increasing continuous curve $u$ with $F(u(t)) \to 1$ as $t \to \infty$, and a function $\sigma(u) > 0$ such that

$$P(X_u \leq x|X_u \not\leq 0) \to \frac{-1}{\log G(0)} \log \frac{G(x)}{G(x \wedge 0)}$$

as $t \to \infty$, for all $x$.

According to this theory if the limit distribution for the normalized maxima converges to MEVD then the normalized exceedances converges to MGPD. The theorem is true conversely as well, and the definition of MGPD in (10) covers all possible limits for the exceedances.

### 3 Multivariate Copula Models

This approach for describing dependence is more general than those mention above as copulas are applicable for any multivariate distribution. The main theory is given by Sklar (1952) where it is shown that for any $d$-variate distribution function $H$, with univariate margins $F_i$ there exits a copula $C$, a distribution over
the $d$-dimensional unit cube, with uniform margins, such that $H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$. Moreover the copula $C$ is unique if the marginal distributions are continuous. For more details see Chapter 4 in Cherubini et al. (2004). In the following we present the characterization of MEVD by copulas and give a brief introduction of some parametric copula families which are not strictly linked to MEVD.

### 3.1 Extreme Value Copulas

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

**Definition 3.1.** The copula is called to be an extreme value copula if it is a copula of a MEVD $G$ as

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

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

**Definition 3.2.** A copula $C$ is called to be stable if $C(u) = C(u_1, \ldots, u_d)$.

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

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

### 3.2 Elliptical Copulas

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

$$
C_{\Sigma}^\Phi(u) = \Phi^{-1}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)) \quad (13)
$$

= \int_{-\infty}^{\Phi^{-1}(u_1)} \ldots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^\frac{d}{2} |\Sigma|^\frac{1}{2}} e^{-\frac{1}{2} x^T \Sigma^{-1} x} \, dx_1 \ldots dx_d.

Another member of the elliptical copula family, which is called the Student's $t$ copula, is similar to (1), but the Gaussian distributions are replaced by $t$-distributions in the formula. The Student's $t$ copula is defined as

$$
C_{\Sigma, \nu}(u) = t_{\Sigma, \nu}(t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_d))
$$

= \int_{-\infty}^{t_{\nu}^{-1}(u_1)} \ldots \int_{-\infty}^{t_{\nu}^{-1}(u_d)} \frac{1}{\Gamma\left(\frac{\nu+2}{2}\right) \nu^\frac{\nu}{2} (1 + \frac{1}{\nu} x^T \Sigma^{-1} x)^\frac{\nu+1}{2}} \, dx_1 \ldots dx_d,

where $\nu$ is the number of degrees of freedom.

### 3.3 Archimedean Copulas

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

$$
C_\phi(u) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right).
$$

Although this family has a very simple construction, it unfortunately suffers from considerable limitations. For instance there is only one parameter (or just few) to capture the entire dependence structure and by its construction all of these copulas are permutation symmetric (exchangeable). This means that all $s < d$ dimensional margins are identically distributed, which is usually a too strict assumption in many applications. Here we present two examples from this family, namely the Clayton and the Gumbel copula.

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

$$
C_{\text{Clayton}}(u) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1\right)^{-\frac{1}{\theta}}.
$$

11
The Gumbel copula has the generator $\phi_\theta(u) = [-\ln(u)]^\theta$, where $\theta \in [1, +\infty)$. Thus, the Gumbel $d$-copula function is given by

$$C_{\text{Gumbel}}(u) = e^{-\left(\sum_{i=1}^{d} - \log(u_i)^\theta\right)^\frac{1}{\theta}}.$$

**Remark 7.** It should be noted that the Gumbel copula also satisfies the stability property in (12) and so it belongs to both of the Archimedean and extreme value copula families at the same time.

## 4 Summary of the Papers

As it is pointed out in the theoretical overview, the way we define multivariate extreme events itself is a very crucial point in modeling. After deciding what is the most appropriate concept for a given case, the limit distribution is basically determined by the underlying dependence structure. In all of the three papers we focus on current problems which are connected in some way to these issues.

- **On Prediction of Bivariate Extremes** In this paper we concentrate on bivariate modeling of extremes and investigate the accuracy of a new concept for modeling bivariate threshold exceedances. We compare the accuracy of prediction regions of the proposed exceedance model with well-known models, assuming wide range of parameters. It turned out that the proposed model performs well in the cases when the association between the original time series reaches a certain level and in some cases its performance is better than the most common ones.

- **Copula-modeling and goodness-of-fit for 3-dimensional wind speed maxima** This paper is concerned with recent validation techniques for copula models. Here we apply models to real three dimensional wind speed data and point out what kind of difficulties can possibly arise in dimensions higher than 2. It turns out that not only finding a reasonable model becomes more complicated, but also to choose a proper method which is capable of detecting errors of the fitted models is problematic as well.

- **Autocopulas: investigating the interdependence structure of stationary time series** In this paper we propose a new field of application of copulas. Here we present a tool for finer assessment of stationary time series models.
5. Future Work

based on non-parametric inference on autocopulas, which are the copulas of the original and the lagged series. This way we explore the interdependence structure within the time series by copulas, which makes possible to adapt many useful procedures from copula theory to detect differences between certain time series models. After a simulation study the proposed methods are applied to models which have been fitted to a river discharge dataset.

It should be emphasized although the procedures we discuss in the above papers are illustrated by environmental examples they can also provide solution for a wide range of applications in e.g. actuarial science, finance, economics or telecommunications. A particular field of application is e.g. spatial modeling. The models in the first paper could be the base of the dependence structure between points which are in the neighborhood of each other. The fit of the chosen copula model as a dependence structure among neighbors could be checked by the methods presented in the second paper.

5 Future Work

There are still numerous open question left in the theory. Here we sketch some directions, which can be possible research goals for the future. Many of the proposed items below are basically concerned with extending some methods to higher dimensions than 2.

- Extension of the spline smoothing method for non-parametric estimates of Pickands dependence functions for higher dimensions than 2.

- Further investigation of the proposed BGPD models including comparisons with other existing threshold exceedance models and extension to higher dimensions.

- Investigation of other parametric families of dependence functions for BGPD and developing estimation procedures. Developing efficient non-parametric estimation provides a major challenge here.

- Merging together the tail distribution with the remaining part of the distribution satisfying differentiability.
Introduction

- Developing and applying non-exchangeable copula models in 2 and higher dimensions for wind data including estimation and testing procedures.

- Handling wind directions in joint modeling of wind speed maxima as well.

- Further investigation of effectiveness of the autocopula tests and comparisons with other nonlinearity tests under more general circumstances.

- Extending the study of 2-dimensional autocopulas. For instance the autocopulas of the set of first \( k \) consecutive lags would be particularly interesting, and could lead to even more stronger tests.
References

REFERENCES


On Prediction of Bivariate Extremes

Pál Rakonczai and Nader Tajvidi

Department of Mathematical Statistics, Lund Institute of Technology
Box 118 SE-22100, Lund, Sweden

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 margins, 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 Capéraa 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.

**Key words:** Multivariate extreme value models, prediction regions, comparative study, wind speed extremes.

### 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

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

\[1 + \frac{x - \mu}{\sigma} > 0.\]

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 \(\xi \in \mathbb{R}\) 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

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

where \( 1 + \frac{x}{\xi} > 0 \); 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 margins 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 Fréchet 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. 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
Figure 1: Bivariate monthly maxima of wind speed at Hamburg and Hanover at different scales

component (see Figure 2 for an illustration). Other possible extensions deal only with the observations which are over the threshold in all components.

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 margins 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
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

Figure 2: The hourly observations of wind speed measurements at Hamburg and Hanover, and the 98% quantiles of the margins as threshold levels 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.
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 component-wise 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 Berlain et al. (2004))

\[
G(x_1, x_2) = \exp \left\{ \log \left( \frac{G_1(x_1)G_2(x_2)}{G_1(x_1)G_2(x_2)} \right) \right\}. \tag{3}
\]

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. \(\max\{1 - t, t\} \leq A(t) \leq t\)
3. \(A(0) = A(1) = 1\).

The lower bound in the second item of (P) corresponds to the complete dependence \(G(x_1, x_2) = \min\{G_1(x_1), G_2(x_2)\}\), 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
\begin{align*}
\bar{G}_*(y_1, y_2) &= P(Y_1 > y_1, Y_2 > y_2) \\
&= P(-\log G_1(X_1) > y_1, -\log G_2(X_2) > y_2) \\
&= P(X_1 \leq G_1^{-1}(e^{-y_1}), X_2 \leq G_2^{-1}(e^{-y_2})) \\
&= \exp\left\{-(y_1 + y_2)A\left(\frac{y_2}{y_1 + y_2}\right)\right\},
\end{align*}
where "\(*\)" denotes the marginal change in the distribution. The benefit of the new simpler form (beyond the handiness 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
\begin{align*}
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)) \\
&= \frac{\partial^2}{\partial x_1 \partial x_2} \bar{G}_*(T_1(x_1), T_2(x_2)) T'_1(x_1) T'_2(x_2) \\
&\times \left(\left(A(\zeta) + (1 - \zeta)A'(\zeta)\right)\left(A(\zeta) - \zeta A'(\zeta)\right) + \eta A''(\zeta)\right),
\end{align*}
where
\begin{align*}
T_i(x) &= -\log G_i(x) = \left(1 + \xi_i \frac{x - \mu_i}{\sigma_i}\right)^{-\frac{1}{\frac{1}{\xi_i} - 1}}, \; i = 1, 2 \\
T'_i(x) &= -\frac{1}{\sigma_i} \left(1 + \xi_i \frac{x - \mu_i}{\sigma_i}\right)^{-\frac{1}{\frac{1}{\xi_i} - 1} - 1}, \; i = 1, 2 \\
\zeta &= \frac{T_2(x_2)}{T_1(x_1) + T_2(x_2)} \\
\eta &= \frac{T_2(x_2)}{(T_1(x_1) + T_2(x_2))^3}.
\end{align*}
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

\[ A_{\log}(t) = \left( (1 - t)^\alpha + t^\alpha \right)^{1/\alpha}, \quad (6) \]

where \( \alpha \geq 1 \). The independence case corresponds to \( \alpha = 1 \).

\[ A_{\text{asy.log}}(t) = \left( (\theta(1 - t))^\alpha + (\phi t)^\alpha \right)^{1/\alpha} + (\theta - \phi)t + 1 - \theta, \quad (7) \]

where \( \theta \geq 0, \phi \leq 1, \alpha \geq 1 \) 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 \( A(t) \)

![Figure 3: Different classes of parametric dependence functions fitted by \{evd\} package of R for observations from Hamburg and Hanover](image-url)

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
2. Bivariate Extreme Value Models

distribution with mean $1/A(t)$ for any $t \in [0, 1]$, as below

$$P\left(\min\left\{ \frac{Y_1}{1-t}, \frac{Y_2}{t} \right\} > x\right) = P(Y_1 > (1-t)x, Y_2 > tx)$$ (8)

$$= \exp\{-xA(t)\}, x \geq 0.$$ 

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

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

However the estimator has the drawback that itself actually is not dependence function according to (P). By proposing some appropriate marginal adjustment, 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)

$$\frac{1}{\hat{A}_n^{HT}(t)} = \frac{1}{n} \sum_{j=1}^{n} \min\left\{ \frac{Y_{1,j}/\bar{Y}_1}{1-t}, \frac{Y_{2,j}/\bar{Y}_2}{t} \right\},$$ (9)

satisfies $\hat{A}_n^{HT}(0) = \hat{A}_n^{HT}(0) = 1$ 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 $\tilde{A}_n^{HT}$, we obtain an estimator, which already satisfies all the necessary criteria of (P). At the expense of its flexibility even $\tilde{A}_n^{HT}$ does not verify the "extra" property of differentiability, so the BEVD density function is still not available by assuming $\tilde{A}_n^{HT}$ 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 division $0 = t_0 < t_1 < \cdots < t_m = 1$ division of the interval $[0, 1]$, and a given a smoothing parameter $\lambda > 0$, we can take $\hat{A}_3$ to be the polynomial smoothing spline of degree 3 or more which minimizes

$$\sum_{j=0}^{m} \left( \hat{A}_n^{HT}(t_j) - \hat{A}_3(t_j) \right)^2 + \lambda \int_0^1 A_3''(t)^2 \, dt,$$
subject to $\tilde{A}_2(0) = \tilde{A}_3(1) = 1, \tilde{A}'_2(0) \geq -1, \tilde{A}'_3(1) \leq 1$ and $\tilde{A}''_g(t) > 0$ on $[0, 1]$. By solving the non-linear optimization problem above one can 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 $\hat{Y}_{i,j} = -\log \hat{G}_i(X_{i,j})$ for $i = 1, 2$ and $j = 1, ..., n$.

The comparison of the exact (symbolically calculated, see Appendix A) derivatives of the parametric dependence functions and the (numerically 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 component-wise 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 curiosity 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 straightforward computations (for details
Figure 4: Upper block: The first and second derivatives for symmetric logistic ($\alpha = 2.06$) and asymmetric logistic ($\alpha = 2.5, \theta = 0.8, \phi = 0.6$) models. (The choice of $\alpha$ 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 $\lambda$ smoothing parameters.

see Appendix B) and it can be expressed with the terms of (5) as follows

$$h(x_1, x_2) = \frac{T'_1(x_1) T'_2(x_2)}{c_0} \times \eta A''(\zeta),$$

(10)
where

\[ T_i(x) = -\log G_i(x) = \left(1 + \frac{x - \mu_i}{\sigma_i}\right)^{-\frac{1}{\xi_i}}, \quad i = 1, 2 \]

\[ T'_i(x) = -\frac{1}{\sigma_i}\left(1 + \frac{x - \mu_i}{\sigma_i}\right)^{-\frac{1}{\xi_i} - 1}, \quad i = 1, 2 \]

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

\[ \eta = \frac{T_1(x_1)T_2(x_2)}{(T_1(x_1) + T_2(x_2))^2} \]

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

### 2.3 Density Based Prediction Regions for BEVD and BGPD

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 probability level. Following the notation in Hall and Tajvidi (2004) a prediction region is defined as

\[ \hat{R}(u) = \{(x, y) : \hat{h}(x, y) \geq u\} \]

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

for any \( \hat{h} \) estimator of the bivariate density \( h \) as e.g in (5) or (10). Given a prediction level \( \gamma \), let \( u = \hat{u}_\gamma \) denote the solution of the equation \( \beta(u) = \gamma \). Then \( \hat{R}(\hat{u}_\gamma) \) is called a \( \gamma \)-level prediction region for the future value of \((X, Y)\). Throughout the later sections this kind of prediction regions have been used in order to evaluate the fitted models.

### 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 margins 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:

Margins: 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 margins, different parameters have been chosen, specifically $X_1 \sim \text{Exp}(2)$ and $X_2 \sim \text{Exp}(3)$ in the exponential case, and $X_1 \sim \text{GPD}(0.08, 0.13)$ and $X_2 \sim \text{GPD}(0.012, 0.09)$ in the GPD case. (For the GPD distribution the first parameter is the scale, 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 $\tau$ 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 $N = 5000, 10000$ and $20000$ have been used for the simulations, illustrating how the estimations can possibly improve by the increased number of observations.

Prediction levels  High $\gamma$-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.
Figure 5: Snapshot picture of the simulation in the case of exponential margins, linked by Gumbel copula with $\tau = 0.3$, 0.5 and 0.7 Kendall’s correlation. The panels of the left block show the block maxima and their estimated prediction regions by BEVD. The panels of the right block show threshold exceedances and their estimated prediction regions by BGPD. The prediction levels for the regions are $\gamma = 0.99$ (solid line) and $\gamma = 0.95$ (dashed line).
3. Simulations Study

3.1 Performance of BEVD Models

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 \( N = 10000 \) 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 \( \tau \)-values are summarized in Tables 1 and 2 for exponential and GPD margins, respectively.

<table>
<thead>
<tr>
<th>Logistic</th>
<th>( \tau=0.3 )</th>
<th>( \tau=0.5 )</th>
<th>( \tau=0.7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Exp.rate</td>
<td>Obs.rate</td>
<td>St.Err.</td>
</tr>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.052</td>
<td>0.010</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.256</td>
<td>0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hall&amp;Tajvidi</th>
<th>( \tau=0.3 )</th>
<th>( \tau=0.5 )</th>
<th>( \tau=0.7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Exp.rate</td>
<td>Obs.rate</td>
<td>St.Err.</td>
</tr>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.049</td>
<td>0.010</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.261</td>
<td>0.020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asym.Logistic</th>
<th>( \tau=0.3 )</th>
<th>( \tau=0.5 )</th>
<th>( \tau=0.7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Exp.rate</td>
<td>Obs.rate</td>
<td>St.Err.</td>
</tr>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.013</td>
<td>0.009</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.062</td>
<td>0.020</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.277</td>
<td>0.028</td>
</tr>
</tbody>
</table>

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

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 fit we found more bias and higher variance as well. This is not that anyone would expect, as the symmetric model is nested in the asymmetric model with $\theta = \phi = 1$. Therefore it should be at least as good as the symmetric. The explanation behind this phenomena is that the distribution we intend to model by an asymmetric model arises originally from a symmetric distribution (Gumbel copula is exchangeable) and so the additional two parameters are redundant. Due to extra parameters, the estimation gets more complicated to carry out. In practice the maximum likelihood rarely converged and often found just a local maxima. In fact the symmetric logistic model led higher likelihood values, and so, we consider the result of the symmetric model valid for the asymmetric one too. After all we displayed the fit of the asymmetric logistic model as well, assuming the results

### Table 2: Expected and observed rates of 3 BEVD models for simulations with GPD distributed margins and different $\tau$ Kendall’s correlations.

<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>Exp. rate</th>
<th>Obs. rate</th>
<th>St. Err.</th>
<th>Obs. rate</th>
<th>St. Err.</th>
<th>Obs. rate</th>
<th>St. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic $\tau=0.3$</td>
<td>99%</td>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
<td>0.011</td>
<td>0.006</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.052</td>
<td>0.012</td>
<td>0.052</td>
<td>0.011</td>
<td>0.049</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.253</td>
<td>0.019</td>
<td>0.250</td>
<td>0.018</td>
<td>0.253</td>
<td>0.015</td>
</tr>
<tr>
<td>Hall&amp;Tajvidi $\tau=0.3$</td>
<td>99%</td>
<td>0.01</td>
<td>0.011</td>
<td>0.006</td>
<td>0.009</td>
<td>0.006</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.052</td>
<td>0.011</td>
<td>0.048</td>
<td>0.012</td>
<td>0.047</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.256</td>
<td>0.022</td>
<td>0.248</td>
<td>0.021</td>
<td>0.225</td>
<td>0.020</td>
</tr>
<tr>
<td>Asym.Logistic $\tau=0.3$</td>
<td>99%</td>
<td>0.01</td>
<td>0.015</td>
<td>0.015</td>
<td>0.016</td>
<td>0.009</td>
<td>0.023</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.063</td>
<td>0.025</td>
<td>0.071</td>
<td>0.019</td>
<td>0.091</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.282</td>
<td>0.036</td>
<td>0.295</td>
<td>0.035</td>
<td>0.340</td>
<td>0.058</td>
</tr>
</tbody>
</table>
3. Simulations Study

for the symmetric model to be unknown. In addition the choice of margins does not have any significant effect on the model performance. Table 1 and Table 2 confirm mainly the same results.

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 $N = 5000, 10000$ and $20000$. 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 ourselves 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.

We found that the BGPD model performed well when $\tau$ was in the range of medium to strong but there has been a clear bias for $\tau = 0.3$. 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 $\alpha$ 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 $\tau = 0.7$ 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.
### Table 3: Expected and observed rates of logistic BGPD models for simulations with exponential margins, different $\tau$ Kendall's correlations and $N$ sample sizes.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$N$ (Sample Size)</th>
<th>Level</th>
<th>Exp.rate</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>5 000</td>
<td>99%</td>
<td>0.01</td>
<td>0.005</td>
<td>0.006</td>
<td>0.004</td>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.038</td>
<td>0.015</td>
<td>0.037</td>
<td>0.008</td>
<td>0.037</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.280</td>
<td>0.024</td>
<td>0.281</td>
<td>0.015</td>
<td>0.281</td>
<td>0.011</td>
</tr>
<tr>
<td>0.5</td>
<td>5 000</td>
<td>99%</td>
<td>0.01</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.004</td>
<td>0.007</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.046</td>
<td>0.015</td>
<td>0.046</td>
<td>0.009</td>
<td>0.043</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.258</td>
<td>0.020</td>
<td>0.257</td>
<td>0.016</td>
<td>0.258</td>
<td>0.011</td>
</tr>
<tr>
<td>0.7</td>
<td>5 000</td>
<td>99%</td>
<td>0.01</td>
<td>0.010</td>
<td>0.008</td>
<td>0.010</td>
<td>0.004</td>
<td>0.011</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95%</td>
<td>0.05</td>
<td>0.050</td>
<td>0.015</td>
<td>0.051</td>
<td>0.009</td>
<td>0.051</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
<td>0.25</td>
<td>0.251</td>
<td>0.022</td>
<td>0.253</td>
<td>0.016</td>
<td>0.251</td>
<td>0.010</td>
</tr>
</tbody>
</table>

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).

By comparing the above tables we can conclude that BEVD model provided 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
3. Simulations Study

### BEVD vs. BGPD: Exponential margins and Gumbel copula

<table>
<thead>
<tr>
<th>Level</th>
<th>Exp.rate</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
<td>0.011</td>
<td>0.005</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.052</td>
<td>0.010</td>
<td>0.052</td>
<td>0.013</td>
<td>0.052</td>
<td>0.011</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.256</td>
<td>0.015</td>
<td>0.253</td>
<td>0.017</td>
<td>0.252</td>
<td>0.018</td>
</tr>
</tbody>
</table>

### BEVD vs. BGPD: GPD margins and Gumbel copula

<table>
<thead>
<tr>
<th>Level</th>
<th>Exp.rate</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.004</td>
<td>0.004</td>
<td>0.008</td>
<td>0.004</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.037</td>
<td>0.008</td>
<td>0.046</td>
<td>0.009</td>
<td>0.051</td>
<td>0.009</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.281</td>
<td>0.015</td>
<td>0.257</td>
<td>0.016</td>
<td>0.253</td>
<td>0.016</td>
</tr>
</tbody>
</table>

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

### BEVD vs. BGPD: GPD margins and Gumbel copula

<table>
<thead>
<tr>
<th>Level</th>
<th>Exp.rate</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
<td>0.011</td>
<td>0.006</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>0.052</td>
<td>0.012</td>
<td>0.052</td>
<td>0.011</td>
<td>0.049</td>
<td>0.011</td>
</tr>
<tr>
<td>75%</td>
<td>0.25</td>
<td>0.253</td>
<td>0.019</td>
<td>0.250</td>
<td>0.018</td>
<td>0.253</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 5: Expected and observed rates of logistic BEVD and logistic BGPD models for simulations with GPD distributed margins

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 $\tau = 0.5$ 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 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
BEVD vs. BGPD: Exponential margins and different copulas

<table>
<thead>
<tr>
<th>Copula</th>
<th>BEVD</th>
<th>Obs.rate</th>
<th>St.Err.</th>
<th>Obs.rate</th>
<th>St.Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gumbel</strong></td>
<td>99%</td>
<td>0.011</td>
<td>0.005</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>95%</td>
<td>0.052</td>
<td>0.013</td>
<td>0.046</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>0.253</td>
<td>0.017</td>
<td>0.257</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td><strong>Clayton</strong></td>
<td>99%</td>
<td>0.010</td>
<td>0.005</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td>95%</td>
<td>0.049</td>
<td>0.010</td>
<td>0.047</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>0.252</td>
<td>0.018</td>
<td>0.251</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td><strong>Student</strong></td>
<td>99%</td>
<td>0.009</td>
<td>0.006</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td>95%</td>
<td>0.048</td>
<td>0.010</td>
<td>0.039</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>0.254</td>
<td>0.020</td>
<td>0.269</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td><strong>Gaussian</strong></td>
<td>99%</td>
<td>0.008</td>
<td>0.005</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>95%</td>
<td>0.046</td>
<td>0.011</td>
<td>0.031</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>0.254</td>
<td>0.018</td>
<td>0.268</td>
<td>0.013</td>
<td></td>
</tr>
</tbody>
</table>

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 $\tau = 0.5$ Kendall's correlation

have been linked by Clayton copula, then also the extremes of their margins are supposed to have strong association. In this case the BGPD has been fairly accurate. However, the accuracy diminishes 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

The prediction methods discussed above have been applied to a 2-dimensional wind speed dataset from north of Germany. The observations have been measured hourly for the recent 50 years observed (from 1958 till 2007). Our interest was in joint modeling of the extreme wind speed in two cities: Hamburg and Hanover. To this end "daily" observations have been used, defined as the daily maxima of the hourly observations. The scatter plot of the data is displayed on Figure 2. Due to the relatively short distance between the cities the measurements are rather strongly correlated. The Kendall’s correlation is $\tau = 0.588$ which can be rated to somewhere between the medium ($\tau = 0.5$) and strong ($\tau = 0.7$) association level according to the simulation study cases. As discussed in the previous section, prediction regions based on both BEVD and BGPD model hold their nominal level for this degree of association. Motivated by this fact, we fitted both of these models to the wind speed data and investigated the constructed prediction regions based on the methodology which was outlined in Section 2.3.

4.1 Prediction by BEVD

For the monthly maxima different prediction regions based on the logistic, asymmetric logistic and the spline smoothed Hall-Tajvidi estimators have been estimated. In the last case the choice of smoothing parameter $\lambda$ can have a considerable effect on the estimated curves. We investigation this in more details and the results are summarized in Appendix D. Here we present only the two alternatives which are arguably the most reasonable choices for this practical case. The curves of the prediction regions (on the original scale) are presented in Figure 6. This makes the models visually comparable, 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 $\gamma$ levels.

As expected, the results are different and seem to depend on $\gamma$-values so this kind of comparison is consequently not enough to decide which one is best alternative. Even if a fitted model was very close to the observation for this given “realization” of the wind speeds, it might not be generally the best choice from prediction point of view. In order to see how these models perform we need some further investigations. Since our focus is on the accuracy of prediction re-
A Logistic model

Hamburg m/s

Hannover m/s

5 10 15 20

5 10 15 20 25 30

α = 2.06

Asym. Logistic model

Hamburg m/s

Hannover m/s

5 10 15 20

5 10 15 20 25 30

θ = 1

φ = 1

α = 2.06

Hall–Tajvidi model

Hamburg m/s

Hannover m/s

5 10 15 20

10 15 20 25 30

λ = 0.08

Hall–Tajvidi model

Hamburg m/s

Hannover m/s

5 10 15 20

10 15 20 25 30

λ = 0.16

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>
<thead>
<tr>
<th>BEVD Models</th>
<th>Logistic</th>
<th>Hall–Tajvidi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Expect.</td>
<td>λ=0.08</td>
</tr>
<tr>
<td>50%</td>
<td>296</td>
<td>316</td>
</tr>
<tr>
<td>75%</td>
<td>148</td>
<td>149</td>
</tr>
<tr>
<td>95%</td>
<td>≈30</td>
<td>34</td>
</tr>
<tr>
<td>99%</td>
<td>≈6</td>
<td>15</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-2531.5</td>
<td>-2533.8</td>
</tr>
</tbody>
</table>

Table 7: Performance of different models at different predictive levels
4. Application to Wind Speed Data

regions (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 split 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 wind speed 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)

As an automatic algorithm was needed for the cross validation method, we started the optimization for the likelihood maximization from parameter vector \((\hat{\xi}_1, \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_2, \hat{\mu}_2, \hat{\sigma}_2, \hat{\alpha}_0)\) given by \(\hat{G}_1, \hat{G}_2\) the univariate EVD estimates for the margins and fixed \(\hat{\alpha}_0 = 2\). The results are presented in Table 8. We have been faced with the same effect as in the case of the simulation study, namely that maximum-likelihood estimation procedure of the asymmetric logistic model usually returned back lower likelihood value than its simple symmetric version. (Technically, the maximum-likelihood estimation has been started from different initial values, and the symmetric turned out to be constantly the best.) Therefore we could conclude that the difference between the symmetric and asymmetric logistic models is negligible in our case. (The asymmetry parameters of (7) are practically \(\theta = \phi = 1\), coinciding with the case of symmetry.) Generally the prediction regions perform acceptably well at \(\gamma = 0.5, 0.75\) and 0.95 prediction levels, perhaps with slightly under-estimation of the quantiles. At highest \(\gamma = 0.99\) level the bias seems to be more serious, in average there are 2 times more observations outside the region than the expected.

4.2 Prediction by BGPD

The prediction regions have been calculated also for the exceedances. First we considered the density given in (10) with a logistic dependence function which has been inherited from the block maxima method and has been kept fixed \(\tau = 2.06\) during the maximum likelihood optimization. Practically by doing this we just adjusted the 6 marginal parameters to an existing dependence function. After this
Cross-validation of BEVD

<table>
<thead>
<tr>
<th>Level</th>
<th>Expect.</th>
<th>Symm./Asym.</th>
<th>( \lambda = 0.08 )</th>
<th>( \lambda = 0.16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>148</td>
<td>154.2(7.3)</td>
<td>156.4(6.6)</td>
<td>155.2(6.6)</td>
</tr>
<tr>
<td>75%</td>
<td>74</td>
<td>71.0(6.4)</td>
<td>73.2(6.3)</td>
<td>71.2(6.0)</td>
</tr>
<tr>
<td>95%</td>
<td>( \approx 15 )</td>
<td>16.3(2.6)</td>
<td>17.3(2.8)</td>
<td>16.3(2.9)</td>
</tr>
<tr>
<td>99%</td>
<td>( \approx 3 )</td>
<td>5.9(1.7)</td>
<td>5.9(1.7)</td>
<td>5.8(1.7)</td>
</tr>
</tbody>
</table>

Table 8: Cross-validation

step we also let the dependence parameter be free and optimized the maximum likelihood for all 7 parameters. Finally asymmetry parameters \( \theta, \phi \) 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 had to be interrupted when there was no significant improvement in the likelihood \((-2202.5)\) value after a large number of iterations. 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 asymmetric 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 summarized in Table 9.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>264</td>
<td>272</td>
<td>278</td>
<td>282</td>
</tr>
<tr>
<td>75%</td>
<td>132</td>
<td>123</td>
<td>119</td>
<td>128</td>
</tr>
<tr>
<td>95%</td>
<td>( \approx 26 )</td>
<td>26</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>99%</td>
<td>( \approx 5 )</td>
<td>8</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Log-likelihood: -2202.5, -2187.8, -2165.6

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

Similar to the block maxima method cross-validation has been carried out for
4. Application to Wind Speed Data

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

The logistic BGPD model as well. In contrast to BEVD model no obvious underestimation were found and the estimated quantiles have been more appropriate compared with the observations, see Table 10 for further notes.

<table>
<thead>
<tr>
<th>Cross-validation of BGPD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic BGPD Model</strong></td>
</tr>
<tr>
<td>Level</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>75%</td>
</tr>
<tr>
<td>95%</td>
</tr>
<tr>
<td>99%</td>
</tr>
</tbody>
</table>

Table 10: Cross-validation for logistic BGPD
Figure 8: Prediction regions estimated by symmetric and asymmetric logistic BGPD at high prediction levels, $\gamma = 0.99$ (solid line) and $\gamma = 0.95$ (dashed line)

Acknowledgments

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


Appendices

A First and second derivatives of two parametric dependence functions

\[ A_{log}'(t) = \left( (1 - t)\alpha + \xi^\alpha \right)^{\frac{1}{\alpha} - 1} \left( (1 - t)^{\alpha - 1} + \xi^{\alpha - 1} \right) \]

\[ A_{log}''(t) = (1 + \beta) \left( (1 - t)\alpha + \xi^\alpha \right)^{\frac{1}{\alpha} - 2} \left( (1 - t)^{\alpha - 2} + \xi^{\alpha - 2} \right) \]

\[ + (1 + \beta) \left( (1 - t)\alpha + \xi^\alpha \right)^{\frac{1}{\alpha} - 1} \left( (1 - t)^{\alpha - 2} + \xi^{\alpha - 2} \right) \]

\[ A_{aleg}(t) = \left( \frac{\theta(1 - t)}{1 - t} \right)^{\alpha - 1} \left( \frac{1}{1 - t} + \frac{\phi t}{t} \right) + \theta - \phi \]

\[ A_{aleg}''(t) = (1 + \beta) \left( \frac{\theta(1 - t)}{1 - t} \right)^{\alpha - 2} \left( \frac{1}{1 - t} + \frac{\phi t}{t} \right)^2 \]

\[ + (1 + \beta) \left( \frac{\theta(1 - t)}{1 - t} \right)^{\alpha - 1} \left( \frac{1}{1 - t} + \frac{\phi t}{t} \right) \]

\[ + (1 + \beta) \left( \frac{\theta(1 - t)}{1 - t} \right)^{\alpha - 1} \left( \frac{1}{1 - t} + \frac{\phi t}{t} \right) \]

\[ + \frac{1}{(1 - t)^2} + \frac{\phi t}{t^2} \]
B Calculation of BGPD Density

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 \( \frac{\partial^2 G(x,y)}{\partial x \partial y} = 0 \). Taking this into account we see that the density is

\[
 b(x, y) = \frac{\partial^2 H(x, y)}{\partial x \partial y} = \frac{\partial^2}{\partial x \partial y} \left( 1 - \frac{\log G(x, y)}{\log G(0, 0)} \right) = C_0 \frac{T_1(x)T_2(y)}{(T_1(x) + T_2(y))^3} A'' T_1'(x)T_2'(y),
\]

Taking the usual marginal transformations we get

\[
 -\log G(x, y) = -\log G_1(-\log G_1(x), -\log G_2(y)) = -\log G_n(T_1(x), T_2(y)) = -\log G_n(t_1, t_2) = -\log \{\exp(-t_1 + t_2)A(\frac{t_1}{t_1 + t_2})\}
\]

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

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

\[
 \frac{\partial^2}{\partial t_1 \partial t_2} \left( (t_1 + t_2) A(\frac{t_2}{t_1 + t_2}) \right) = \frac{\partial}{\partial t_2} \left( A(\mu) + (t_1 + t_2)A'(\mu) \frac{-t_2}{(t_1 + t_2)^2} \right) = \frac{\partial}{\partial t_2} \left( A(\mu) - \mu A'(\mu) \right) = A'(\mu) \mu' - A'(\mu) - \mu A''(\mu) \mu' = -\frac{t_1 t_2}{(t_1 + t_2)^3} \mu''(\mu).
\]

C Simulation results for different copula families

Snapshot pictures of simulations assuming exponential margins, linked by Clayton, Student-t and Gauss copula with \( \tau = 0.5 \) 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 \( \gamma = 0.99 \) (solid line) and \( \gamma = 0.95 \) (dashed line).
Figure 9: Prediction regions for simulations from exponential margins linked by Clayton, Student-t and Gauss copula with $\tau = 0.5$ Kendall’s correlation. The prediction levels for the regions are $\gamma = 0.99$ (solid line) and $\gamma = 0.95$ (dashed line).
D Effect of smoothing parameter on H&T estimator

Figure 10: The effect of $\lambda$ smoothing parameter on the maximum likelihood values of BEVD assuming H&T estimators

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.0025</th>
<th>0.005</th>
<th>0.01</th>
<th>0.02</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>320</td>
<td>317</td>
<td>317</td>
<td>317</td>
<td>317</td>
</tr>
<tr>
<td>75%</td>
<td>158</td>
<td>158</td>
<td>158</td>
<td>158</td>
<td>153</td>
</tr>
<tr>
<td>95%</td>
<td>39</td>
<td>39</td>
<td>38</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>99%</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>-Loglik.</td>
<td>2536.2</td>
<td>2536.0</td>
<td>2535.7</td>
<td>2535.1</td>
<td>2534.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.06</th>
<th>0.08</th>
<th>0.1</th>
<th>0.12</th>
<th>0.16</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>317</td>
<td>316</td>
<td>313</td>
<td>310</td>
<td>309</td>
</tr>
<tr>
<td>75%</td>
<td>151</td>
<td>149</td>
<td>144</td>
<td>144</td>
<td>140</td>
</tr>
<tr>
<td>95%</td>
<td>35</td>
<td>34</td>
<td>34</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>99%</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>-Loglik.</td>
<td>2534</td>
<td>2533.8</td>
<td>2533.8</td>
<td>2534</td>
<td>2534.6</td>
</tr>
</tbody>
</table>

Table 11: The effect of smoothing parameter $\lambda$ on the prediction regions constructed by non-parametric estimator of dependence function by Hall&Tajvidi
B
Abstract

Recent years have seen more applications of copula models aimed at describing the dependence of multivariate data sets. However, there is no generally accepted method for checking the validity of the fitted copula model, especially for the case of higher dimensions. The most recent literature provides goodness of fit procedures which, seemingly, are easily extensible to any dimensions, but in practice many problems are coming up with the increased dimensionality. The majority of the reviews or comparative studies investigate examples only in the bivariate case. Even the very prudent work of Genest et al. (2009) is limited to bivariate simulations. To the best of our knowledge the papers which compare performance of goodness of fit tests in higher dimension are Berg and Bakken (2006) and Berg (2009), but there is no practical application suggested. In this paper we present different approaches for three dimensional copula models with particular focus on jointly occurring extreme values. One alternative approach is based on Kendall’s transform which reduces the multivariate problem to one dimension. Another one uses the Rosenblatt’s transform, which was recently re-invented for this testing purpose, and basically converts the problem into testing for independence.
Although the approaches we apply are not completely new, we suggest specific weights for test statistics in order to increase the efficiency of the tests for the extremes. We apply these methods to explore a three dimensional environmental data set, namely 50 years of daily maximal wind speed measurements for three German cities.

**Key words:** copulas, goodness-of-fit test, probability integral transformation, Kendall’s transform, Rosenblatt’s transform, wind speed maxima

## 1 Introduction

In the last decade the question of multivariate modeling of high-dimensional data has become also tractable, mainly due to the vast number of recorded data and the powerful computing equipment readily available. However, the methodology has not always been kept pace with the available resources. For instance, one can easily fit many multivariate models by one software or another, but often there is no really suitable method for checking the goodness of the fit. In the model fitting aspect the rapidly developing open-source R package plays a leading role. Particularly, for modeling dependence structures of multivariate extremes the \{copula\} and \{evd\} packages are available. In this paper we focus on checking the validity of the copula models. Copulas are simple but yet powerful tools for modeling, which ensure the separation of marginal modeling and dependence structure. They have been re-invented in the 1990s and their use has been expanded rapidly since then. One natural area of their applications is in the financial mathematics, where they are often used to model the dependence structure of assets or losses, stock indices and so on. Here we apply the methodology to another, equally important field, namely the environmental data.

In Section 2 we first briefly review the needed elements of copula theory and present the notations. In Section 3 we summarize the most recent approaches for measuring the goodness of fit (GoF) for copula models. We suggest some practical modifications of the test statistics as our focus is on the joint behavior of extreme events, and we intend to reach more sensitivity in that respect. In the higher dimensional cases the computational requirements can easily blow up, so we need some effective dimension reducing methods. One of our proposed GoF tests is based on the Kendall’s transform of the joint distribution (see Genest et al. (2006)), which reduces the multivariate problem to one dimension. Further, we suggest an alternative testing methodology aimed at revealing differences in
2. Copula Concepts

We also show some methods based on the Rosenblatt’s transform (often called conditional probability integral transform), which converts the GoF problem into checking the hypothesis of independence for the transformed variables. However, one may see that testing “only” the independence is not an obvious task either. We present a recent approach of Berg and Bakken (2006), derived originally from the approach of Breyman et al. (2003), where an additional transform is performed after Rosenblatt’s to ensure consistency of the method. We compare different possible weight functions that allow us to concentrate on discrepancies occurred in the joint extremes. In Section 4 we apply the presented methods to the dependence structure of a real wind data set. Here, we start the analysis with methods offered by the \{copula\} package of the R for preliminary understanding of the problems. Finally, we give some conclusions in Section 5.

2 Copula Concepts

Consider a continuous random vector \( \mathbf{X} = (X_1, \ldots, X_d) \) with joint distribution function \( \mathbf{H} \) and margins \( F_1(x_1), \ldots, F_d(x_d) \). Due to Sklar’s theorem we know that to any \( d \)-variate distribution function \( \mathbf{H} \), with univariate margins \( F_i \) there exits a copula \( \mathbf{C} \), a distribution over the \( d \)-dimensional unit cube, with uniform margins, such that \( \mathbf{H}(x_1, \ldots, x_d) = \mathbf{C}(F_1(x_1), \ldots, F_d(x_d)) \). Moreover the copula \( \mathbf{C} \) is unique if the marginal distributions are continuous. This fact allows us to capture the dependence structure without specifying the marginal distributions. In the recent literature various families of copulas have been introduced, for an overview and examples see the textbooks of Cherubini et al. (2004), Joe (1997), McNeil et al. (2005) and Nelsen (2006).

2.1 Elliptical Copulas

Elliptical copulas are simply the copulas of elliptical distributions as multivariate Gaussian or Student-\( t \) distributions. The main advantage of this class is that one can specify different levels of correlation between the margins. Unfortunately they do not have closed form expressions and are restricted to have radial symmetry. The Gaussian copula family can simply be derived from the multivariate Gaussian distribution function with mean zero and correlation matrix \( \Sigma \), transforming the
margins by the inverse of the standard normal distribution function $\Phi$ as

$$
C^-_{\Sigma}(u) = \Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)
$$

This model is suitable for the cases where there is no reason to assume dependence among the extremes. Roughly speaking this means if we observe a high value at one margin, the values of other margins are not likely to be high at the same time. For further details of the concept of tail dependence measures see Coles et al. (1999). Another member of the elliptical copula family, which is called the Student’s $t$ copula, does not concern this kind of independence at the tails. Its structure is similar to (1) but the Gaussian distributions are replaced by $t$-distributions in the formula. The Student’s $t$ copula is defined as

$$
C^T_{\Sigma}(u) = t^{-1}_d(u_1), \ldots, t^{-1}_d(u_d))
$$

where $\nu$ is the number of degrees of freedom.

### 2.2 Archimedean Copulas

Another broad class of copulas is called the Archimedean copula-family. It is also used frequently due to its very convenient structure. Let us consider a copula generator function: $\phi(u) : [0, 1] \to [0, \infty]$, which is continuous and strictly decreasing with $\phi(1) = 0$. Then a $d$-variate Archimedean copula function is

$$
C_{\phi}(u) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right).
$$

It has a very simple construction but it also suffers from considerable limitations. There is only one parameter (or just few) to capture the entire dependence structure and by its construction all of these copulas are permutation symmetric (exchangeable). This means that all $s < d$ dimensional margins are identically distributed, which is usually a too strict assumption in real life. In the course of the next sections we will be limited to present the Clayton and Gumbel copula.
family, but we emphasize that the presented methods can be adapted to any Archimedean models exactly in the same way. The Gumbel copula has the generator 
\[ \phi_\theta(u) = [-\ln(u)]^\theta, \]
where \( \theta \in [1, +\infty) \). Thus, the Gumbel \( d \)-copula function is given by
\[
C_{\text{Gumbel}}(u) = e^{-\left(\sum_{i=1}^{d} - \log(u_i)^\theta\right)^\frac{1}{\theta}}.
\]

We should notice that the Gumbel copula belongs to another important family too. A copula \( C \) is called \textit{extreme value copula} if \( C(u_1^t, \ldots, u_d^t) = C'(u_1, \ldots, u_d) \) for all \( t > 0 \). This family consists of copulas of multivariate extreme value distributions. The Gumbel copula is the only Archimedean copula, which is included in the extreme value copula family at the same time and actually it coincides with the logistic dependence structure. The generator function of the Clayton copula (also known as Cook and Johnson’s family) is given by \( \phi_\theta(u) = u^{-\theta} - 1 \), where \( \theta > 0 \). Thus, the Clayton \( d \)-copula function is the following
\[
C_{\text{Clayton}}(u) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1\right)^{-\frac{1}{\theta}}.
\]

For parameter estimation the pseudo-maximum likelihood estimation is the most widely accepted method in the above cases. For more details and for simulation methods see the Chapter 5-6 in Cherubini et al. (2004). An illustration of the 4 described copula models is shown on Figure 1, where scatter plots of 2 dimensional simulations with given sample size (\( n=600 \)) and given strength of association (Kendall’s \( \tau = 0.5 \)) are displayed.

3 Goodness-of-Fit Tests in 3D

After estimating the model parameters one must be able to check the GoF. Formally we intend to test the hypothesis
\[
\mathcal{H}_0 : C \in \mathcal{C}_0 = \{C_\theta, \theta \in \Theta\},
\]
that the dependence structure of the copula arises from a specific parametric family \( \mathcal{C}_0 \) of copulas. The most obvious way for testing GoF is to consider multidimensional \( \chi^2 \) approaches, but in this case we need to discretize the data, losing
valuable information. For high dimensional data, gridding yields substantial computational difficulties too. In order to avoid the use of such methods we would rather choose some handy dimension reducing methods for our purposes. As usual in this context we consider the $F_i$ marginal distributions as nuisance parameters and base all of the tests on ranks. Basically in a preliminary step we perform the probability integral transformation (PIT) for the observations mapping them into the $d$-dimensional unit cube as

$$X_i = (X_{i1}, \ldots, X_{id}) \sim H \rightarrow_{\text{PIT}} U_i = (U_{i1}, \ldots, U_{id}) \sim C, \text{ for } i = 1, \ldots, n.$$
3. Goodness-of-Fit Tests in 3D

otherwise the testing procedures fail to hold their nominal level. Another relevant issue mentioned is the option of choosing suitable weights. We do believe that this is an important option, since statisticians must be able to choose additional parameters for their tests if they aim to focus on a given region of the tested distribution. While their detailed investigation is carried out for the bivariate case, we go one step further and present three dimensional tests. Although this step does not really look like a major one, there are different new effects arise. First of all, we focus on cdf-based methods and exclude methods based on multivariate kernel density estimation such as it is proposed in Fermanian (2005) and Scaillet (2007). In higher dimension the latter techniques are extremely computation exhaustive and also because our given sample size is too small for them to be really effective.

3.1 Univariate GoF Tests

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

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

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

\[ T_{GoM} = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x) \]  

(3)

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

\[ T_{AD} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1-F(x))} dF(x), \]  

(4)

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

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

(5)
for the case of maximum and with \( \Phi(x) = F(x) \) in the place of \( \Phi(x) = (1 - F(x)) \) for minimum. The advantage of (4) in comparison to (3) is that its sensitivity is concentrated to discrepancies at the relevant tail of the distribution see Zempléni’s test in Kotz and Nadarajah (2000), p.77. The computation of these statistics is straightforward and for arbitrary weight function \( \Phi(x) \) can be numerically approximated.

\[ (4) \]

3.2 Empirical Copula Process

The information that the pseudo-observations \( U_1, ..., U_n \) carry is well captured by their empirical distribution

\[
C_n(u) = \frac{1}{n} \sum_{i=1}^{n} 1(U_{i1} \leq u_1, ..., U_{id} \leq u_d)
\]

for any \( u = (u_1, ..., u_d) \in [0, 1]^d \). This \( C_n \) is usually called the "empirical copula", even if it is not exactly a copula. Under certain conditions (see Fermanian (2004)) it is a consistent estimator of the true underlying copula \( C \), which is tested to be in \( C_0 = \{ C_\theta, \theta \in \Theta \} \) according to \( H_0 \). So in order to check \( H_0 \) a very natural approach can be obtained by measuring the distance in some sense between the \( C_n \) and the fitted element of the assumed copula family \( C_\theta \). Formally we can base a GoF test on the so-called empirical copula process

\[
C_n = \sqrt{n}(C_n - C_\theta)
\]

possibly using e.g. Cramér-von Mises statistics, analogously to the univariate case, by the following statistics

\[
G_n = \int_{[0,1]^d} C_n(u)^2 dC_n(u).
\]

As \( G_n \) represents kind of a distance between the true and observed copula, its large values lead to the rejection of \( H_0 \). The method has been recently implemented in the \{copula\} package of R. It can be found under the command \texttt{gofCopula}, where in default the approximate \( p \)-values for the test statistics are obtained using the parametric bootstrap. Further details can be found in the papers of Genest and Remillard (2008) and Genest et al. (2009). As the method is fairly time consuming, there are options for speeding it up by using the fast multiplier approach (described in Kojadinovic and Yan (2008a, 2008b)) instead of slow parametric bootstrapping. We call this test later on a G-test referring to the authors.
3. Goodness-of-Fit Tests in 3D

3.3 Rosenblatt’s Transform

There is a well-known mapping due to Rosenblatt (1952) on which a GoF procedure could be based. The Rosenblatt’s (probability integral) transform decomposes a random vector into mutually independent components having uniformly distributed margins on the unit interval. Let \( R : (0, 1)^d \rightarrow (0, 1)^d \) be defined for a copula \( C \) and for \( u = (u_1, ..., u_d) \in (0, 1)^d \) as \( R(u) = (e_1, ..., e_d) \), where \( e_1 = u_1 \) and for \( i \geq 2 \)

\[
e_i = \frac{\partial^{i-1} C(u_1, ..., u_i, 1, 1, ..., 1)}{\partial u_1 \cdots \partial u_{i-1}} / \frac{\partial^{i-1} C(u_1, ..., u_{i-1}, 1, 1, ..., 1)}{\partial u_1 \cdots \partial u_{i-1}}.
\]

In this setting \( U \) is distributed as \( C \) if and only if the copula of \( R(U) \) is the independence copula. This means that the original problem of testing the model fit can be converted to testing independence. Technically (1) is equivalent to

\[ H_0^* : R(\theta) \sim \Pi_d, \]

where \( \Pi_d \) is the \( d \)-dimensional independence (product) copula. A promising approach by Breymann et al. (2003) proposes dimensional reduction, considering that the transformed pseudo-observations \( E_1, ..., E_n \) of \( E = R(U) \) are uniformly distributed on \( U[0, 1]^d \) under \( H_0 \). Therefore instead of testing \( H_0 \) one can test \( H_0^* \) after the following transformation.

For the dimension reduction let \( \Phi^{-1} \) denote the inverse distribution function of a standard \( \mathcal{N}(0, 1) \) random variable so the variables \( \Phi^{-1}(E_i) \) for \( i = 1, ..., d \) are i.i.d. \( \mathcal{N}(0, 1) \). By taking the sum of squares of them

\[ Y_B = \sum_{i=1}^d \Phi^{-1}(E_i)^2, \]

we must “ideally” get a \( \chi^2_d \) distributed \( Y_B \). Define \( W_B = F_{\chi^2_d}(Y_B) \), which is an i.i.d. \( \mathcal{U}(0, 1) \) vector under \( H_0 \). Finally let the distribution function of \( W_B \) be denoted by

\[ F_B(w) = P(W_B \leq w), w \in [0, 1]. \]
Since under the null-hypothesis $F_B(w) = w$ one can get a reasonable test based on the comparison of $F_B$ with its empirical counterpart $\hat{F}_B$ computed from the observations. This univariate GoF test can already be carried out by (2), (3) or (4). Although there are some problematic issues with the approach, since due to the transformation the weights do not have the same effect as in the usual univariate setup. Some important remarks on the method:

1. The word "ideally" is emphasized above as the basic assumptions are too optimistic according to the fact that the pseudos are not exactly mutually independent and they are only approximately uniformly distributed in the unit cube. However the idea of dimension reduction by (7) is appealing and the method can be further refined by approximating the real distribution of $Y_B$ by simulation.

2. Another weakness is the inconsistency in the sense that the resulted test statistics are not strictly increasing for every deviation from the null hypothesis. This phenomena is caused by the projection in (7).

3. If there is a particular interest in the tail of the copula there is no option offered to be more specifically focused on it.

Berg and Bakken (2005) proposed a solution to these questions. As an extension of the above method proposed by Breyman et al. (2003), they defined a additional transformation before the dimension reduction, which tackles the inconsistency issue arising from the projection to one dimension. Let $B : (0, 1)^d \rightarrow (0, 1)^d$ be defined for $e = (e_1, ..., e_d) \in (0, 1)^d$ as $B(e) = (e_1^*, ..., e_d^*)$, where $e_1^* = e_1$ and for $i \geq 2$

$$e_i^* = 1 - \left( \frac{1 - \hat{e}_i}{1 - \hat{e}_{i-1}} \right)^{d-(i-1)},$$

(8)

where $\hat{e} = (\hat{e}_1, \hat{e}_2, ..., \hat{e}_d)$ is the sorted counterpart of $e$. So the procedure can be presented as follows

\[
\begin{align*}
\text{Rosenblatt's transform} & \quad \begin{array}{l}
\tilde{E}_i = (E_{i1}, ..., E_{id}) \sim U[0, 1]^d \\
\end{array} \\
\text{Berg&Bakken transform} & \quad \begin{array}{l}
\hat{E}_i^* = (E_{i1}^*, ..., E_{id}^*) \sim U[0, 1]^d \\
\end{array}
\end{align*}
\]

$64$
Modify (7) by replacing the "new" coordinates from $E^* = B(E)$ into the formula and also by adding a weight function as follows

$$Y_B^* = \sum_{i=1}^{d} \gamma(E_i; \omega) \Phi^{-1}(E_i^*)^2,$$

where $\gamma$ is a weight function and $\omega$ is the set of the weighting parameters. Further let $F_{Y_B^*}$ be the cdf. of $Y_B^*$, which is not assumed to be $\chi^2$ anymore, but can be obtained by simulation. After this step we proceed analogously to the original idea and define $F_{Y_B^*}$ be the cdf. of $Y_B^*$, which is not assumed to be $\chi^2$ anymore, but can be obtained by simulation. After this step we proceed analogously to the original idea and define $W_B^*$ as an i.i.d. $U(0,1)$ vector under $H_0$.

Finally let the cdf. of $W_B^*$ denoted by $F_{W_B^*}$, which is an i.i.d. $U(0,1)$ vector under $H_0$. The B-test can similarly be defined as above by the comparison of $F_{W_B^*}$ to its empirical counterpart $\hat{F}_{W_B^*}$. The effect of the permutation order for different methods based on Rosenblatt’s transform is investigated in Berg (2009) and found that although there is an extra variation added in $p$-values of the those tests which are using weight functions, the practical consequence is negligible.

### 3.4 Kendall’s Transform

Other choice for GoF statistics can be based on the Kendall’s transform, which is the distribution function of the probability integral transformation of the joint distribution

$$K(\theta, t) = P(C_{d}(F_1(X_1), ..., F_d(X_d)) \leq t) = P(C_{d}(U_1, ..., U_d)) \leq t).$$

In the case of Archimedean copula family, (5) can be computed as follows

$$K(\theta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \left[ \phi_{\theta}(t) \right] f_i(\theta, t),$$

where $f_i(\theta, t) = \left. \frac{\partial}{\partial \theta} \phi_{\theta}^{-1}(x) \right|_{x=\phi_{\theta}(t)}$. Note that actually $f_{i+1}(\theta, t) = f_i(\theta, t) \frac{\partial}{\partial \theta} f_i(\theta, t)$, $i \in \{1, ..., d-1\}$. The $K$ function defined this way is invariant on the marginal distributions, hence it depends only on the copula of $X$. The empirical version of $K$ can be computed by the rank based pseudo-observations as

$$K_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1(E_{in} \leq t), t \in [0,1],$$
Figure 2: Comparison between parametric $K_{\theta}$ from the 3D Gumbel family with different dependence parameters and the empirical version $K_n$ for the wind data where

$$E_{in} = \frac{1}{n} \sum_{j=1}^{n} 1(U_{i1} \leq U_{j1}, ..., U_{jd} \leq U_{jd}).$$

For illustrations see Figure 2.

Known tests for checking the match of the theoretical and empirical version of the Kendall’s transform $K$ use continuous functionals of Kendall’s process $x_n(t) = \sqrt{n}(K(\theta_n, t) - K_n(t))$ having favorable asymptotic properties. There are two different kind of approaches investigated in Genest et al. (2006), Cramér-von Mises type $S_n = \int_0^1 (x_n(t))^2 \, dt$ and Kolmogorov-Smirnov type $T_n = \sup_{0 \leq t \leq 1} |x_n(t)|$ statistics. As the second approach is proved to be generally less powerful in detecting discrepancies near the tails, we based our inference on the test statistics (analogously to (2), (3) and (4)) summarized by Table 1, where $(t_i)_{i=1}^m$ is an appropriately fine division of the interval (0, 1).

Moreover, it is a natural idea to combine this approach with the Rosenblatt’s transform, so testing the independence among the Rosenblatt coordinates by the K-test. The main advantage of doing this is having the same Kendall’s transform...
4. Application to Wind Speed Maxima

<table>
<thead>
<tr>
<th>Focused Regions</th>
<th>Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>$K_1 = \frac{1}{n} \sum_{t_i \in [0+\epsilon, 1-\epsilon]} (K(\theta_n, t_i) - K_\theta(t_i))^2$</td>
</tr>
<tr>
<td>Upper Tail</td>
<td>$K_2 = \frac{1}{n} \sum_{t_i \in [0+\epsilon, 1-\epsilon]} \frac{(K(\theta_n, t_i) - K_\theta(t_i))^2}{1-K_\theta(t_i)}$</td>
</tr>
<tr>
<td>Lower Tail</td>
<td>$K_3 = \frac{1}{n} \sum_{t_i \in [0+\epsilon, 1-\epsilon]} \frac{(K(\theta_n, t_i) - K_\theta(t_i))^2}{K_\theta(t_i)}$</td>
</tr>
<tr>
<td>Lower and Upper Tail</td>
<td>$K_4 = \frac{1}{n} \sum_{t_i \in [0+\epsilon, 1-\epsilon]} \frac{(K(\theta_n, t_i) - K_\theta(t_i))^2}{(1-K_\theta(t_i))(1-K_\theta(t_i))}$</td>
</tr>
</tbody>
</table>

Table 1: Numerically approximated test statistics

for any hypothetical copula because they are supposed to play a role only when performing the Rosenblatt’s transform. So for independent coordinates the $K$ can be written as

$$K_{\text{indep}}(t) = P(I_d(E) \leq t) = P\left(\prod_{i=1}^{d} E_i \leq t\right) = t + t \sum_{i=1}^{d-1} \frac{(-1)^i}{i} \log(t)^i.$$ (11)

3.5 Testing Algorithm

The algorithm for the above described test procedures have been performed by repeating the following steps.

1. Simulate a sample from the copula model $C_\theta$ under the null-hypothesis
2. Re-estimate $\hat{\theta}$ from the simulation by maximum likelihood method
3. Transform the sample when needed by $R$ or $B$ based on the parameter $\hat{\theta}$
4. Approximate the required test statistics.

Finally compute the test statistics also for the observations and compare with the simulated ones by $p$-values.

4 Application to Wind Speed Maxima

We applied the described GoF procedures when fitting the copula models of Section 2 for a 3-dimensional wind speed dataset. The observations have been measured hourly for the recent 50 years at some locations in North-Germany (from 1958 till 2007). Our interest was in joint modeling of the monthly maxima of
wind speed in three cities: Bremerhaven, Hamburg and Hanover. These observations are shown on Figure 3 for each pair of the cities and altogether. The points in the upper right corners represent months when there was extremely high winds measured at least at 2 cities, indeed many of them occurred jointly at all of the 3 cities. These kind of events could cause rather dangerous consequences (e.g. from insurance point of view) so we were more focused on whether the fitted model was appropriate in these upper regions. As it was mentioned in Section 3, for fitting the different models the pseudo-observations have been used. Therefore in the first step of the analysis we transformed the data into the unit cube with the help of the empirical univariate margins. The dependence structure among the pseudos is shown on Figure 4 and the parameters of the fitted models are given in Table 2. In the scatter plots of the pseudo-observations there is no flagrant difference among the bivariate margins, so the possible use of an Archimedean copula, in which case all the bivariate margins are identical (i.e. exchangeable) $C_{\phi,\theta}(1, u_2, u_3) = C_{\phi,\theta}(u_1, 1, u_3) = C_{\phi,\theta}(u_1, u_2, 1)$, looks at least visually acceptable. From this point of view the elliptical ones are considered to be better espe-
4. Application to Wind Speed Maxima

Figure 4: Pseudo-observations after marginal transformations by the empirical distribution functions

especially for higher dimensions, because there is a parameter for every pair of margins to capture the dependence. Another crucial issue is the fit of the tails, as the each of the 4 presented copulas model it differently. The proposed GoF procedures have to be able to reflect on all the above problems or preferably on the mixture of them.

In the rest of the paper we concentrate on the GoF test described in Section 3. There have been three basically different approaches applied, which are presented in the next three subsections. The first approach is a test for checking the global fit (abbreviated as G-test, see 3.2 in Section 3) by the \{copula\} package of R. The second one is the Berg and Bakken test (abbreviated as B-test, see 3.3 in Section 3), which is based on checking independence hypothesis after the Rosenblatt's transform. (Applied only for Archimedean families.) Two different versions of B-test have been applied with weights, focusing on the fit near the upper tail. Finally, the third presented approach (abbreviated as K-test, see 3.4) is also a weighted one, based on the Kendall's transform.
4.1 G-Test

The GoF analyses have been started by applying the one of the most recent methods `gofCopula` from the `{copula}` package of R to have some preliminary understanding of the global fit of our models. For the wind maxima the default setting, using parametric bootstrap in the G-test turned out to be not very practical, as it has been very time-consuming even in this 3 dimensional case and as a result all of the proposed 4 copula models have been rejected. However another version of it using the multiplier approach instead of parametric bootstrapping has shown substantial differences among the proposed models. The results are shown in Table 2, from where we can conclude that the elliptical copulas (not rejected) look to be closer to the structure of the wind data than the Archimedean ones (both rejected).

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter(s)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel</td>
<td>$\theta = 1.858$</td>
<td>0.0000079</td>
</tr>
<tr>
<td>Clayton</td>
<td>$\theta = 1.312$</td>
<td>0.000049</td>
</tr>
</tbody>
</table>
| Gaussian| $\rho_1 = 0.72\ 
\rho_2 = 0.66\ 
\rho_3 = 0.66$ | 0.21        |
| Student | $\rho_1 = 0.74\ 
\rho_2 = 0.68\ 
\rho_3 = 0.69$ | 0.08        |

Table 2: G-Tests for the 3D wind speed data by `gofCopula` using multiplier method

As discrepancies could arise just because of assuming incorrect pair-wise association between the coordinates, it is recommended to dissect the problem. For better understanding, the model fit for the margins have been investigated as well. In this step we fitted the same type of models separately for the 3 possible pairs of cities and checked how they could perform in capturing the bivariate wind speed structure. The association have been measured by Kendall's correlation $\tau$ (instead of the usual Pearson's correlation $\rho$) because it is invariant under the PIT, ensuring that the original observations and the pseudo-observations have the same correlation. The estimated correlations are presented in the upper block of Table 3 which are very close to each other. Although this fact would really permit to use exchangeable copulas like the Archimedean ones, the fit is not adequate for any of them. By checking the $p$-values in Table 3 we can see that both of the Gumbel and Clayton family are rejected at $\alpha = 0.05$ level in all cases. Specifically the Clayton
model is always highly rejected and the Gumbel is slightly better with not being rejected at $\alpha = 0.01$ level 2 times out of 3. Between the elliptical copulas the overall fit looks better for the Student copula with 1 rejection out of 3 at $\alpha = 0.05$ in face with the Gaussian which is rejected 2 times. The most problematic dependence structure is between Hanover and Bremerhaven in which case none of the models could be really "accepted". As the picture is pretty mixed, based on this,

<table>
<thead>
<tr>
<th>Model</th>
<th>Param.</th>
<th>p-value</th>
<th>Param.</th>
<th>p-value</th>
<th>Param.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel</td>
<td>2.04</td>
<td>0.048</td>
<td>1.87</td>
<td>0.006</td>
<td>1.87</td>
<td>0.03</td>
</tr>
<tr>
<td>Clayton</td>
<td>1.62</td>
<td>0</td>
<td>1.32</td>
<td>0</td>
<td>1.34</td>
<td>0</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.74</td>
<td>0.035</td>
<td>0.68</td>
<td>0.023</td>
<td>0.68</td>
<td>0.15</td>
</tr>
<tr>
<td>Student</td>
<td>0.72</td>
<td>0.142</td>
<td>0.67</td>
<td>0.02</td>
<td>0.67</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 3: G-Tests for the 2D wind speed data by gofCopula using multiplier method

we can not state that any of the models is clearly better than the others. E.g. even if the Gumbel copula turned out to be weak in the overall fit could be fairly good in modeling the tail behavior. Therefore from now on we will be more focused on the tail behavior and apply dimension reduction approaches, to make the whole procedure easier to compute.

### 4.2 B-Test

As it was described before, the approach proposed by Berg and Bakken is an alternative way of testing the fit using the Rosenblatt’s transform. Here we investigate the Clayton and Gumbel copula, which turned out to underachieve according to the G-test. The formulas of the Rosenblatt’s transform for these copulas are easy to obtain by straightforward computation. In our 3 dimensional case the new coordinates can be computed as below.
1. Rosenblatt’s transform for Gumbel copula

\[
\begin{align*}
\epsilon_1 &= u_1 \\
\epsilon_2 &= -\frac{(-\ln(u_1))^{-(\theta+1)}}{u_1} \times Q_1^{\theta-1} \times C_\phi(u_1, u_2) \\
\epsilon_3 &= \frac{-1 + \theta + Q_2^{\frac{1}{\theta}}}{-1 + \theta + Q_1^{\frac{1}{\theta}}} \times \left(\frac{Q_2}{Q_1}\right)^{\theta-2} \times \frac{C_\phi(u_1, u_2, u_3)}{C_\phi(u_1, u_2)} \\
Q_i &= \sum_{j=1}^{i+1} \phi(u_j) = \sum_{j=1}^{i+1} (-\ln(u_j))^{-\theta}, \quad i = 1, 2.
\end{align*}
\]

2. Rosenblatt’s transform for Clayton copula

\[
\begin{align*}
\epsilon_1 &= u_1 \\
\epsilon_2 &= \frac{u_1^{-(\theta+1)}}{Q_1} \times C_\phi(u_1, u_2) \\
\epsilon_3 &= \left(\frac{Q_1}{Q_2}\right)^2 \times \frac{C_\phi(u_1, u_2, u_3)}{C_\phi(u_1, u_2)} \\
Q_i &= \sum_{j=1}^{i+1} \phi(u_j) = \sum_{j=1}^{i+1} u_j^{-\theta} - i, \quad i = 1, 2.
\end{align*}
\]

The Rosenblatt coordinates before the suitable transformation according to (8) are shown marginally in Figure 5. Investigating these figures in detail we can find some systematic deviance from uniformity in both of the cases. To decide quantitatively whether these deviances are significant or not, two different version of the B-test have been used. The differences between the procedures lie in the weightings:

- \( B_1 \)-test: There are no weights added on the points before the dimension reduction step \( \gamma(E_i, \alpha) = 1 \), but for the univariate test we used Anderson-Darling statistics with tail weights, more specifically (4) has been used.

- \( B_2 \)-test: The weights are focused on the joint upper tail \( \gamma(E_i, \alpha) = E_i^2 \) and we used general Cramér von-Mises statistics (viz. no additional weights have been used in the univariate test statistic).
4. Application to Wind Speed Maxima

Figure 5: Scatterplots of the 2D margins after the Rosenblatt’s transform assuming Gumbel (upper block) and Clayton (lower block) hypothesis

The results of the B-tests are summarized in Table 4, which gives the quantiles of the simulated statistics as well as the actual values. We found that the $B_1$-test is not so sensitive in detecting the deviations of the fitted models from the observations as neither the Gumbel ($p$-value=0.67) nor the Clayton copula ($p$-value=0.26) were rejected. However it is also clear, that the Gumbel copula seems to be much closer to the wind data. The reason behind the insensitivity could be due to the dimension reduction step (9) where no weights were added and so the information got possibly too strongly aggregated for the Anderson-Darling test. We conclude this since in the other case when we focused on the points around the upper tail in (9) the deviances were already more obviously detected. So generally the $B_2$-test seemed to be more reliable from this point of view. The $B_2$-test rejected the Clayton model ($p$-value=0.02) but did not reject the Gumbel ($p$-value=0.12).
We get the same conclusions if we check the independence of the transformed pseudo-observations by \( K \)-test according to (11). The Gumbel model was never rejected, in contrast to the Clayton model which was strongly rejected by all of the 4 versions of the proposed test statistics. However the different statistics did not make any significant difference from the inference point of view. We consider the upper tail weighted version as the sensible approach for testing tail dependence of extremes. So according to the \( K_2 \) statistics the \( p \)-value is 0.17 for the Gumbel model and it is 0.001 for the Clayton. The detailed results are summarized in Table 5. Despite of the known weakness due to the inconsistency of the approach we think it provides useful information about the model’s goodness. It is a rather serious problem with this approach that the Rosenblatt’s transform is not easily obtainable in higher dimensions for any models.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Gumbel copula</th>
<th>Clayton copula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( B_1 )</td>
<td>( B_2 )</td>
</tr>
<tr>
<td>95%</td>
<td>1.356</td>
<td>0.445</td>
</tr>
<tr>
<td>99%</td>
<td>1.853</td>
<td>0.598</td>
</tr>
<tr>
<td>Max.</td>
<td>2.986</td>
<td>0.893</td>
</tr>
<tr>
<td>Obs.</td>
<td>0.273</td>
<td>0.314</td>
</tr>
<tr>
<td>( p )-value</td>
<td>0.668</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Table 4: B-Test: Summary of the simulated test statistics under the hypothesis of independence, the observed test statistics and \( p \)-values for the wind data set
4. Application to Wind Speed Maxima

<table>
<thead>
<tr>
<th>Summary</th>
<th>Gumbel copula</th>
<th>Clayton copula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
</tr>
<tr>
<td>95%</td>
<td>0.0004</td>
<td>0.003</td>
</tr>
<tr>
<td>99%</td>
<td>0.0005</td>
<td>0.006</td>
</tr>
<tr>
<td>Max.</td>
<td>0.0008</td>
<td>0.019</td>
</tr>
<tr>
<td>Obs.</td>
<td>0.0001</td>
<td>0.002</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.317</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Table 5: K-Test: Summary of the simulated test statistics under the hypothesis of independence, the observed test statistics and $p$-values for the wind data set.

Figure 6: Fitted parametric K-functions
4.3 K-Test

In the final step of the study we investigated the GoF by different versions of K-tests. The approach is more universally applicable even though that the formula of the Kendall transform is not explicitly available for the elliptical copulas. This problem could be solved by simulation from the given model. For the elliptical copulas we approximated the $K(\theta, t)$ in (5) by taking the average of many $K_n$ values, simulated from the given model with fixed $\theta$ parameter. The fitted $K(\theta_n, t)$ functions are displayed in Figure 6 together with $K_n(t)$ arisen from the observations. Globally the elliptical models look to be closer to the observations as we have already concluded by performing G-tests. But we currently intend to concentrate to the upper tail so we need some finer assessment here. In the figures there is no evident tail-wise difference among the models. Clayton copula, it has lower tail dependence so the data have been flipped before the model fit.

After performing the tests we found that the Clayton copula is highly rejected by every statistics. In fact the observed test statistics were higher than any of the simulated ones. The Gumbel copula is also rejected, however if we just focus only on the upper tail by $K_2$ the fit looks to be slightly better ($p$-value=0.002). The elliptical copulas perform significantly better. The global fit is never rejected, but according to the $K_2$ statistics while the Gaussian copula is rejected with $p$-value=0.037, the Student with d.f. 12 can not be rejected ($p$-value=0.186).
## Table 6: K-Test: Summary of the simulated test statistics under different copula hypotheses, the observed test statistics and p-values for the wind data set

<table>
<thead>
<tr>
<th>Summary</th>
<th>Gumbel copula</th>
<th>Clayton copula</th>
<th>Gaussian copula</th>
<th>Student copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>$K_1$</td>
<td>$K_2$</td>
<td>$K_3$</td>
<td>$K_4$</td>
</tr>
<tr>
<td>95%</td>
<td>0.00030</td>
<td>0.00130</td>
<td>0.00070</td>
<td>0.00190</td>
</tr>
<tr>
<td>99%</td>
<td>0.00040</td>
<td>0.00170</td>
<td>0.00100</td>
<td>0.00250</td>
</tr>
<tr>
<td>Max.</td>
<td>0.00060</td>
<td>0.00310</td>
<td>0.00120</td>
<td>0.00400</td>
</tr>
<tr>
<td>Obs.</td>
<td>0.00110</td>
<td>0.00240</td>
<td>0.00310</td>
<td>0.00550</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00020</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>Summary</td>
<td>Gaussian copula</td>
<td>Student copula</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.00030</td>
<td>0.00140</td>
<td>0.00070</td>
<td>0.00200</td>
</tr>
<tr>
<td>99%</td>
<td>0.00040</td>
<td>0.00170</td>
<td>0.00100</td>
<td>0.00250</td>
</tr>
<tr>
<td>Max.</td>
<td>0.00070</td>
<td>0.00270</td>
<td>0.00150</td>
<td>0.00400</td>
</tr>
<tr>
<td>Obs.</td>
<td>0.00010</td>
<td>0.00150</td>
<td>0.00010</td>
<td>0.00160</td>
</tr>
<tr>
<td>p-value</td>
<td>0.78400</td>
<td>0.03700</td>
<td>0.95800</td>
<td>0.83100</td>
</tr>
</tbody>
</table>
5 Conclusions

Table 7: Simulated and observed statistics for the wind data set

<table>
<thead>
<tr>
<th>Model</th>
<th>$G$-Test</th>
<th>$B_1$-Test</th>
<th>$B_2$-Test</th>
<th>$K$-Test ($II_2$)</th>
<th>$K$-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Clayton</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Gaussian</td>
<td>✓</td>
<td>n.a</td>
<td>n.a</td>
<td>n.a</td>
<td>x</td>
</tr>
<tr>
<td>Student</td>
<td>✓</td>
<td>n.a</td>
<td>n.a</td>
<td>n.a</td>
<td>✓</td>
</tr>
</tbody>
</table>

We can summarize our findings as follows.

- As there is no clearly best test in our setup, one has to use several GoF tests in order to get a picture of the behavior of the fitted models, including the faster version of the $G$-test, possibly with weights. In our case the Student copula turned out to be the best model among those investigated.

- The presented analysis about the Rosenblatt’s transform-based methods was rather preliminary, but our results have not proven their superiority. Its combination with the $K$-test turned out as especially problematic, as here even the weights were not having their usual meaning.

- A lot of work need to be done until one gets a clear picture about the properties of these statistics for real, 3-dimensional and relatively small data sets. The order of the conditioning as well as the choice of statistics and the dimension reduction method for a given type of problems is not yet solved. One may need to check alternative models as unexchangeable copulas and pair-wise copula construction methods, which all provide new GoF challenges.

Acknowledgments

The work of the first named author was mainly supported by a grant from the Marie Curie RTN Seamocs project "Applied stochastic models for ocean engineering, climate and safe transportation". We wish to express our particular thanks to Nader Tajvidi for his helpful suggestions and comments.
References


REFERENCES


C
Paper C

Autocopulas: investigating the interdependence structure of stationary time series

Pál Rakonczai, László Márkus, András Zempléni

Department of Mathematical Statistics, Lund Institute of Technology
Box 118 SE-22100, Lund, Sweden
and
Department of Probability Theory and Statistics, Eötvös Loránd University
Budapest, Hungary

Abstract

Here we present a novel approach to the description of the lagged interdependence structure of stationary time series. The idea is to extend the use of copulas to the lagged (one-dimensional) series, to the analogy of the autocorrelation function. The use of such autocopulas can reveal the specifics of the lagged interdependence in a much finer way. However, the lagged interdependence is resulted from the dynamics, governing the series, therefore the known and popular copula models have little to do with that type of interdependence. True though, it seems rather cumbersome to calculate the exact form of the autocopula even for the simplest nonlinear time series models, so we confine ourselves here to an empirical and simulation based approach. The advantage of using autocopulas lays in the fact that they represent nonlinear dependencies as well, and make it possible e.g. to study the interdependence of high (or low) values of the series separately. The presented methods are capable to check whether autocopulas of an observed process can be distinguished significantly from the autocopulas of a given time series.
model. The proposed approach is based on the Kendall's transform which reduces the multivariate problem to one dimension. After illustrating the use of our approach in detecting conditional heteroscedasticity in the AR-ARCH vs. AR case, we apply the proposed methods to investigate the lagged interdependence of river flow time series with particular focus on model choice based on the synchronized appearance of high values.

**Key words:** copulas, stationary time series, autocopulas, goodness-of-fit test, Kendall's transform, river flow series

### 1 Introduction

Our goal in the present paper is to go beyond linearity and hence autocorrelation or its transforms in the description of the lagged interdependence structure of a stationary time series. Driven by that aim, to the analogy of autocorrelations we define for an arbitrary lag or set of lags the so-called autocopula, which is the copula of the original and the lagged series. Autocopulas are ordinary copulas related to a time series, and as such they describe the interdependence structure in more detail than autocorrelations do, specifically, they take into account non-linear interdependencies as well. Lagged interdependence is closely related to the dynamical model of the series so the importance to get more detailed and accurate information about it is obvious. Hence, model adequacy can be evaluated more sensitively by testing the fit of autocopulas.

Though copula theory was not particularly popular at the time of its creation (Hoeffding 1940), and early stage (Sklar 1959), it has been re-invented in the 1990’s and underwent a rapid expansion, creating sometimes heated debates (Mikosch 2006; Genest and Rémillard 2006), ever since then. We share the common view, that the main advantage of using copulas for describing the interdependence structure is that it incorporates all types of dependence, - not only linear ones like Pearson’s correlation - and it handles the margins and the interdependence separately. Though note here immediately, that e.g. the same linear combinations \( U = aX + bY, \ V = cX + dY \) of independent variables \( (X, Y) \) will result in different copulas \( C_{U, V} \) for different margins of \( (X, Y) \). This fact has the important consequence that the autocopulas (specified in the next section) of linear processes are dependent not only on the dynamics of the series, represented by the coefficients, but also on the marginal distribution of the generating noise.
1. Introduction

There are commonly used parametric copula families, but again, aside of the simplest (e.g. the Gaussian) cases, these are non-conformable with parametric time series dynamics and vice versa. On the other hand (Ghoudi and Rémillard 2004; Genest and Rémillard 2004) introduce serial copulas of strictly stationary time series, what is the copula of say $m$ consecutive values of the series. Serial copulas are time-shift invariant by virtue of the strict stationarity, and they are associated with the $m$-dimensional joint distribution of the series. In fact the same object is introduced in the ARCH case (Diks and Panchenko 2008), and called the ARCH copula. The mentioned authors use serial copulas for testing serial independence of observations, Diks and Panchenko being specific against the ARCH alternative. The review article of copula theory for finance (Patton 2008) also singles out the study of the copula of a sequence of consecutive observations i.e. the same object as the previous authors, as a branch of current time series modeling and mentions a few further authors whose work can be related to the subject. The other branch mentioned is the study of cross series dependence in multidimensional models, that is not related to our work here. Going the other way, Markov time series are built up in (Chen and Fan 2006) by prescribing parametric copula families to the lag one interdependence. Nonlinear AR(1) series with appealing properties can be created this way but it seems to be quite complicated to interpret the corresponding dynamics, especially, when this method is generalized beyond AR(1).

Being associated with the full $m$ dimensional joint distribution, up to lag $m$ the serial copula incorporates, of course, all information on the dependence of the lagged series, too. However, the growing lag necessitates the increase of the dimension of the joint distribution, and with that grows the dimension of the serial copula, making the latter a rather inconvenient, and unnecessarily complicated tool for characterizing lagged interdependences. It is true, that neither of the mentioned publications use it for such purpose. We come up here with an other approach to represent lagged interdependence, that is two dimensional only, hence simpler but still effective for the goal, and also easier to visualize: the concept of autocopula, as defined in Section 2.

In a simulation experiment we demonstrate the use of autocopulas by distinguishing between observations of AR-ARCH and simple AR series. The importance of the chosen model is demonstrated by the fact that AR-ARCH or slightly more generally AR-GARCH series with possibly skewed and fat tailed innovations are used in mathematical finance to describe e.g. the price evolution of daily
returns on individual stock indices, as in (van den Goorbergh 2004). Our procedure does have a practical value, too, as AR-ARCH series are easy to misspecify as a higher order AR series, but autocopulas detect the difference pretty reliably. About AR-ARCH series it is proven in (Borkovec and Klüppelberg 2001) that the AR(1) model driven by a quadratic ARCH(1) innovation has stationary distribution with regularly varying tail for a very general class of noise distributions. Condition on geometric ergodicity - at the same instance, stationarity - of non-linear AR(p) models with GARCH(1,1) innovation were established in (Meitz and Saikkonen 2008). The estimation properties (consistency and asymptotic normality of the AR parameter estimation) of AR-GARCH series were also studied (Francq and Zakoïan 1998 and Francq et al. 2005). AR-ARCH parameter estimation is readily available e.g. in the “tseries” package of R.

Elek and Márkus (Elek and Márkus 2008) model the daily river discharge data of Danube and Tisza Rivers in Hungary with the ultimate aim of inundation risk evaluation, and study the extremal behavior of the model processes. They model the dynamics of the entire flow series and conclude on the extremes from the extremal characteristics of those models. Their suggested dynamics differs from conventional ARMA-GARCH models as the variance of innovations of the linear filter is conditioned on the lagged values of the generated process instead of the innovations themselves and the variance is asymptotically a linear (instead of a quadratic) function of the past values. We compare the model proposed in (Elek and Márkus 2008) to a FARIMA alternative when both are fitted to a real hydrological data set. Our point of view in evaluating the models is the interdependence structure description, represented by the autocopulas.

As the applications demonstrate we strongly believe that copula theory may contribute significantly to the analysis of time series by revealing the character of the nonlinear interdependence, when inherent in the series. So, in our view copulas, and in that respect autocopulas, do not model time series but rather they are important characteristics of various models, just as autocovariance functions are, and we intend to use them in that spirit, despite of the lack of a well elaborated theory for autocopulas as yet.

The paper is organized as follows. We first briefly recall in Section 2 the needed elements of copula theory and present the notations. Here we define the so-called autocopulas for stationary time series. In Section 3 we present an effective approach for measuring the goodness-of-fit (GoF) for copula models. Then we suggest some practical modifications of the test-statistics as our focus is
on the joint behavior of extreme events, and we intend to reach more sensitivity in that respect. The reason is that differentiating between models e.g. in upper tail dependence necessitates a very exact exploration of the copula around the upper right corner of the unit square. Our proposed GoF test is based on the Kendall’s transform of the joint distribution, which from the one hand reduces the multivariate problem to one dimension and from the other hand provides possibility to use specific weights as well, focusing on the desired region of the distribution. Section 4 contains the results of the above mentioned simulations, and distinctions of AR and AR-ARCH series by their autocopulas. Section 5 presents the mentioned comparison of hydrological models. For its basis, we address the GoF of the autocopulas of the fitted models to the ones of the observed discharge series by the tests suggested in Section 3.

2 From Copulas to Autocopulas

In what follows, we extend the use of copulas to the interdependence structure of stationary time series, to the analogy of the autocorrelation function. Consider a bivariate continuous random vector \((X, Y)\) with joint distribution function \(H\) and margins \(F\) and \(G\) respectively. Due to Sklar’s theorem (Sklar 1959) the joint distribution function \(H\) can be represented as \(H(x, y) = C(F(x), G(y))\), where \(C(x, y)\) the so called copula of \((X, Y)\) is a bivariate distribution function with uniform margins. Moreover, the copula \(C\) is unique if the marginal distributions are continuous. For more details see the books (Joe 1997 and Nelsen 2006) where various copula families have been introduced. Further illustrations of copula methods can be found in (Cherubini et al. 2004 and McNeil et al. 2005).

We now give a general definition of the autocopula and also a simplified (bivariate) version of it, which is largely sufficient for our practical purposes.

Definition 2.1. Given a strictly stationary time series \(Y_t\) and \(L = \{l_i \in \mathbb{Z}^+, i = 1, ..., d\}\) a set of lags, the autocopula \(C_{Y, L}\) is defined as the copula of the \(d + 1\)-dimensional random vector \((Y_t, Y_{t-l_1}, ..., Y_{t-l_d})\).

Let us remark that the supposed strict stationarity implies that the autocopula does not, indeed, depend on \(t\).

Definition 2.2. Given a strictly stationary time series \(Y_t\) and \(l \in \mathbb{Z}^+\) the \(l\)-lag autocopula \(C_{Y,l}\) is the copula of the bivariate random vector \((Y_t, Y_{t-l})\). The \(l\)-lag autocopulas as the function of the lag \(l\) give the autocopula function.

One of the most commonly used parametric copula family is the so-called
Gauss copula family. This family can simply be derived from a Gaussian distribution function $\Phi_\Sigma$ with mean zero and correlation matrix $\Sigma$ by transforming the margins by the inverse of the standard normal distribution function $\Phi$. In the bivariate case it has the form

$$C(x, y) = \Phi_\Sigma(\Phi^{-1}(x), \Phi^{-1}(y)).$$

It is obvious that the dependence in a Gauss copula is completely determined by the covariances. In the time series setup, the Gaussian copula plays a very specific role, as linear processes with Gaussian innovations give rise to Gaussian copulas as both their serial copulas and autocopulas. We emphasize here the dependence of the autocopula on the marginal distribution of the noise by mentioning (cf. introduction) that linear processes with non-Gaussian innovations do not, in general, have Gaussian autocopulas. Gaussian linear time series with autocorrelation decaying quicker than $1/\log(t)$ do have Gaussian-like interdependence, e.g. their "very" extremes cannot appear to be synchronized. This property can be quantified by the tail-dependence $\chi$, which is 0 for the Gaussian copula. For a detailed description of dependence measures see (9.5 in Berlaint 2004). While covariances are fully representative of interdependence in the Gaussian case, they are insufficient for non-Gaussian distributions, especially fail to represent the extremal dependence.

3 Goodness-of-Fit Tests for Autocopulas

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

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

(1)
e.g., the dependence structure of the investigated autocopula arises from $\mathcal{C}_{0,l}$, which is a copula family defined by a specific time series model. One very crucial question before starting off is what exactly should be considered as an appropriate sample for the later inference. As the usual inference theory for copulas works with the assumption of i.i.d. observations from copula models one should not

\(^1\)Or if yes, somehow the dependence should be taken into account.
use all single pairs of observations from the autocopula

\[ \left\{ \left( Y_i, Y_{i-l} \right) : i \in \{ l + 1, \ldots, n \} \right\}, \]

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

\[ \left\{ \left( Y_i, Y_{i-l} \right) : i \in T \right\} \]

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

\[ \left\{ \left( \frac{\sum_{j \in T} 1(Y_j \leq Y_i)}{|T| + 1}, \frac{\sum_{j \in T} 1(Y_{j-l} \leq Y_{i-l})}{|T| + 1} \right) : i \in T \right\}. \]

Therefore the \( \{(U_i, V_i) : i \in T\} \) can be interpreted as an i.i.d. sample from the underlying autocopula \( C_{Y_{l}, l} \). Before discussing the GoF tests for copulas we give a short overview of the univariate case, which later will be utilized for the bivariate case after having used a feasible dimension reduction method.

### 3.1 Univariate GoF Tests

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

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

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

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

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

\[ T_{AD} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x), \tag{3} \]

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

\[ T_{uAD} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{1 - F(x)} dF(x) \tag{4} \]

for the case of maximum and with \( \Phi(x) = F(x) \) in the place of \( \Phi(x) = 1 - F(x) \) for minimum. The advantage of (4) in comparison to (3) is that its sensitivity is concentrated to discrepancies at the relevant tail of the distribution see Zempléni’s test (Kotz and Nadarajah 2000, p.77). The computation of these statistics is straightforward and for arbitrary weight function \( \Phi(x) \) can be numerically approximated.

### 3.2 Copula GoF Tests

As it has been already mentioned before, one might use the univariate techniques described above, after performing an appropriate dimension reduction procedure on the copula distribution. For this purpose we suggest the use of the so-called Kendall’s transform as follows

\[ \mathcal{K}(t) = P(H(X, Y) \leq t) = P(C(F(X), G(Y))) \leq t = P(C(U, V)) \leq t. \tag{5} \]
The empirical version of $K$ can be computed as

$$K_n(t) = \frac{1}{n} \sum_{j=1}^{n} 1(E_{in} \leq t), \quad t \in [0, 1],$$

where

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

Although a closed formula for (5) is only available for some specific copula families, $K(t)$ can be easily approximated by simulation from any given model with the desired accuracy. For illustrations see Figure 1. In the later sections we refer to the approximated version of $K(t)$ simply as K-function.

![K(t) for Gauss copula](image)

Figure 1: Simulated $K(t)$ for Gauss copulas with different Pearson’s correlation

Known tests for checking the match of the “theoretical” $K(t)$ and its empirical version use continuous functionals of Kendall’s process

$$X_n(t) = \sqrt{n}(K(\theta_n, t) - K_n(t)),$$

having favorable asymptotic properties. There are two different kind of approaches
investigated in (Genest et al. 2006), Cramér-von Mises type and Kolmogorov-Smirnov type as
\[ S_n = \int_0^1 (x_n(t))^2 \, dt \]
\[ T_n = \sup_{0 \leq t \leq 1} |x_n(t)|. \]
statistics. Since the second statistic \( T_n \) is proved to be generally less powerful in detecting discrepancies near the tails, we suggest to base the inference on the test statistics according to \( S_n \), analogously to (2), (3) and (4). The proposed test statistics are summarized in Table 1, \( (t_i)_{i=1}^m \) is an appropriately fine division of the interval (0, 1)).

<table>
<thead>
<tr>
<th>Focused Regions</th>
<th>Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>( S_1 = \frac{1}{m} \sum_{t \in [0,1]} (\mathcal{K}(\theta_n, t) - \mathcal{K}_n(t))^2 )</td>
</tr>
<tr>
<td>Upper Tail</td>
<td>( S_2 = \frac{1}{m} \sum_{t \in [0,1]} \left( \frac{\mathcal{K}(\theta_n, t) - \mathcal{K}_n(t)}{1 - \mathcal{K}(\theta_n, t)} \right)^2 )</td>
</tr>
<tr>
<td>Lower Tail</td>
<td>( S_3 = \frac{1}{m} \sum_{t \in [0,1]} \left( \frac{\mathcal{K}(\theta_n, t) - \mathcal{K}_n(t)}{\mathcal{K}(\theta_n, t)} \right)^2 )</td>
</tr>
<tr>
<td>Lower and Upper Tail</td>
<td>( S_4 = \frac{1}{m} \sum_{t \in [0,1]} \left( \frac{\mathcal{K}(\theta_n, t) - \mathcal{K}_n(t)}{1 - \mathcal{K}(\theta_n, t)} \right)^2 )</td>
</tr>
</tbody>
</table>

Table 1: Numerically approximated Cramér-von Mises type test statistics on the Kendall’s process

### 4 Example: Testing for Heteroscedasticity in the Innovation of an AR Model

Here we illustrate the use of the suggested test procedure through a simple but important practical example. Time series having the same weak AR representation, i.e. complying the same AR model driven by uncorrelated innovations do have identical autocovariance structure: the one of the "classical" AR series, generated from that particular AR model by i.i.d. innovations. (When referring simply to AR or AR(p) series in the sequel we always mean innovations to be i.i.d.) As a consequence, no test based on autocovariances can really make a distinction between an ARCH- (or eventually GARCH) innovation driven AR (hereinafter AR-ARCH or AR-GARCH) and an i.i.d. one driven AR series. The identification of
4. Example: Testing for Heteroscedasticity in the Innovation of an AR Model

The autocopula may serve this end, and this is what we intend to show here. So, we use the concept of autocopula for model selection purpose just as the ACF/ACVF is most frequently used. One may argue that the rejection of the AR hypothesis can be achieved e.g. by linearity tests, too, but then the possible alternative is not restricted to one particular type of models, while a well-identified autocopula may point to that within the bounds of reliability inherent in the sample.

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

\[
Y(t) = \phi \cdot Y(t-1) + \epsilon(t)
\]

\[
\epsilon(t) = \sigma(t) \cdot Z(t)
\]

\[
\sigma^2(t) = \omega + \alpha \cdot \epsilon^2(t-1),
\]

where \( |\phi| < 1, \omega > 0, \alpha > 0, \alpha + \omega < 1 \) and \( Z(t) \) is an i.i.d. standard normal series. These conditions imposed on the parameters guarantee the existence of a stationary solution with finite variance (though the study would have been possible under more relaxed assumptions, including infinite variance, as well). As for the terminology we call \( \epsilon(t) \) to be the innovation and \( Z(t) \) the generating noise. We test, and try to reject the null-hypothesis that \( Y(t) \) is an AR(1) series with i.i.d. innovations, having the same marginal distribution as the stationary distribution of the ARCH innovation \( \epsilon(t) \).

Since the simulations serve only the illustration of the proposed method, we are not aimed at a full scale investigation of the behavior of various AR-ARCH samples. So, we only consider one fixed setup of the parameters, in particular, we choose the autoregressive coefficient in the equation as \( \phi = 0.5 \) and the two parameters for the heteroscedastic innovation as \( \omega = 0.1 \) and \( \alpha = 0.85 \). This choice results in a definite but not too strong autocorrelation, a significant ARCH effect, and a moderate variance. We simulate relatively large time series samples with size \( N = 50000 \) from the AR-ARCH series \( Y(t) \). In doing this, we first generate the stationary ARCH innovations from standard Gaussian white noise, letting a sufficiently long (50000) step burn in period before we store the actual 50000 innovation values. We then create the AR-ARCH sample from these innovations by R’s "arimasim" function. We also generate AR series, by first taking a random reordering of the ARCH sample, to destroy interdependence. This way we obtain
an innovation, that various tests accept for an i.i.d. sample. Among them is the BDS-test (Brock-Dechert-Scheinkman) testing serial independence. For all the simulated ARCH samples the BDS-test is highly significant, i.e. rejects the null of serial independence by practically all-zero p-values, and the test is insignificant for the resampled innovations, giving typical p-values in the range of 0.4 and 0.6. This is not surprising, as the sample size is pretty high. The same can be inferred from the autocopula of the resampled series. As we noted above the autocopula of a linear process is dependent on the marginal of the innovation, so, it is of utmost importance that the newly created i.i.d. innovations have the same marginal distribution as the stationary ARCH series, and from this resampled innovation we create the required AR series. Re-generating the ARCH sample, and repeating the latter procedure to obtain independent innovations, we get 500 of such AR series and compute from each of them the \( l \)-lag autocopulas, and their K-functions. Averaging the K-functions out we obtain a good estimation of the theoretical \( K \) in (5). This kind of empirical approach is necessary since it seems hopeless to compute an exact closed formula of the true theoretical \( K \). In order to explore the difference we compare the autocopulas of the AR and the AR-ARCH series at the same set of lags, basically by checking the distance of the empirical K-function of the AR-ARCH series from the averaged K-function of the AR series by the help of the test statistics presented in Table 1. For the formal test the 95% quantiles of the different test statistics have been computed as critical values. Again, as the distributions of the test statistics are not known the 95% quantiles under the null hypothesis were determined from the simulated AR samples. Practically these critical values represent the maximal "distance" which is still acceptable for a given simulation to be considered as an AR according to the null-hypothesis at \( \gamma = 0.05 \) significance level. The autocovariance functions of the AR and AR-ARCH series, displayed in the left panels of Figure 2, are very similar as expected, and their differences in the first few lags are insignificant, according to the limit distribution of the estimator.

In contrast to the almost full coincidence of the autocovariance functions, the autocopulas differ substantially. For instance the discrepancies between the 1-lag autocopulas can even be detected visually, as is shown in the right panels of Figure 2. We have taken a sample from both the AR-ARCH and the AR series created by the above described simulations, and thinned it for every considered lag by choosing every 10th \( l \)-lag-apart pairs of values (this means e.g. consecutive pairs in the case of \( l = 1 \)), creating a sample of \( (n = 500) \) pairs from every process.
4. Example: Testing for Heteroscedasticity in the Innovation of an AR Model

Figure 2: Differences between AR and AR-ARCH: The left block shows the estimated ACF-s, whereas the right block shows thinned 1-lag autocopula samples for both time series models, with thinning parameter $s = 10$ and sample size $n = 1500$.

(Note that in our notation $N$ refers to the sample size of the original time series and $n$ refers to the sample size of the thinned copula sample.) In view of the fading away interdependence within the time series the $l$-lag-apart pairs in the thinned sample for small lags hardly (insignificantly) differ from an independent sample of pairs of variables, with the same pairwise interdependence. This is the reason why we use those thinned samples to estimate the autocopula of the given lag. A hint for the choice of the thinning can be obtained from the autocovariance function which is practically zero for $l = 10$, so the elements of autocopula sample do not suffer from the interdependence effect of the time series. (The effect of the choice of thinning is discussed later at the end of this section.) The difference between
higher lag autocopulas is not so clear visually, see Figure 3, so there is a definite need for more quantitative investigation, based on formal tests.

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

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

1. Simulate AR time series 500 times as described above

2. Obtain their $l$-lag autocopula sample from the thinned by $s = 10$ series

3. Calculate the test statistics $S_{i,j}$, $i = 1, \ldots, 4$ and $j = 1, \ldots, 500$ and choose the 0.95 quantiles as critical values $Q_{i,0.95}$, $i = 1, \ldots, 4$

4. Simulate AR-ARCH time series 500 times

5. Obtain now their $l$-lag autocopula sample from the thinned by $s = 10$ series

6. Calculate the test statistics $S_{i,j}$, $i = 1, \ldots, 4$ and $j = 1, \ldots, 500$ and in every case reject $H_0$ when $S_{i,j} > Q_{i,0.95}$

The results are summarized in Table 2.

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

Table 2: Rejection percentages of the hypothesis of AR autocopula, if the true model is an AR-ARCH, at $\gamma = 0.05$ significance level and sample size $n = 500$ (out of 500 simulation)

Although all 4 tests gave very similar results for $l = 1$, we found that the tail-sensitive test statistics performed definitely better, especially for higher lags. Namely $S_2$ (with upper weights) and $S_4$ (both upper and lower weights) turned out to be the most effective. The efficiency was not too high at this sampling level, even so the tests were able to separate almost 70% of the AR-ARCH models from the AR ones by the 1-lag autocopula, and this separation level can be increased by taking into account more lags simultaneously. Of course, one could improve the separation rate by increasing the significance level but by doing so one would
reject more AR models incorrectly. In order to reach real and significant improvement in separation one should increase the sample size for the autocopulas, when possible, creating a finer resolution image better approximating the real copula. Our large simulation sample size enabled us to perform the same algorithm with larger autocopula samples. Still applying the same $s = 10$ thinning parameter the tests have been recomputed with new sample sizes as $n = 1000, 1500$. The rejection rates for $S_2$ and $S_4$ are shown in Table 3.

<table>
<thead>
<tr>
<th>lag</th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 1500$</th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 1500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l = 1$</td>
<td>65.7%</td>
<td>93.4%</td>
<td><strong>99.3%</strong></td>
<td>70.9%</td>
<td>96.0%</td>
<td><strong>99.6%</strong></td>
</tr>
<tr>
<td>$l = 2$</td>
<td>28.0%</td>
<td>49.7%</td>
<td><strong>68.6%</strong></td>
<td>27.8%</td>
<td>50.4%</td>
<td><strong>69.9%</strong></td>
</tr>
<tr>
<td>$l = 3$</td>
<td>23.7%</td>
<td>39.6%</td>
<td><strong>51.6%</strong></td>
<td>22.4%</td>
<td>37.7%</td>
<td><strong>50.3%</strong></td>
</tr>
<tr>
<td>$l = 4$</td>
<td>19.5%</td>
<td>28.2%</td>
<td><strong>34.8%</strong></td>
<td>17.8%</td>
<td>25.5%</td>
<td><strong>32.2%</strong></td>
</tr>
<tr>
<td>$l = 5$</td>
<td>12.0%</td>
<td>15.9%</td>
<td><strong>20.4%</strong></td>
<td>11.2%</td>
<td>14.1%</td>
<td><strong>18.4%</strong></td>
</tr>
<tr>
<td>$l = 6$</td>
<td>9.8%</td>
<td>11.4%</td>
<td><strong>11.5%</strong></td>
<td>9.3%</td>
<td>10.3%</td>
<td><strong>10.3%</strong></td>
</tr>
<tr>
<td>$l = 7$</td>
<td>7.4%</td>
<td>8.0%</td>
<td><strong>8.8%</strong></td>
<td>6.9%</td>
<td>7.8%</td>
<td><strong>7.8%</strong></td>
</tr>
</tbody>
</table>

Table 3: Rejection rates computed from samples with sample size $n = 500, 1000$ and $1500$ (out of 500 simulations)

We conclude that the sample size has a very significant effect on the results. Although the computations assuming larger autocopula samples are more exhaustive, the larger copula samples yield definitely stronger tests. For example the $S_4$ based test improves by 30% for $l = 1$, resulting in, that almost all of the ARCH time series has been separated from AR successfully. For higher order autocopulas ($l = 2, 3, 4, 5$) the rejection rates increase roughly 1.5 − 2 times, as well. For greater lags ($l \geq 5$) there is no such improvement any more, but the explanation behind this phenomenon is simply the negligible association in the autocopulas. As with the increasing lags the strengths of association fades out totally, there remains practically no measurable dependence to model and compare. As it has already been mentioned, a theoretical assumption must be fulfilled when using standard results about copulas, namely that the multivariate observations are supposed to be independent. As usual in time series models the assumption of independence between close pairs: $(X_t, X_{t-l})$ and $(X_{t+s}, X_{t+s-l})$, $s$ being small, is not realistic. (In this notation $t \geq 1$ and $l \geq 1$ denotes the time and the lag
as before, moreover $s \geq 1$ is what we have earlier referred to as the thinning parameter.) Of course the larger $s$ is the better, but in practical applications there is always a limit defined by the given sample size. So one needs to find the optimal $s$ which is large enough for acceptable independence and small enough for having sufficiently large sample for the proper inference. The decision can be based on the ACF by choosing such a large $s$ for which the ACF is essentially zero.

In the given situation there is no limitation on the sample size as we can generate time series with arbitrary length. In order to investigate the effect of the thinning parameter on the test performance $n = 500$ has been chosen constantly as sample size, avoiding differences appearing in the comparison just because of the different sample sizes. The results for the relevant lags are shown in Table 4. By analyzing it, we can find a remarkable relapse in the test performance when the criteria of independence of sample elements is presumably violated, i.e. the observation pairs have been chosen too close to each other, as in the case of $s = 1$, or to a lesser extent for $s = 3$.

<table>
<thead>
<tr>
<th>lag</th>
<th>$S_2$ statistics</th>
<th>$S_4$ statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l = 1$</td>
<td>$s = 10$</td>
<td>65.7%</td>
</tr>
<tr>
<td></td>
<td>$s = 5$</td>
<td>66.8%</td>
</tr>
<tr>
<td></td>
<td>$s = 3$</td>
<td>63.9%</td>
</tr>
<tr>
<td></td>
<td>$s = 1$</td>
<td>59.2%</td>
</tr>
</tbody>
</table>

Table 4: Rejection rates by decreasing $s = 10, 5, 3$ and 1 thinning parameter (out of 500 simulations, with sample size $n = 500$)

5 An Application in Hydrology

Turning now to the hydrological application we recall here a model of daily river discharge data for Danube and Tisza Rivers in Hungary, as described in (Elek and Márkus 2008). Our aim is to analyze the fit of that model, in terms of the lagged interdependence structure, as compared to a fitted FARIMA model. For this end we compare the autocopulas of the observed river flow data with simulated ones from the two models. The empirical autocopulas for the observed data are shown
in the first row of Figure 5. In order to be self-contained we briefly sketch the model given in the mentioned paper, where further analyzes are also presented. The deseasonalized river flow series \( Y_t = X_t - c_t \) (with seasonal component \( c_t \)) has skewed, leptokurtic and light-tailed marginal distribution. The autocorrelated squares and absolute values of the innovations of a fitted ARMA filter and its nonseasonal periods of high and low variances point to conditionally heteroscedastic (GARCH-type in wide sense) modeling. Elek and Márkus (2008) suggest the model:

\[
Y_t = \sum_{i=1}^{\rho} a_i (Y_{t-i}) + \sum_{i=1}^{q} b_i \varepsilon_{t-i}, \\
\varepsilon_t = \sigma (Y_{t-1}) Z_t, \\
\sigma^2(x) = (\alpha_0 + \alpha_1 (x - m)_+).
\]

with positive constants \( a_i, b_i, \alpha_0, \alpha_1, m \), innovation \( \varepsilon_t \) and i.i.d. noise \( Z_t \). The model differs from conventional ARMA-GARCH ones as the variance of innovations is conditioned on the lagged values of the generated process instead of the innovations themselves and is asymptotically linear function of its past values. Hereinafter, in the frames of the present paper we refer to this model as H-GARCH pointing by H to the hydrological application. The model is certainly heteroscedastic, in a way it is a generalization of the ARCH idea, and contains an ARMA filter, and the conditional variance of the innovation contains a feedback from the linearly filtered process. An alternative FB-ARMA-GARCH notation would be rather clumsy, however. We have the daily water discharge data of River Tisza, at Vásárosnamény gauge (North-east of Hungary) for the period of 100 years (1901-2000) available for the present analysis. We fit the specified model to the deseasonalized (in both mean and variance) data and simulate from it 100 samples of the same length following the methodology described in (Elek and Márkus 2008) in detail.

As alternative modeling we also fit a fractionally differenced ARIMA(2,d,1) model to the same deseasonalized data on the basis of the Whittle-estimator and also simulate 100 FARIMA samples of the same size. As an illustration, Figure 4 displays in its upper row a five years long section of the deseasonalized daily observations, together with simulations from the fitted models. In the lower row there are the ACF-s respectively.

We compute the autocopulas, average out their K-functions for both models, and compute the 95 % confidence bound to both averages, just like in the pre-
5. An Application in Hydrology

Figure 4: Upper block: Five years windows of deseasonalized (daily) water discharge time series for River Tisza and simulated artificial discharges by Hydro-GARCH and FARIMA models. Lower block: The estimation of ACF-s for the deseasonalized observations and the two models.

In the previous section, we compute the autocopulas and their K-functions for the observed sample as well. The autocopulas are shown in Figure 5 for the first 3 lags. In the next step we compare the distances of the K-function of the observed sample from the averaged K-functions of the two models according to the GoF measures of Section 3. Our proposed formal tests decide whether the autocopula of the observations differs significantly from the autocopulas of the models. The same settings as of Section 4 have been applied, namely we used $s = 10$ as the thinning parameter and run the tests for both models assuming different autocopula sample sizes $n = 500, 1000$. The results are summarized in Table 5, 6. There are two vertical blocks in each tables, separating the results concerning the two applied models, H-GARCH and FARIMA respectively. The consecutive columns contain the 95% critical values and the observed test statistics for the 4 different tests, based on the statistics presented in Table 1. From top to bottom we see the changes resulted in by the increasing lags. (The boldfaced numbers correspond to cases, when the test do not reject the null that the observed sample complies...
Figure 5: Comparison of autocopulas: Thinned samples from the observed and simulated autocopulas with thinning parameter \( s = 10 \), sample size \( n = 1500 \)

the model, i.e. the test statistics do not exceed the critical values.) We found that the less powerful tests (based on relatively low amount of copula data, \( n = 500 \)) do not reject at \( \gamma = 0.05 \) significance level for the first 3 lags in the H-GARCH, and for all considered lags, in the FARIMA case. However if we require higher accuracy of exploration and consider larger, \( n = 1000 \), copula sample for the tests we find that the H-GARCH model can not be rejected for the crucial \( l = 1 \)-lag whereas the FARIMA one is pretty much rejected. Had we gone further in the resolution, both models would have been rejected at all lags, meaning that neither of them are perfect.

Summarizing, we found that while the H-GARCH fits better in terms of the dominant 1-lag interdependence, the FARIMA captures better the approximate (as the weaker test was rather low resolution) decay of the interdependence, and this is quite understandable by the very nature of the models. The dominance of the 1-lag interdependence may explain, however, why does the FARIMA underestimate the quantiles of the observed series, in contrast to the H-GARCH, as shown in (Elek and Márkus 2004).
### Table 5: Autocopula Goodness-of-Fit Test: 95% critical values and the observed test statistics computed from samples with thinning parameter $s = 10$, sample size $n = 500$ for the first 5 lags

<table>
<thead>
<tr>
<th>Model</th>
<th>Hydro-GARCH, $n=500$</th>
<th>FARIMA, $n=500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l = 1$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00008</td>
<td>0.00026</td>
</tr>
<tr>
<td>Test St.</td>
<td><strong>0.00003</strong></td>
<td><strong>0.00013</strong></td>
</tr>
<tr>
<td>$l = 2$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00016</td>
<td>0.00057</td>
</tr>
<tr>
<td>Test St.</td>
<td><strong>0.00014</strong></td>
<td><strong>0.00038</strong></td>
</tr>
<tr>
<td>$l = 3$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00023</td>
<td>0.00080</td>
</tr>
<tr>
<td>Test St.</td>
<td><strong>0.00027</strong></td>
<td><strong>0.00068</strong></td>
</tr>
<tr>
<td>$l = 4$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00026</td>
<td>0.00096</td>
</tr>
<tr>
<td>Test St.</td>
<td><strong>0.00051</strong></td>
<td><strong>0.00122</strong></td>
</tr>
<tr>
<td>$l = 5$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00029</td>
<td>0.00142</td>
</tr>
<tr>
<td>Test St.</td>
<td><strong>0.00088</strong></td>
<td><strong>0.00209</strong></td>
</tr>
<tr>
<td>Model</td>
<td>Hydro-GARCH, n=1000</td>
<td>FARIMA, n=1000</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>$l = 1$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00004</td>
<td>0.00012</td>
</tr>
<tr>
<td>Test St.</td>
<td>0.00003</td>
<td>0.00010</td>
</tr>
<tr>
<td>$l = 2$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00008</td>
<td>0.00025</td>
</tr>
<tr>
<td>Test St.</td>
<td>0.00017</td>
<td>0.00041</td>
</tr>
<tr>
<td>$l = 3$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00010</td>
<td>0.00040</td>
</tr>
<tr>
<td>Test St.</td>
<td>0.00031</td>
<td>0.00070</td>
</tr>
<tr>
<td>$l = 4$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00012</td>
<td>0.00049</td>
</tr>
<tr>
<td>Test St.</td>
<td>0.00072</td>
<td>0.00151</td>
</tr>
<tr>
<td>$l = 5$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
<tr>
<td>95% c.v.</td>
<td>0.00012</td>
<td>0.00053</td>
</tr>
<tr>
<td>Test St.</td>
<td>0.00106</td>
<td>0.00248</td>
</tr>
</tbody>
</table>

Table 6: Autocopula Goodness-of-Fit Test: 95% critical values and the observed test statistics computed from samples with thinning parameter $s = 10$, sample size $n = 1000$ for the first 5 lags
6 Conclusions

We have introduced the concept of autocopula for the description of the lagged interdependence of time series. We have shown that autocopulas are efficient tools that provide deep insight into the structure and dynamics of times series. We have presented several goodness of fit procedures to check whether autocopulas of observed samples correspond to the ones originated from various models. The presented methods and tests are not exhaustive, as there are other options in the literature, for example those, based on Rosenblatt’s transform, cf. (Genest et al, 2009). Though our aim was to come up with tailor-made statistics for extreme-value analysis, comparisons and generalizations may well be of interest. We have demonstrated through quantitative evaluations that the proposed goodness of fit statistics help to choose between alternative models.

Although we limited our interest to the study of 2-dimensional autocopulas, there is no theoretical obstacle to use higher dimensional ones to obtain even stronger tests by handling certain combination of lags at the same time. For instance the set of k consecutive lags \( L = \{1, 2, \ldots, k\} \) would be particularly interesting. The presented computations can be carried out in a similar fashion for autocopulas of arbitrary set of lags \( L \), but of course at the expense of increased computation time, and with less visual appeal.

Our demonstrative example provided certainty that within a suitable parameter range the suggested tests for autocopulas are capable to detect the nonlinear ARCH character of and ARMA-ARCH series. The range of effectiveness of the tests is yet to be investigated, as are also comparisons with other nonlinearity tests.

In the context of hydrological application we chose to compare the fit of the specified heteroscedastic H-GARCH model and the linear FARIMA one. There are of course other popular models for describing hydrological data sets but the detailed investigation was beyond the scope of this paper. We found a further evidence, demonstrated by the fitted 1-lag autocopulas that the interdependence of daily river discharge series contain a significant nonlinear component. The studied H-GARCH model is better at describing short term interdependence, while the FARIMA does slightly better at longer terms. As short term interdependence is more dominant here, the H-GARCH fits better from other aspects as well, cf. (Elek and Márkus 2008). The obtained result underlines the need for model fit evaluation from the general dependence point of view, as an acceptable marginal fit may not be sufficient when characteristics associated with a whole temporal
period of observations e.g. flood duration or flood volume are considered. Having increased the accuracy of exploration, the fit of both models is rejected, meaning that neither of them are perfect. So, there remains room for other models to be considered. Among others, there is the one based on regime switches as suggested in Vasas et al. 2007 that we intend to study in a future work.

Acknowledgments

This research was partially supported by a grant provided for the first author within the framework of the Marie Curie RTN Seamocs project "Applied stochastic models for ocean engineering, climate and safe transportation".
References


REFERENCES


