Journal of Multivariate Analysis 101 (2010) 2103–2117

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# The pairwise beta distribution: A flexible parametric multivariate model for extremes

Daniel Cooley [a,*], Richard A. Davis [b], Philippe Naveau [c]

[a] Department of Statistics, Colorado State University, Fort Collins, CO, USA
[b] Department of Statistics, Columbia University, New York, NY, USA
[c] Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, France

ARTICLE INFO

*Article history:*
Received 3 July 2009
Available online 27 April 2010

*AMS 2010 subject classifications:*
62F99
62H99
62P12

*Keywords:*
Angular measure
Spectral measure
Multivariate regular variation
Threshold exceedances

ABSTRACT

We present a new parametric model for the angular measure of a multivariate extreme value distribution. Unlike many parametric models that are limited to the bivariate case, the flexible model can describe the extremes of random vectors of dimension greater than two. The novel construction method relies on a geometric interpretation of the requirements of a valid angular measure. An advantage of this model is that its parameters directly affect the level of dependence between each pair of components of the random vector, and as such the parameters of the model are more interpretable than those of earlier parametric models for multivariate extremes. The model is applied to air quality data and simulated spatial data.



## 1. Introduction

Fig. 1 shows three scatterplots of air pollutant measurements taken in the city centre of Leeds, UK. These data were analyzed in a discussion paper of Heffernan and Tawn [18] and were again recently analyzed by Boldi and Davison [2]. Since it is likely that the compound effects of high levels of multiple pollutants have more severe health consequences than the effects resulting from high levels of the individual pollutants, there is a need to model the data's joint upper tail. Interest in modeling multivariate extremes extends to many disciplines such as hydrology, finance, and engineering. There is a critical need to develop statistical methodologies for multivariate extremes for disciplines in which the assessment of risk associated with high levels of multiple components is of importance.

The probability theory which underlies the statistical practice for studying multivariate extremes is well developed. A classical work in multivariate extremes is [30], and the recent books by Beirlant et al. [1], de Haan and Ferreira [10] and Resnick [31] have large portions devoted to the multivariate case. Although the theory is well developed, there is still much room for work in developing statistical methodologies for analyzing and modeling multivariate extremes. In this paper we present a new and flexible parametric model for multivariate extremes of any order.

There are a number of flexible parametric models that exist for bivariate data such as the Gaussian model [19,35], bilogistic [21], and polynomial [26]. There are fewer models for higher-dimensional data, and many of these models have weaknesses such as a lack of flexibility, or conversely over-parametrization; these models are discussed in more depth in Section 3.

The parametric models mostly appeared in the literature in the late 1980s and early 1990s, and since then attention seems to have turned to other aspects of the study multivariate extremes. One area of interest has been in developing non-parametric and semi-parametric models for multivariate extremes. Much of this work has focused on the bivariate case.

* Corresponding author.
*E-mail address:* cooleyd@stat.colostate.edu (D. Cooley).

0047-259X/$ – see front matter 
doi:10.1016/j.jmva.2010.04.007

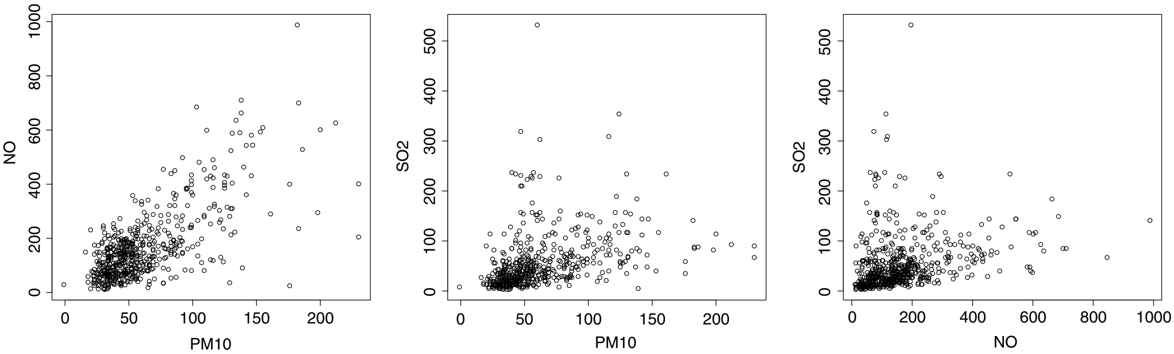


**Fig. 1.** Scatterplots of NO vs. PM10 (left), SO2 vs. PM10 (center), and SO2 vs. NO (right). The extremes of PM10 and NO appear to have relatively strong dependence, while the extremes of SO2 and the other two pollutants appear to have much weaker dependence.

Early non-parametric work [14] did not attempt to meet the required moment conditions of a multivariate extreme value model which are discussed in Section 2. In more recent work, Einmahl and Segers [15] use a non-parametric maximum empirical likelihood approach to fit an angular density to bivariate data which does meet the requirements of a multivariate extreme value distribution. The only semi- or non-parametric work done for dimension greater than 2 is that of [2]. Using a semi-parametric approach, they create a model for the angular density of a multivariate extreme value distribution by applying mixtures of Dirichlet distributions that meet the required moment conditions. Boldi and Davison describe two different fitting procedures for their model: the first is a Bayesian approach in which posterior draws are obtained via a reversible jump Markov chain Monte Carlo algorithm, and the second is an EM approach with AIC-based model selection. Either approach requires considerable effort to fit the model.

There has been separate work in describing the levels of dependence found in multivariate extremes. Several metrics that quantify the level of dependence in traditional max-stable random vectors have been suggested: the extremal coefficient studied extensively by Schlather and Tawn [33], the dependence measure $\chi(u)$ that appears in [5], the dependence measure $d(u, v)$ in [9], and the madogram, a first-order variogram, in [8]. With the exception of the extremal coefficient, these measures all quantify bivariate dependence. The complete bivariate dependence structure can be described by the Pickands dependence function, and estimators of this function have been proposed by Deheuvels [12], Capéraà and Fougéres [3] and Hall and Tajvidi [17], and an equivalent function, the $\lambda$-madogram has been studied by Naveau et al. [27]. Estimation of the Pickands dependence function is closely related to the bivariate non-parametric angular density estimation described above.

There has also been interest in developing models for max-stable processes, particularly for spatial problems. In a now famous unpublished manuscript, Smith [35] created a model for max-stable random processes using a point process to locate "storm centres" and "storm intensities". Schlather [32] extended Smith's point process idea to create a different spatial model, and we use Schlather's model to simulate fields in Section 5.2. Most recently, de Haan and Pereira [11] provided several models for spatial extremes. While all these are models for max-stable processes, only the bivariate joint distribution is known in closed form.

Others have turned their attention to modeling dependence under the class of asymptotic independence. A bivariate couple $(X_1, X_2)$ is termed asymptotically independent if $\lim_{x\to\infty} \mathbb{P}(X_2 > x | X_1 > x) = 0$. Papers by Ledford and Tawn [23–25] spurred interest in describing and modeling dependence under the class of asymptotic independence. Ledford and Tawn [25], Peng [28] and Draisma et al. [13] all developed measures for the coefficient of tail dependence which describes the amount of dependence under the case of asymptotic independence for the bivariate case. A paper by Heffernan and Tawn [18] provided models for extremes which included the case of asymptotic independence, but the models were developed via bivariate conditional relationships, and higher-dimensional relationships have not been made explicit. A recent paper by Ramos and Ledford [29] offers a parametric model which captures both asymptotic dependence and independence again in the bivariate case.

The work presented here extends the early work done in modeling multivariate extremes. We believe that developing useful multivariate models that can readily be applied by practitioners in various fields is important work, and that a need exists for new parametric models for multivariate extremes of dimension greater than two. An advantage of a parametric approach is that it allows for easy model fitting procedures that do not require advanced computational techniques. Another advantage is that often the parameters lend themselves to interpretation, and this is the case for the pairwise beta model. The model is classical in the sense that it is a model under the case of asymptotic dependence. The constructive approach we use for model formulation is novel.

The remainder of the article is organized as follows. In Section 2 we briefly summarize the necessary background in multivariate extreme value theory. Section 3 first reviews the parametric models for dimension $p > 2$ that have previously appeared in the literature, and then introduces a new parametric model, the pairwise beta. Section 4 details how the model can be fit to high observations and tests the procedure on data simulated from the model. Section 5 applies the model to the air quality data referenced above and also to simulated spatial data. The paper concludes with a summary section.

## 2. Multivariate extreme values and the angular measure

The aim of a multivariate extreme value analysis is to characterize the joint upper tail of a distribution. It is common practice to analyze only the data which are considered to be extreme. Two approaches for choosing the subset of the data to be analyzed are to extract block (e.g., annual) maximum values or alternatively, to retain only observations which exceed some threshold. In both cases, asymptotic results from probability theory provide a framework for modeling the selected data. One way of characterizing the dependence for the limiting distributions of both block maxima and threshold exceedances is via an angular (or spectral) measure.

### 2.1. Regular variation and threshold exceedances

An approach used to characterize threshold exceedances is via the concept of regular variation. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^T \geq \boldsymbol{0}$ be a random vector with distribution $F$, and define $\mathcal{C}$ to be the set $[\boldsymbol{0}, \infty] \setminus \boldsymbol{0}$. Then $\boldsymbol{Z}$ is regularly varying if

$$\frac{\mathbb{P}(t^{-1}\boldsymbol{Z} \in \cdot)}{\mathbb{P}(\|\boldsymbol{Z}\| > t)} \xrightarrow{v} \nu(\cdot), \tag{1}$$

where $v$ denotes vague convergence [31] and $\|\cdot\|$ is any norm on $\mathcal{C}$[1]. The measure $\nu$ has the scaling property

$$\nu(sA) = s^{-\alpha}\nu(A) \tag{2}$$

for all Borel sets $A \in \mathcal{C}$, and $\alpha$ is called the tail index. Choosing the sequence $\{a_n\}$ such that $\mathbb{P}(\|\boldsymbol{Z}\| > a_n) \sim n^{-1}$, we have the sequential relation of (1)

$$n\mathbb{P}\left(\frac{\boldsymbol{Z}}{a_n} \in \cdot\right) \xrightarrow{v} \nu(\cdot). \tag{3}$$

The scaling property (2) suggests a transformation to polar coordinates. Define $R = \|\boldsymbol{Z}\|$ and $\boldsymbol{W} = \boldsymbol{Z}\|\boldsymbol{Z}\|^{-1}$, and let $S_{p-1} = \{\boldsymbol{z} \in \mathcal{C} : \|\boldsymbol{z}\| = 1\}$ be the unit sphere in $p$-dimensional space under the chosen norm. Then there exists a probability measure $H$ on $S_{p-1}$ such that

$$n\mathbb{P}\left(\frac{R}{a_n} > r, \boldsymbol{W} \in B\right) \to r^{-\alpha}H(B), \tag{4}$$

for all $H$-continuity sets $B$. Conceptually, it is often helpful to view the angular measure $H$ as the limiting distribution of $\boldsymbol{W}$ for $R$ large, i.e.,

$$\mathbb{P}(\boldsymbol{W} \in \cdot | R > t) \xrightarrow{d} H(\cdot)$$

as $t \to \infty$. If $H$ is absolutely continuous on $S_{p-1}$, then we denote its density by $h(\boldsymbol{w})$.

There is a link between multivariate regular variation and point process convergence. Specifically, if $\{\boldsymbol{Z}_n\}$ is an iid sequence then (4) is equivalent to convergence in distribution of the point process with points at $\boldsymbol{Z}_1/a_n, \ldots, \boldsymbol{Z}_n/a_n$ to a Poisson random measure (PRM) with intensity measure $\nu(\cdot)$. Transforming to polar coordinates, we have $\{(a_n^{-1}R_i, \boldsymbol{W}_i), i = 1, \ldots, n\}$ converges to a PRM$(r^{-(\alpha+1)}\mathrm{d}r \times \mathrm{d}H(\boldsymbol{w}))$. It is via this point process representation that we will fit our angular measure model to threshold exceedances in Section 4

We aim to construct a parametric model for $H$. For the general case of multivariate regular variation described above, $H$ can be any probability measure, and thus constructing a parametric model for the angular measure is infeasible. However, it is common practice in multivariate extremes to assume that the components $Z_i$, $i = 1, \ldots, p$ of the random vector have a common marginal distribution, not just the common tail index that is required under the general conditions of multivariate regular variation. We assume that $Z_i$, $i = 1, \ldots, p$ have a common marginal distribution $F_1$ which is regularly varying with index $\alpha = 1$. If the data we intend to model arise from a random vector $\boldsymbol{Y}$ for which this is not the case, we assume probability integral transforms $T_i$ are applied so that $T_i(Y_i) = Z_i$ and $Z_i$ has the marginal $F_1$. In [2] it was assumed that $F_1(z) = \exp(-z^{-1})$, the standard unit Fréchet distribution.

Assuming $\alpha = 1$, then we have for the $i$th marginal component,

$$\begin{aligned} n\mathbb{P}\left(\frac{Z_i}{a_n} > z\right) &\to \nu\{\boldsymbol{x} \in \mathcal{C} : x_i > z\} \\ &= \int_{S_{p-1}} \int_{\frac{z}{w_i}}^{\infty} r^{-2}\mathrm{d}r\mathrm{d}H(\boldsymbol{w}) \\ &= \frac{1}{z}\int_{S_{p-1}} w_i \mathrm{d}H(\boldsymbol{w}). \end{aligned}$$

[1] Others choose to normalize by $\mathbb{P}(Z_1 > t)$, see [31, p. 174].

Since we have assumed a common marginal, this implies that

$$\int_{S_{p-1}} w_1 \mathrm{d}H(\boldsymbol{w}) = \int_{S_{p-1}} w_j \mathrm{d}H(\boldsymbol{w}) \tag{5}$$

for all $j = 2, \ldots, p$. The moment conditions in (5) are the only requirements for a valid angular measure model with identically distributed marginals.

When $\alpha = 1$ it is particularly useful to choose the $L_1$ norm: $\|\boldsymbol{z}\| = z_1 + \cdots + z_p$, for which the unit sphere is the simplex $S_{p-1} = \{\boldsymbol{w} \in \mathcal{C} : w_1 + \cdots + w_p = 1\}$. With this norm, $1 = \int_{S_{p-1}} \mathrm{d}H(\boldsymbol{w}) = \int_{S_{p-1}} (w_1 + \cdots + w_p)\mathrm{d}H(\boldsymbol{w})$ and hence $\int_{S_{p-1}} w_i \mathrm{d}H(\boldsymbol{w}) = p^{-1}$.

Using the $L_1$ norm, the moment conditions (5) on the angular measure have a helpful geometric interpretation. They imply that the center-of-mass ($\int_{S_{p-1}} w_i \mathrm{d}H(\boldsymbol{w})$ for $i = 1, \ldots, p$) of $H$ must be at the point $\boldsymbol{w} = (1/p, \ldots, 1/p)$. Dependence increases between all components as the mass of $H$ moves to the center of the simplex, the components become less dependent as the mass increases near the vertices of the simplex. This geometric interpretation leads to our constructive approach in Section 3.2.

### 2.2. Multivariate extreme value distributions and block maxima

Rather than modeling threshold exceedances, the more classical approach to studying extremes is to model block maximum data. In the multivariate case, the definition of maximum is ambiguous. Classical multivariate extreme value theory describes the behavior of the vector constructed from the componentwise maxima. The family of multivariate extreme value distributions (MEVDs) are the limiting distributions of componentwise block maxima, and the MEVDs can again be characterized by the angular measure.

We first characterize the MEVD which corresponds to the random vector $\boldsymbol{Z}$ described in the previous section. Let $\boldsymbol{Z}_m = (Z_{m,1}, \ldots, Z_{m,p})^T$, $m = 1, 2, \ldots$ be independent and identically distributed copies of $\boldsymbol{Z}$, and let the vector of componentwise maxima be denoted by $\boldsymbol{M}_n = \left(\bigvee_{m=1,\ldots,n} Z_{m,1}, \ldots, \bigvee_{m=1,\ldots,n} Z_{m,p}\right)^T$, where $\bigvee$ denotes max. We assume there exists a distribution function $G$ such that

$$\mathbb{P}\left(\frac{\boldsymbol{M}_n}{a_n} \le \boldsymbol{z}\right) = \mathbb{P}^n\left(\frac{\boldsymbol{Z}}{a_n} \le \boldsymbol{z}\right) \xrightarrow{d} G(\boldsymbol{z}), \tag{6}$$

where $\{a_n\}$ is defined as above. Taking logarithms and applying Taylor series approximations to (6), we obtain

$$n \log\left[1 - \mathbb{P}\left(\frac{\boldsymbol{Z}}{a_n} > \boldsymbol{z}\right)\right] \approx -n\mathbb{P}\left(\frac{\boldsymbol{Z}}{a_n} \in [\boldsymbol{0}, \boldsymbol{z}]^c\right) \approx \log G(\boldsymbol{z}), \tag{7}$$

which combined with (3) gives us

$$G(\boldsymbol{z}) = \exp(-\nu[\boldsymbol{0}, \boldsymbol{z}]^c) = \exp\left(-\int_{S_{p-1}} \max_{i=1,\ldots,p}\left(\frac{w_i}{z_i}\right) \mathrm{d}H(\boldsymbol{w})\right), \tag{8}$$

relating the MEVD to the angular measure $H$.

Eq. (8) differs slightly from the representation of the family of MEVDs given by Coles and Tawn [6]. The representations' marginals differ by a constant, and this difference can be attributed to how the normalizing sequence $\{a_n\}$ is chosen. Choosing to have standard unit Fréchet marginals, Coles and Tawn [6] characterize the family of MEVDs as

$$G^*(\boldsymbol{z}) = \exp(-V(\boldsymbol{z})), \tag{9}$$

where $V(\boldsymbol{z})$ is termed the exponent measure function. In terms of our choice of $\{a_n\}$ above, the exponent measure function is

$$V(\boldsymbol{z}) = p \int_{S_{p-1}} \max_{i=1,\ldots,p}\left(\frac{w_i}{z_i}\right) \mathrm{d}H(\boldsymbol{w}). \tag{10}$$

The exponent measure function is simply a way of relating the angular measure, which is best understood in polar coordinates, to the distribution function, which requires Cartesian coordinates. Rather than via the regular variation argument described in the previous section, it is via this representation of the family of MEVDs that Coles and Tawn [6] and others have described the moment conditions (5) of the angular measure.

## 3. Parametric models for multivariate extremes

It is not possible to construct a parametric family of models that exhausts the entire class of angular measures $H$ satisfying (5). Nevertheless, there have been some parametric subfamilies suggested which can capture important behavior and which have been used successfully in many modeling applications. The book by Kotz and Nadarajah [22] gives a collection of models for MEVDs. As mentioned in the introduction, several models have been suggested for the bivariate case, in this section we concentrate on models which can handle any finite dimension. In what follows, we use $\boldsymbol{\theta}$ generically to denote the vector of parameters associated with a model.

### 3.1. Previous parametric models

The parametric models can be divided into two classes. The first class gives a parametric model for the exponent measure function, $V(\boldsymbol{z}; \boldsymbol{\theta})$, which provides closed-form expression for the joint distribution of the MEVD. The second class parametrically models the angular density $h(\boldsymbol{w}; \boldsymbol{\theta})$ directly.

Defining a multivariate extreme value distribution via a parametric exponential measure function is challenging, and few models for dimension $p > 2$ have been suggested. The most widely known MEVD is the logistic [16] which has exponent measure function $V(\boldsymbol{z}; \gamma) = (\sum_{i=1}^{p} z_i^{-1/\gamma})^{\gamma}$ for $0 < \gamma \leq 1$. Dependence between the components increases as $\gamma$ decreases. The logistic is easy to work with because its exponential measure function is relatively simple and leads to an easy representation for its angular measure which (if $\gamma \neq 1$) exists entirely on the interior of $S_{p-1}$. However, because it is characterized by a single parameter $\gamma$, in higher dimensions it is inadequate to model situations where dependence between components differs. As such, it has primarily been used in bivariate applications (e.g. [36]).

The asymmetric logistic [37] and the negative logistic [20] are similar models which extend the logistic model to allow different levels of dependence between the components. Both models give explicit definitions for the exponent measure function $V(\boldsymbol{z}, \boldsymbol{\theta})$. One difficulty of these models is that they have a large number of parameters; in the three-dimensional case the asymmetric logistic has sixteen parameters, twelve of which can be freely chosen. Another potential limitation of these models is that both achieve the center-of-mass condition by putting mass on the edges and vertices of the simplex $S_{p-1}$, resulting in a discontinuous angular measure. No asymmetric parametric model for $V(\boldsymbol{z}, \boldsymbol{\theta})$ has been proposed with a continuous angular measure.

Rather than defining a parametric model for $V(\boldsymbol{z}; \boldsymbol{\theta})$, one can alternatively define parametric models for the angular density $h(\boldsymbol{w}; \boldsymbol{\theta})$. Coles and Tawn [6] describe one method of obtaining a model for $h(\boldsymbol{w}; \boldsymbol{\theta})$. They show that if $h^*$ is a positive function on $S_{p-1}$ with finite first moments $m_i = \int_{S_{p-1}} w_i h^*(\boldsymbol{w}; \boldsymbol{\theta}) \mathrm{d}(\boldsymbol{w})$, then

$$h(\boldsymbol{w}; \boldsymbol{\theta}) = \frac{1}{p} (\boldsymbol{m} \cdot \boldsymbol{w})^{-(p+1)} \prod_{j=1}^{p} m_j h^* \left( \frac{m_1 w_1}{\boldsymbol{m} \cdot \boldsymbol{w}}, \ldots, \frac{m_p w_p}{\boldsymbol{m} \cdot \boldsymbol{w}}; \boldsymbol{\theta} \right) \tag{11}$$

is a valid angular density which has all its mass on the interior of $S_{p-1}$. In effect, if one thinks of $h^*$ as a (perhaps unnormalized) density on $S_{p-1}$, then (11) alters the density so that it has center-of-mass at $(1/p, \ldots, 1/p)$ and total mass of 1. Coles and Tawn [6] used their technique to create a multivariate extreme value model from the Dirichlet density, a well-known density on the unit simplex which in $p$-dimensions is parameterized by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$ and whose pdf is given by

$$h^*(\boldsymbol{w}; \boldsymbol{\alpha}) = \frac{\Gamma(\boldsymbol{\alpha} \cdot \boldsymbol{1})}{\prod_{j=1}^{p} \Gamma(\alpha_j)} \prod_{j=1}^{p} w_j^{\alpha_j - 1}, \quad \alpha_j > 0, j = 1, \ldots, p.$$

As the Dirichlet density has moments $m_i = \alpha_i / \left( \sum_{j=1}^{p} \alpha_j \right)^{-1}$, applying (11) one obtains the angular density

$$h(\boldsymbol{w}; \boldsymbol{\alpha}) = \frac{1}{p} \prod_{j=1}^{p} \frac{\alpha_j}{\Gamma(\alpha_j)} \frac{\Gamma(\boldsymbol{\alpha} \cdot \boldsymbol{1} + 1)}{(\boldsymbol{\alpha} \cdot \boldsymbol{w})^{p+1}} \prod_{j=1}^{p} \left( \frac{\alpha_j w_j}{\boldsymbol{\alpha} \cdot \boldsymbol{w}} \right)^{\alpha_j - 1}$$

which can be asymmetric. This angular density model has been termed the tilted Dirichlet to distinguish it from the Dirichlet density above.

Compared to the parametric models for $V(\boldsymbol{z}, \boldsymbol{\theta})$, modeling $h(\boldsymbol{w}; \boldsymbol{\theta})$ directly allows for more flexibility in how the angular measure behaves in the interior of the simplex. Consequently, Coles [4] found the tilted Dirichlet model preferable to the logistic and negative logistic models when fitting spatial rainfall extreme data. A disadvantage of the tilted Dirichlet model is that after application of (11), the angular density's parameters become largely uninterpretable. Given any model for $h(\boldsymbol{w}; \boldsymbol{\theta})$, one must perform the integration in (8) to obtain an expression for the extreme value distribution. Since this integration must generally be done numerically, models for $h(\boldsymbol{w}; \boldsymbol{\theta})$ are more useful for describing threshold exceedances than for block maxima. To date, the tilted Dirichlet is the only parametric model for $h(\boldsymbol{w}, \boldsymbol{\theta})$ which has appeared in the literature.

### 3.2. A constructive approach for angular density models

The center-of-mass interpretation of condition (5) provides us with inspiration for directly constructing models for the angular density $h(\boldsymbol{w}; \boldsymbol{\theta})$. Our approach for constructing an angular density model is geometric and differs from that of Coles and Tawn [6].

Consider the function

$$h_{i,j}(\boldsymbol{w}; \beta_{i,j}) = \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j}-1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j}-1}$$

where $\boldsymbol{w} \in S_{p-1}$ and $\beta_{i,j} > 0$. We refer to $h_{i,j}(\boldsymbol{w}; \beta_{i,j})$ as the pairwise beta function as it is simply a symmetric beta density between two components on the simplex. By following reasoning similar to Theorem 1 in the Appendix, it can be shown that for any parameter $\beta_{i,j}$, the pairwise beta function has center-of-mass at $(1/p, \ldots, 1/p)$.

The pairwise beta function provides a foundation for constructing angular density functions. An obvious method for constructing a valid angular density is to sum each of the $\binom{p}{2}$ pairwise beta functions for each pair of components in our random vector. However, a simple sum of pairwise beta functions yields a model which is not entirely satisfying. Using a simple sum, if just one of the pairwise beta functions has relatively strong dependence, this will create mass at the center of the simplex, which in turn causes all of the components to have some minimum level of dependence. To alleviate this issue, in the model below we add an additional global parameter $\alpha$ to help control the overall dependence in the model.

Let

$$h(\boldsymbol{w}; \alpha, \boldsymbol{\beta}) = K_p(\alpha) \sum_{1 \le i < j \le p} h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}), \tag{12}$$

$$\text{where} \quad h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}) = (w_i + w_j)^{2\alpha-1}(1 - (w_i + w_j))^{\alpha(p-2)-p+2} \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j}-1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j}-1},$$

$$\text{and} \quad K_p(\alpha) = \frac{2(p-3)!}{p(p-1)\sqrt{p}} \frac{\Gamma(\alpha p + 1)}{\Gamma(2\alpha + 1)\Gamma(\alpha(p-2))}$$

be known as the pairwise beta model. It is shown in Theorem 1 in the Appendix that the pairwise beta meets (5) and thus is a valid angular density.

The function $h_{i,j}$ in (12) consists of two pieces. The first piece of $h_{i,j}$ is determined by the parameter $\alpha > 0$ which simply draws the mass of the density toward the center of the simplex as it increases. The second piece is the pairwise beta function defined above. A feature of the model is that its parameters are easily interpretable; $\alpha$ is a global parameter which controls the overall level of dependence in the model and each of the $\beta_{i,j}$ parameters controls the level of dependence between the $i$th and $j$th components of the random vector. The factor $K_p(\alpha)$ is a normalizing constant.

Fig. 2 shows four examples of angular measures given by the three-dimensional pairwise beta model. The top left figure has parameters $\alpha = 1$ and $\boldsymbol{\beta} = (2, 4, 15)$, and the angular measure clearly shows strong dependence between the second and third components due to the large value of $\beta_{2,3}$. The top right figure increases the global parameter ($\alpha = 4$) while leaving the $\boldsymbol{\beta}$ values unchanged, and one can see how the mass of the angular measure is pulled to the center of the simplex. The lower left shows a plot where the global parameter is small ($\alpha = 7/12$) but one still sees that there is still relatively strong dependence between the second and third components, as the mass is either located near the point $(1, 0, 0)$ or near $(0, 1/2, 1/2)$. The lower right plot has $\alpha = 1$ and $\boldsymbol{\beta} = (2, 2, 1/2)$ and the low value for $\beta_{2,3}$ drives the mass of the angular measure to the boundaries indicating that large values of components two and three are unlikely to occur at the same time.

To fully describe the dependence in the extremes requires knowledge of $H(\boldsymbol{w})$ or alternatively $V(\boldsymbol{z})$. However, as mentioned in Section 1, several related dependence metrics have been proposed which describe the level of dependence for max-stable random vectors. If $\boldsymbol{Z}$ is max-stable with unit Fréchet margins, the pairwise extremal coefficient [35,34] is given by $\phi_{i,j}(\boldsymbol{Z}) = V(\boldsymbol{c}_{i,j})$, where $\boldsymbol{c}_{i,j}$ is the $p$-dimensional vector with components $c_i = 1$, $c_j = 1$ and $c_k = \infty$ for $k \neq i$ or $j$. The dependence metric can be understood via the relation

$$\mathbb{P}(\max(Z_i, Z_j) \le z) = \exp[-V(z\boldsymbol{c}_{i,j})] = \exp[-z^{-1}V(\boldsymbol{c}_{i,j})] = (\exp[-z^{-1}])^{\phi_{i,j}} = (\mathbb{P}(Z_1 \le z))^{\phi_{i,j}};$$

that is, the pairwise extremal coefficient is the effective number of independent random variables in the bivariate couple. Hence $\phi_{i,j} \in [1, 2]$ and as $\phi_{i,j}$ increases, the amount of dependence decreases. Schlather and Tawn [34,33] describe higher-order extremal coefficients, but below we focus only on pairwise dependence.

Theorem 2 in the Appendix more clearly illustrates how the parameter $\beta_{i,j}$ affects the pairwise dependence between the $i$th and $j$th components. Let $\boldsymbol{Z}$ be a max-stable random vector with angular measure described by the pairwise beta model with parameters $(\alpha, \boldsymbol{\beta})$, and let $\boldsymbol{Z}^*$ be another pairwise beta max-stable random vector with parameters $(\alpha, \boldsymbol{\beta}^*)$, where $\beta^*_{i,j} > \beta_{i,j}$ and $\beta^*_{k,l} = \beta_{k,l}$ for $(k, l) \neq (i, j)$. Theorem 2 shows that $\phi_{i,j}(\boldsymbol{Z}^*) < \phi_{i,j}(\boldsymbol{Z})$; that is, the dependence between the $i$th and $j$th components of the random vectors as measured by the extremal coefficient is stronger for $\boldsymbol{Z}^*$ than for $\boldsymbol{Z}$.

Several comments should be made regarding the dependence in the pairwise beta model. The existing extremal dependence metrics such as the extremal coefficient are all defined in terms of max-stable random vectors and the exponent measure function and are only indirectly related to the angular measure. Consequently, the proof of Theorem 2 is somewhat

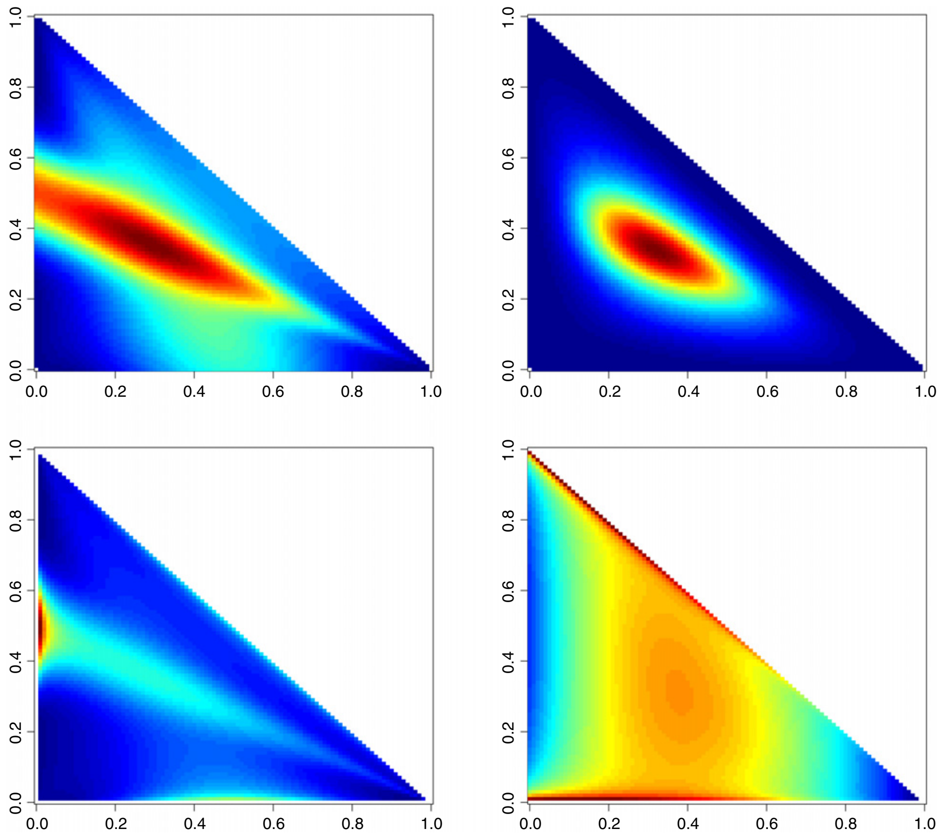


**Fig. 2.** Four examples of the pairwise beta density. The lower right corner of the triangle in each panel corresponds to an event where the first element of the random vector is large while the other two are small. The upper and lower left corners correspond to the second and third elements being large. The parameter values for the top left panel are $\alpha = 1$, $\boldsymbol{\beta} = (2, 4, 15)$. One sees in the plot that the second and third components are likely to be large at the same time, due to parameter $\beta_{2,3} = 15$. The parameters for the other plots are $\alpha = 4$, $\boldsymbol{\beta} = (2, 4, 15)$ (top right), $\alpha = 7/12$, $\boldsymbol{\beta} = (2, 4, 15)$ (bottom left), and $\alpha = 1$, $\boldsymbol{\beta} = (2, 2, 1/2)$ (bottom right).

circuitous. Second, the additive nature of the pairwise beta model implies that the $\beta_{i,j}$ parameters do not act independently. It would be very satisfying to find an parametric angular measure model where the $i$th and $j$th components were completely controlled by a single parameter and unaffected by the others. Despite the fact that the $\beta_{i,j}$'s do not act independently, the pairwise beta has a flexibility not found in other existing models. The $\alpha$ parameter sets a level of overall dependence which can then be adjusted either up or down by the $\beta_{i,j}$ parameters allowing one to achieve an appropriate dependence level for each pair of components.

## 4. Estimation procedure

Fitting a model for an angular density is a relatively straightforward exercise. Given a set of iid observations $\boldsymbol{y}_m$, $m = 1, \ldots, n$, one first fits distributions to the marginals, and then transforms $\boldsymbol{z}_m = T(\boldsymbol{y}_m)$ to have a common marginal with tail index $\alpha = 1$. One then makes a further transformation to pseudo-polar coordinates yielding points $(r_m, \boldsymbol{w}_m)$ where $r_m = \|\boldsymbol{z}_m\|$ and $\boldsymbol{w}_m = \boldsymbol{z}_m \|\boldsymbol{z}_m\|^{-1}$. A high threshold $t_0$ is selected and the points $\{(r_m, \boldsymbol{w}_m), m = 1, \ldots, n : r_m > t_0\}$ are retained. Let $(r_{(m)}, \boldsymbol{w}_{(m)})$, $m = 1, \ldots, N_{t_0}$ denote the reindexed threshold exceedances. Given that $t_0$ is large enough, we assume that the points $(r_{(m)}, \boldsymbol{w}_{(m)})$ approximately follow a Poisson process with intensity measure $\nu$ given in (3). Letting $A = \{(r, \boldsymbol{w}) : r > t_0\}$ the approximate likelihood [1, pp. 170–171] of the points $(r_{(m)}, \boldsymbol{w}_{(m)})$, $m = 1, \ldots, N_{t_0}$ is given by

$$L(\boldsymbol{\theta}; (r_{(m)}, \boldsymbol{w}_{(m)}), m = 1, \ldots, N_{t_0}) \approx \exp(-\nu(A)) \prod_{m=1}^{N_{t_0}} d\nu(r_{(m)}, \boldsymbol{w}_{(m)}) = \exp(-t_0^{-1}) \prod_{m=1}^{N_{t_0}} r_{(m)}^{-2} h(\boldsymbol{w}_{(m)}, \boldsymbol{\theta}),$$

where $h(\boldsymbol{w}; \boldsymbol{\theta})$ is any parametric model for the angular measure. To find $\boldsymbol{\theta}$ which maximizes this likelihood, we need to only note that $L(\boldsymbol{\theta}; (r_{(m)}, \boldsymbol{w}_{(m)}), m = 1, \ldots, N_{t_0}) \propto \prod_{m=1}^{N_{t_0}} h(\boldsymbol{w}_{(m)}, \boldsymbol{\theta})$. The estimate $\hat{\boldsymbol{\theta}}$ can then be found via numerical optimization. This estimation procedure was used by both [7] and [2]. Since the pairwise beta angular measure is a smooth function of $\alpha \in (0, \infty)$ and $\boldsymbol{\beta} \in (0, \infty)^{\binom{p}{2}}$ and has bounded support on the unit simplex $S_{p-1}$, if one assumes the marginal distributions are known, standard asymptotics hold for the maximum likelihood estimators $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$.

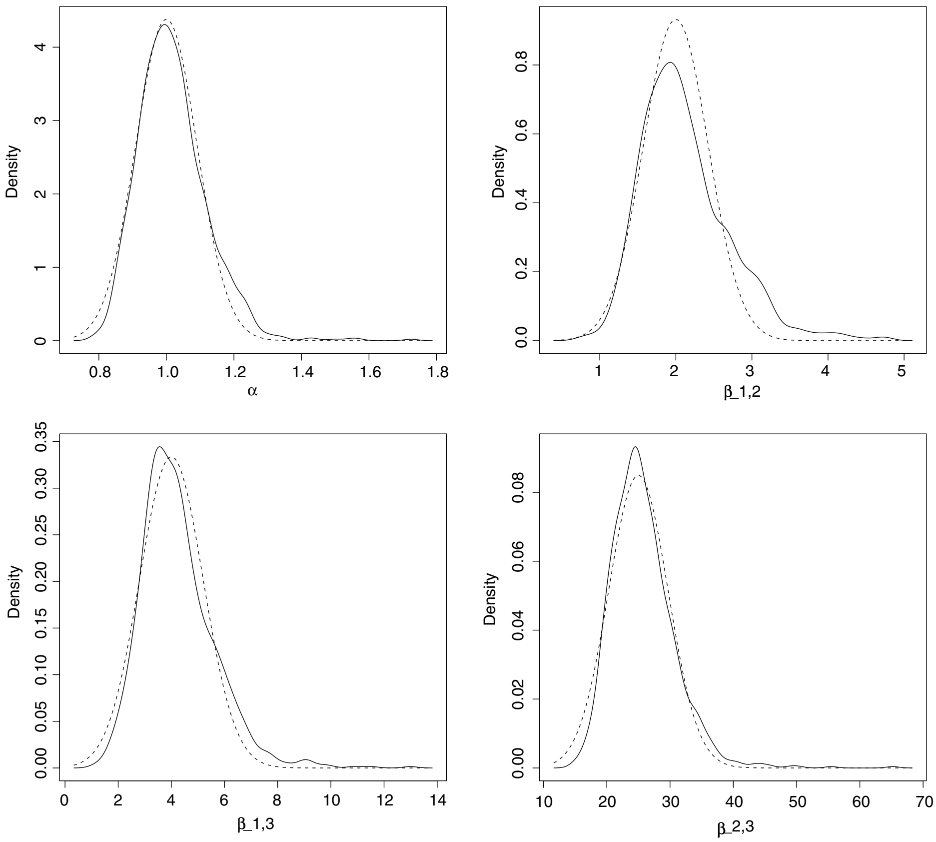


**Fig. 3.** Density estimates from 1000 simulations, each consisting of 200 realizations of the pairwise beta angular measure with parameters $\alpha = 1$, and $\boldsymbol{\beta} = (2, 4, 15)$. All figures show that the maximum likelihood estimators for $N_{t_0} = 200$ appear to have a distribution with mean and mode near the actual parameter value and are approaching normality, although all appear to be slightly positively skewed. Dashed line is the (numerically obtained) normal distribution suggested by the asymptotics for MLEs with $N_{t_0} = 200$.

**Table 1**

Summary of the pairwise beta simulation results. Table gives the true values of the parameters, the mean of the MLE estimates for the 1000 simulations, the standard error of the estimates suggested by the asymptotics, and the sample standard error of the estimates. The means of the estimates are slightly larger than the actual parameter values and the sample standard errors are slightly larger than the asymptotic estimates; both results are presumably due to the skewness seen in Fig. 3.

| | $\alpha$ | $\beta_{1,2}$ | $\beta_{1,3}$ | $\beta_{2,3}$ |
|---|---|---|---|---|
| Value | 1 | 2 | 4 | 15 |
| Mean | 1.018 | 2.119 | 4.233 | 15.707 |
| Asymp SE | 0.091 | 0.428 | 1.196 | 4.699 |
| SD of ests | 0.103 | 0.586 | 1.380 | 5.023 |

To test the estimation procedure for this model, a simulation exercise was performed. For each simulation, two hundred realizations of angular components $\boldsymbol{w}_m$ were generated according to the pairwise beta angular measure via an accept–reject algorithm. For the simulation, the parameters of the pairwise beta were set at $\alpha = 1$, $\boldsymbol{\beta} = (2, 4, 15)$. These realizations of the angular components $\boldsymbol{w}_m$ were assumed to correspond with realizations of $\boldsymbol{z}_m$ with large radial components. The pairwise beta model was then fit via the method described above. This experiment was repeated 1000 times and the maximum likelihood estimates were recorded. Typically, the angular measure only describes the angular component in the limit as $r_m \to \infty$; however, for these simulated points, all observations $(r_m, \boldsymbol{w}_m)$ such that $r_m > t_0$ follow the pairwise beta model exactly.

Fig. 3 shows density estimates formed from the maximum likelihood estimates of these 1000 simulations. The information matrix associated with this likelihood is analytically intractable but is easily numerically approximated for this three-dimensional case, and the dashed lines show the normal distribution suggested by the asymptotics. All panels show that the distribution of the estimators appears to have a mean and mode near the actual parameter values and that, although slightly positively skewed, the distributions are approaching normality for $N_{t_0} = 200$. Table 1 summarizes the results of the simulations and indicates that the standard deviations of the estimates are slightly greater than the asymptotics suggest, presumably due to the skewness seen in Fig. 3. The coverage probabilities for the asymptotic 95% confidence intervals for these 1000 simulations were 0.967 for $\alpha$ and (0.934, 0.935, 0.937) for $(\beta_{1,2}, \beta_{1,3}, \beta_{2,3})$.

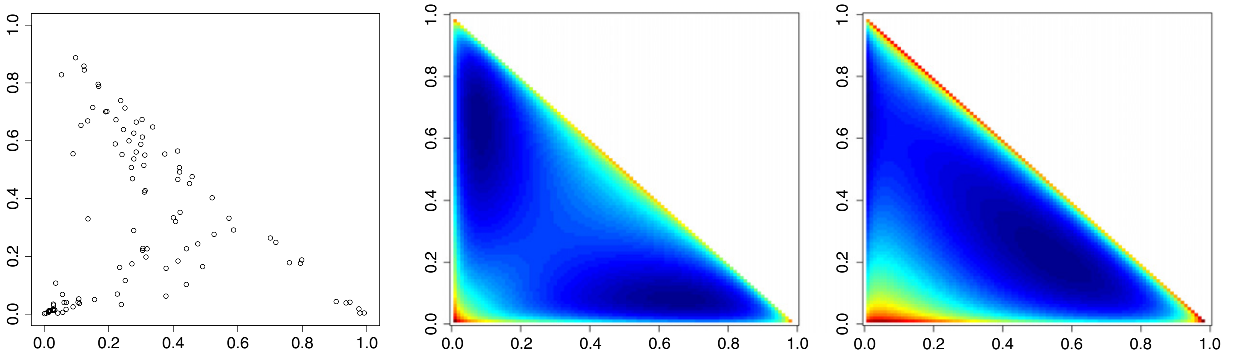


**Fig. 4.** Left plot shows the angular components of the largest 100 observations of the trivariate air quality data. The lower right corner corresponds to large values of PM10 and small values of NO and SO2. The upper corner corresponds to large values of NO, and the lower left corresponds to large values of SO2. As points tend to lie along the hypotenuse of the triangle and in the lower left corner, this indicates that large values of PM10 and NO tend to occur together, while large values of SO2 occur independently. The center plot shows the log density of the fitted pairwise beta model, while right plot shows the log density of the fitted Dirichlet model.

## 5. Applications

### *5.1. Air quality data*

We examine a set of air quality data which has been analyzed by Heffernan and Tawn [18] and more recently by Boldi and Davison [2]. The data were taken in the city centre of Leeds, UK and are daily maximum measurements for five different air pollutants: particulate matter (PM10), nitrogen oxide (NO), nitrogen dioxide (NO2), ozone (O3), and sulfur dioxide (SO2). The data were downloaded from http://www.airquality.co.uk. We focus on the data collected during the winter season (November–February) and to be consistent with the previous studies, we examine data for the years 1994–1998.

For illustrative purposes, we restrict our attention to the trivariate data of PM10, NO, and SO2. The scatterplots in Fig. 1 indicate that PM10 and NO exhibit relatively strong extremal dependence, while SO2 has much weaker dependence with the other two pollutants. Heffernan and Tawn [18] used a conditional approach to model the dependence found in each pair of pollutants. Like [2], we approach this data in the manner of a traditional extreme value problem and fit both the pairwise beta and tilted Dirichlet models to the trivariate observations. Similar to [18], we transform each marginal distribution by fitting a generalized Pareto distribution to the exceedances of the empirical 0.7 quantile, and using the empirical distribution function below the threshold; however, we transform to a unit Fréchet marginals whereas Heffernan and Tawn transform to Gumbel marginals. The 100 trivariate observations with the largest radial components were then selected, and their angular components were used to fit the two models. The plot of the angular components can be seen in Fig. 4. Many of the points lie along the hypotenuse of the triangle, indicating large simultaneous values of PM10 and NO and a small value for SO2; conversely, the cluster of points at the lower left corner indicates large values of SO2 and small values of PM10 and NO.

When fit to the data, the pairwise beta model yields a log-likelihood of 41.02 and an AIC value of −74.04. Its parameter estimates are $\hat{\alpha} = 0.68$, and $\hat{\boldsymbol{\beta}} = (3.21, 0.47, 0.45)$, with the large $\hat{\beta}_{1,2}$ value indicating the stronger dependence between PM10 and NO as expected. The estimated standard errors of these estimates are respectively 0.009, 0.101, 0.010, and 0.010. We note that these standard errors account only for uncertainty due to estimation of the parameters of the angular measure model and do not reflect the uncertainty generated by first estimating the marginals, the use of a limiting angular measure to fit threshold exceedances, or the assumption of the parametric model choice. The middle panel of Fig. 4 shows the fitted pairwise beta angular measure, which has increased mass along the hypotenuse and in the lower left vertex as expected. The tilted Dirichlet model yields a log-likelihood of 34.84 and an AIC value of −63.68, indicating that the pairwise beta model yielded a better fit. The tilted Dirichlet model parameter estimates are $\hat{\boldsymbol{\alpha}} = (1.20, 0.67, 0.42)$, and while these parameters have the same ranks and the $\hat{\boldsymbol{\beta}}$'s above, it is not clear what these parameters represent. The right panel of Fig. 4 shows the fitted tilted Dirichlet angular measure which, while exhibiting similar behavior to the pairwise beta model, has some asymmetries in the dependence between PM10 and NO which do not appear to be reflected in the data.

An aim of a multivariate extreme model is to appropriately estimate probabilities of jointly large events. Given a $p$-dimensional angular measure model $h(\boldsymbol{w})$,

$$\begin{aligned}\mathbb{P}(T_1(Y_1) > z_1, \dots, T_p(Y_p) > z_p) &\approx \int_{S_{p-1}} \int_{r=\max_{i=1,\dots,p} w_i z_i^{-1}} r^{-2} h(\boldsymbol{w}) \mathrm{d}r \mathrm{d}\boldsymbol{w} \\ &= \int_{S_{p-1}} \min_{i=1,\dots,p} w_i z_i^{-1} h(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}.\end{aligned}$$

For known angular measure models such as the tilted Dirichlet and the pairwise beta, this integral must be done numerically or with Monte Carlo methods. A straightforward approach to estimating this integral is to use importance sampling.

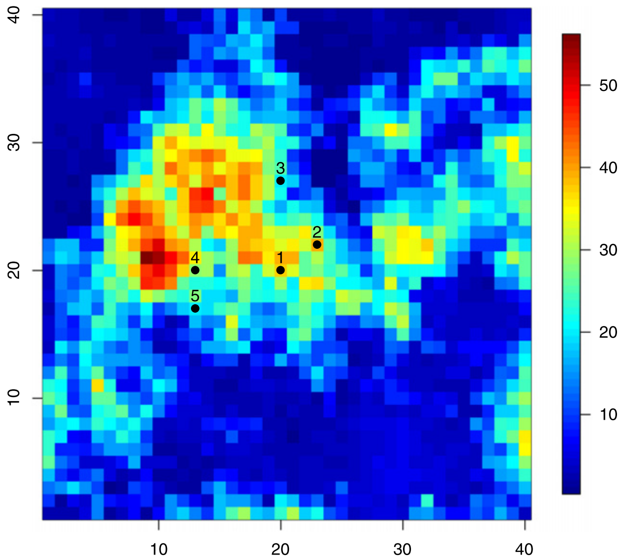


**Fig. 5.** Shows a realization of the Schlather method for simulating max-stable random fields along with the five locations, labeled 1–5, for which we want to construct a multivariate extreme model.

**Table 2**
Summary of the estimates of the three events that appear in Section 5.1. In all three cases the probability of the event as calculated from the fitted pairwise beta model falls well within the 95% confidence interval associated with the empirical estimates.

| | Event 1<br>PM10 > 100, NO > 400 | Event 2<br>PM10 > 85, NO > 230, SO2 > 120 | Event 3<br>PM10 > 185, NO > 600 |
|---|---|---|---|
| $n$ | 543 | 528 | 543 |
| Exceedances | 32 | 16 | 2 |
| Emp. Pt. Est. | 0.0589 | 0.0303 | 0.0037 |
| 95% CI | (0.0391, 0.0787) | (0.0157, 0.0449) | (0.0004, 0.0132) |
| Model Pt. Est. | 0.0465 | 0.0411 | 0.0064 |

We assess the pairwise beta's ability to model days which have large amounts of air pollution in more than one component. We look at three different events: (1) PM10 > 100, NO > 400; (2) PM10 > 85, NO > 230, SO2 > 120; and (3) PM10 > 185, NO > 600. The first and third events answer questions about the joint behavior of PM10 and NO, and we have seen evidence that these components are frequently large together. In the first event, both PM10 and NO are relatively large: a measurement of 100 corresponds to roughly the 0.86 empirical quantile of PM10 and 400 corresponds to roughly the 0.94 quantile of NO. The third event corresponds to a very large event in both components: the marginal quantiles are roughly 0.99 and 0.98 for PM10 and NO respectively. The second event answers a question that might be posed about the joint behavior of all three components, namely what is the probability that SO2 is large (120 roughly corresponds to the 0.90 quantile) when both PM10 and NO are relatively large (both at roughly their 0.80 quantiles)?

Table 2 summarizes the results for both the empirical estimates of these three events as well as the probabilities of these three events as estimated by the fitted pairwise beta model. The 95% confidence intervals are generated by the normal approximation to the binomial for the first two events and the exact binomial confidence interval is given for the third event because its estimate was near 0. The model point estimates were calculated via importance sampling with 10 000 points from a uniform density on $S_2$. The uncertainty of the importance sampling estimates (not including uncertainty due to the earlier parameter estimation) is not reported in the table to avoid confusion, but it was much smaller than that of the empirical estimates. In all three cases, the point estimate given by the model is well inside the 95% confidence intervals indicating that the pairwise beta model has sufficient flexibility to model these three large joint events.

Although it is possible to fit the pairwise beta to the five-dimensional air quality data, since the model assumes asymptotic dependence, it is probably inappropriate to include the O3 data as Heffernan and Tawn [18] found that O3 was negatively correlated with all the other winter pollutants. Rather, to illustrate that the pairwise beta can be fit to data of higher dimension, we instead model simulated spatial data.

### 5.2. Simulated spatial fields

Our exploration into multivariate extreme value models was originally motivated by spatial problems. Dependence in spatial data is usually modeled to decrease as distance between the observations increases, thus it is natural to describe the dependence between pairs of observations. Fig. 5 shows a random field generated by the process developed by Schlather [32].

**Table 3**
The log-likelihood ($\ell(\boldsymbol{\theta})$), the number of parameters ($k$), and AIC values for the pairwise beta and Dirichlet models as fit to the five locations of simulated spatial fields. Results are for 750, 500, and 250 exceedances which correspond to the empirical 85%, 90%, and 95% quantiles of the radial components for the five locations of the 5000 simulated fields. The AIC shows that the pairwise beta model outperforms the Dirichlet model for this spatial data.

| | Pairwise Beta | | | Dirichlet | | |
|---|---|---|---|---|---|---|
| | $\ell(\boldsymbol{\theta})$ | $k$ | AIC | $\ell(\boldsymbol{\theta})$ | $k$ | AIC |
| $N_{t_0} = 750$ | 2127.2904 | 11 | −4232.581 | 1829.3867 | 5 | −3648.773 |
| $N_{t_0} = 500$ | 1425.2265 | 11 | −2828.453 | 1200.0731 | 5 | −2390.146 |
| $N_{t_0} = 250$ | 726.2325 | 11 | −1430.465 | 602.8029 | 5 | −1195.606 |

**Table 4**
Gives the maximum likelihood parameter estimates for the pairwise beta model as fit to the five observed locations from the simulated fields. Standard errors are in parentheses. To more clearly show that larger values of $\beta_{i,j}$ correspond to pairs of points with shorter distances, the $\beta_{i,j}$ parameters have been listed in increasing order of distance between locations.

| Parameter | $\alpha$ | $\beta_{4,5}$ | $\beta_{1,2}$ | $\beta_{2,3}$ | $\beta_{1,3}$ | $\beta_{1,4}$ | $\beta_{1,5}$ | $\beta_{3,4}$ | $\beta_{2,4}$ | $\beta_{2,5}$ | $\beta_{3,5}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Distance | – | 3.00 | 3.61 | 5.83 | 7.00 | 7.00 | 7.62 | 9.90 | 10.20 | 11.18 | 12.21 |
| $N_{t_0} = 750$ | 1.15 | 88.58 | 40.45 | 35.88 | 26.87 | 31.43 | 18.53 | 0.54 | 0.55 | 0.47 | 0.41 |
| | 0.00 | 0.83 | 0.35 | 0.45 | 0.49 | 0.47 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 |
| $N_{t_0} = 500$ | 0.94 | 68.14 | 32.85 | 30.83 | 21.38 | 51.30 | 20.77 | 0.49 | 0.47 | 0.40 | 0.36 |
| | 0.00 | 1.02 | 0.44 | 0.62 | 0.51 | 1.72 | 0.37 | 0.01 | 0.01 | 0.00 | 0.00 |
| $N_{t_0} = 250$ | 0.84 | 81.20 | 34.87 | 24.33 | 40.07 | 16.89 | 17.48 | 0.39 | 0.38 | 0.32 | 0.29 |
| | 0.01 | 2.47 | 0.98 | 1.06 | 2.98 | 0.65 | 0.71 | 0.01 | 0.01 | 0.01 | 0.01 |

The field is max-stable with unit Fréchet marginal distributions and all finite-dimensional distributions are regular varying. The field is created by multiplying Gaussian fields by realizations from a point processes with specific intensities to yield the desired marginal distribution, and then taking the pointwise maxima of these fields. The bivariate distribution of the field is given by

$$\mathbb{P}(Z(x) \le y_1, Z(x+h) \le y_2) = \exp\left[-\frac{1}{2}\left(\frac{1}{y_1} + \frac{1}{y_2}\right)\left(1 + \sqrt{1 - 2(\rho(h)+1)\frac{y_1 y_2}{(y_1+y_2)^2}}\right)\right],$$

where $\rho(h)$ gives the spatial covariance function of the Gaussian field. The fields we simulate have $\rho(h) = \exp(-h/20)$. A closed-form expression for the multivariate joint distribution for $p > 2$ is not known. These fields are known to be asymptotically dependent for all distances $h$.

We wish to model the joint distribution of the five locations marked in Fig. 5. Five-thousand fields are simulated and, using the sum of the observations at the five observed locations, the largest 750, 500, and 250 are selected; that is, we set the threshold at the 0.85, 0.90, and 0.95 empirical quantiles of the norm of the observations at the five locations. We fit the tilted Dirichlet and pairwise beta models to these realizations as before. With $p = 5$, the pairwise beta model has 11 parameters and the tilted Dirichlet model has only five. Table 3 gives the log-likelihood and AIC values of the two models for each of the three thresholds. In all cases the pairwise beta has a lower AIC value indicating that even when penalized for its additional complexity, the pairwise beta outperforms the tilted Dirichlet. Table 4 gives the parameter estimates for the pairwise beta model and it is clear that the parameters $\beta_{i,j}$ which correspond to points closer together have larger estimated values. The estimated parameters for the tilted Dirichlet model were $\hat{\boldsymbol{\alpha}} = (1.52, 1.28, 1.04, 1.27, 1.08)$ but again these estimates yield no interpretation as to the relative dependence between the pairs of locations.

As with the air pollution data, we assess the fit by comparing the estimated probabilities of rare joint events from the fitted pairwise beta model to the actual probabilities from Schlather's model. We again test three events. The first event is $Z(x_1) > 100, Z(x_2) > 100$, corresponding to a large event at two locations in close proximity. The probability of this event according to Schlather's model is 0.0071 and the estimated probability according to the fitted pairwise beta model is 0.0058. The second event we test is $Z(x_3) > 100, Z(x_5) > 100$, corresponding to a large event at the two locations farthest apart. The exact and estimated probabilities are 0.0056 and 0.0050. The third event we test is $Z(x_i) > 100$, for $i = 1, \ldots, 5$. The probability of this event cannot be calculated directly for Schlather's model, however an empirical estimate based on 50000 simulated fields was 0.0029. The estimated probability from the fitted pairwise beta model was 0.0022.

## 6. Discussion

A current challenge of modeling multivariate extremes is finding an adequate model for either the exponent measure function, or alternatively, the angular measure. In this work we have introduced a new parametric model for angular measure, the pairwise beta model. An advantage of this model is that it is largely specified by parameters that relate to the amount of dependence between pairs of components in the random vector, which allows for easier interpretation of the parameter estimates. The constructive approach of this model is novel. In both the air quality and spatial examples, the pairwise beta model proved useful and the parameter estimates agreed with the known relative dependence between the components of the data. In both applications, estimated probabilities of jointly large events compared well to either the empirical or true probabilities of these events.

Although the pairwise beta model seems to be more flexible than the tilted Dirichlet model of Coles and Tawn [6], its additive nature does not give it the complete flexibility we had initially hoped to achieve. The global $\alpha$ parameter sets an overall level of dependence and the $\beta_{i,j}$ parameters allow the model to increase or decrease the pairwise dependence level. However, the model's additive nature means that these $\beta_{i,j}$ parameters do not work independently: adjusting a single $\beta_{i,j}$ parameter affects the level of dependence among all pairs of components in the random vector. Consequently, the model is unable to achieve complete independence between a pair of components if other $\beta_{i,j}$ parameters are non-zero. This implies that the pairwise beta, like the tilted Dirichlet, should be used when a level of dependence is assumed between all components.

Another desirable property of a multivariate model is for it to be closed over dimensions. A parametric family of $p$-dimensional multivariate extreme value distributions is closed if its exponent measure function $V_p(\boldsymbol{z}, \boldsymbol{\theta})$ is closed. Only some of the existing parametric models for an exponent measure function exhibit this property. The negative logistic model [20] can be shown to be closed, whereas the asymmetric logistic model [37] is not closed.

There are two ways one can think of closure for angular density models. Since angular density models are probability distributions on the simplex $S_{p-1}$, one can ascertain if these probability densities are closed on $S_{p-1}$. That is, letting $h_p^*(\boldsymbol{w}; \boldsymbol{\alpha}_p)$ be a density on $S_{p-1}$, then it is closed if $\int_0^{1-\sum_{i=1,\ldots,p-2} w_i} h_p^*(\boldsymbol{w}; \boldsymbol{\alpha}_p)\mathrm{d}w_{p-1} = h_{p-1}^*(\boldsymbol{w}; \boldsymbol{\alpha}_{p-1})$. It is well known that the Dirichlet density exhibits this type of closure; however without a restriction of being completely symmetric, the Dirichlet density does not meet the moment conditions (5) required of an angular density model. With regards to valid angular density models, neither the tilted Dirichlet of [6] nor the pairwise beta model are closed on $S_{p-1}$.

In terms of describing extremes, closure on $S_{p-1}$ is probably not the property of interest, as it does not correspond to closure of a multivariate extreme value distribution. If an angular density model $h_p(\boldsymbol{w}; \boldsymbol{\theta})$ corresponds to a closed multivariate extreme value distribution, then

$$\lim_{\zeta\to\infty} V_p(z_1, \ldots, z_{p-1}, \zeta; \boldsymbol{\theta}) = p \int_{S_{p-1}} \max_{i=1,\ldots,p-1} \left(\frac{w_i}{z_i}\right) h_p(\boldsymbol{w}; \boldsymbol{\theta})\mathrm{d}(\boldsymbol{w}) = V_{p-1}(z_1, \ldots, z_{p-1}, ; \boldsymbol{\theta}).$$

The integrand in the middle expression above is intractable for both the tilted Dirichlet and pairwise beta models. Therefore, like other extremes models, the pairwise beta is dimension-specific, which is indeed a shortcoming, especially if one wishes to model data with missing observations.

Finally, while the pairwise beta model could theoretically be fit to data of any finite dimension, it is clear that the model is most useful for problems of modest dimension.

Obviously the pairwise beta model does not answer all questions about modeling multivariate extremes. Nevertheless, the model should prove quite useful to people who need to model the joint extremal behavior of a random vector in a parsimonious manner. Angular measure models such as the pairwise beta are well suited for fitting threshold exceedance data, and thus practitioners do not have to construct and work with vectors of componentwise block maxima which are difficult to understand and explain. The model can be fit via the straightforward approach in Section 4. The model has all its mass on the interior of the simplex and thus does not have the discontinuities seen in models for the exponent measure function. As demonstrated, the pairwise beta model achieves a level of flexibility and an interpretability not before seen in parametric angular measure models. Perhaps most importantly, the pairwise beta model represents a step in a new direction of creating more useful models for multivariate extremes.

## Acknowledgments

The authors would like to thank two anonymous referees whose thoughtful comments greatly improved the manuscript. Dan Cooley's work has been supported in part by a National Science Foundation grant DMS 0905315, by the PRIMES program at Colorado State University (NSF-IGERT grant DGE-0221595), and by the Geophysical Statistics Project at the National Center for Atmospheric Research (NSF DMS-03055474). Richard Davis' work was supported in part by NSF grant DMS-0743459. Philippe Naveau's work was supported in part by the FP7-ACQWA, GIS-PEPER, ANR-MOPERA and ANR-AssimilEx projects.

## Appendix

**Theorem 1.** *The pairwise beta model, $h(\boldsymbol{w}; \alpha, \boldsymbol{\beta})$ as defined (12), is a valid angular measure.*

It suffices to show that $\int_{S_{p-1}} w_k h(\boldsymbol{w}; \alpha, \boldsymbol{\beta})\mathrm{d}\boldsymbol{w} = 1/p$ for all $k = 1, \ldots, p$. Without loss of generality, we only consider the case $k = 1$. From (12) we have

$$\begin{aligned} &\int_{S_{p-1}} w_1 h(\boldsymbol{w}; \alpha, \boldsymbol{\beta})\mathrm{d}\boldsymbol{w} \qquad (13) \\ &\quad = \int_{S_{p-1}} w_1 \qquad K_p(\alpha) \sum_{1\le i<j\le p} h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j})\mathrm{d}\boldsymbol{w} \end{aligned}$$

$$= K_p(\alpha) \sum_{1\le i<j\le p} \int_{S_{p-1}} w_1 h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}) \mathrm{d}\boldsymbol{w}$$

$$= K_p(\alpha) \left[ \sum_{j=2}^{p} \int_{S_{p-1}} w_1 h_{1,j}(\boldsymbol{w}; \alpha, \beta_{1,j}) \mathrm{d}\boldsymbol{w} + \sum_{2\le i<j\le p} \int_{S_{p-1}} w_1 h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}) \mathrm{d}\boldsymbol{w} \right]$$

$$= K_p(\alpha) \left[ \sum_{j=2}^{p} I_{1,1,j} + \sum_{2\le i<j\le p} I_{1,i,j} \right], \tag{14}$$

where $I_{1,1,j} = \int_{S_{p-1}} w_1 h_{1,j}(w_1, w_j; \alpha, \beta_{1,j}) \mathrm{d}\boldsymbol{w}$ and $I_{1,i,j} = \int_{S_{p-1}} w_1 h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}) \mathrm{d}\boldsymbol{w}$. We first examine $I_{1,1,j}$ and consider the case $j = 2$. Suppressing the dependence of $h_{1,2}$ on $\alpha$ and $\beta_{1,2}$, we have

$$\begin{aligned}
I_{1,1,2} &= \sqrt{p} \int_{w_1=0}^{1} \int_{w_2=0}^{1-w_1} \int_{w_3=0}^{1-(w_1+w_2)} \cdots \int_{w_{p-1}=0}^{1-(w_1+w_2+\cdots w_{p-2})} w_1 h_{1,2}(\boldsymbol{w}) \mathrm{d}w_{p-1} \ldots \mathrm{d}w_2 \mathrm{d}w_1 \\
&= \sqrt{p} \int_{w_1=0}^{1} \int_{w_2=0}^{1-w_1} w_1 h_{1,2}(\boldsymbol{w}) \int_{w_3=0}^{1-(w_1+w_2)} \cdots \int_{w_{p-1}=0}^{1-(w_1+w_2+\cdots w_{p-2})} \mathrm{d}w_{p-1} \ldots \mathrm{d}w_2 \mathrm{d}w_1 \\
&= \sqrt{p} \int_{w_1=0}^{1} \int_{w_2=0}^{1-w_1} w_1 (w_1+w_2)^{2\boldsymbol{\alpha}-1} (1-(w_1+w_2))^{\boldsymbol{\alpha}(p-2)-p+2} \\
&\quad \times \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})} \left(\frac{w_1}{w_1+w_2}\right)^{\beta_{1,2}-1} \left(\frac{w_2}{w_1+w_2}\right)^{\beta_{1,2}-1} \frac{(1-(w_1+w_2))^{p-3}}{(p-3)!} \mathrm{d}w_2 \mathrm{d}w_1.
\end{aligned}$$

After the change of variables $\theta = w_1/(w_1+w_2)$ and $r = w_1 + w_2$, which has Jacobian $|J| = \begin{vmatrix} \partial w_1/\partial\theta & \partial w_1/\partial r \\ \partial w_2/\partial\theta & \partial w_2/\partial r \end{vmatrix} = r$, we obtain

$$\begin{aligned}
I_{1,1,2} &= \sqrt{p} \int_{\theta=0}^{1} \int_{r=0}^{1} r\theta r^{2\boldsymbol{\alpha}-1} (1-r)^{\boldsymbol{\alpha}(p-2)-p+2} \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})} \theta^{\beta_{1,2}-1} (1-\theta)^{\beta_{1,2}-1} \frac{(1-r)^{p-3}}{(p-3)!} r \mathrm{d}r \mathrm{d}\theta \\
&= \frac{\sqrt{p}}{(p-3)!} \int_{\theta=0}^{1} \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})} \theta^{\beta_{1,2}} (1-\theta)^{\beta_{1,2}-1} \mathrm{d}\theta \int_{r=0}^{1} r^{2\boldsymbol{\alpha}+1} (1-r)^{\boldsymbol{\alpha}(p-2)-1} \mathrm{d}r \\
&= \frac{\sqrt{p}}{(p-3)!} \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})} \frac{\Gamma(\beta_{1,2})\Gamma(\beta_{1,2}+1)}{\Gamma(2\beta_{1,2}+1)} \frac{\Gamma(2\alpha+2)\Gamma(\alpha(p-2))}{\Gamma(\alpha p+2)} \\
&= \frac{\sqrt{p}}{2(p-3)!} \frac{\Gamma(2\alpha+2)\Gamma(\alpha(p-2))}{\Gamma(\alpha p+2)}.
\end{aligned} \tag{15}$$

Similarly, for the case where $i = 2$ and $j = 3$,

$$\begin{aligned}
I_{1,2,3} &= \sqrt{p} \int_{w_2=0}^{1} \int_{w_3=0}^{1-w_2} \int_{w_4=0}^{1-(w_2+w_3)} \cdots \int_{w_{p-1}=0}^{1-(w_2+w_3+\cdots w_{p-2})} \int_{w_1=0}^{1-(w_2+w_3+\cdots w_{p-1})} w_1 h_{2,3}(\boldsymbol{w}) \mathrm{d}w_1 \mathrm{d}w_{p-1} \ldots \mathrm{d}w_3 \mathrm{d}w_2 \\
&= \sqrt{p} \int_{w_2=0}^{1} \int_{w_3=0}^{1-w_2} h_{2,3}(\boldsymbol{w}) \int_{w_4=0}^{1-(w_2+w_3)} \cdots \int_{w_{p-1}=0}^{1-(w_2+w_3+\cdots w_{p-2})} \int_{w_1=0}^{1-(w_2+w_3+\cdots w_{p-1})} w_1 \mathrm{d}w_1 \mathrm{d}w_{p-1} \ldots \mathrm{d}w_3 \mathrm{d}w_2 \\
&= \sqrt{p} \int_{w_2=0}^{1} \int_{w_3=0}^{1-w_2} (w_2+w_3)^{2\boldsymbol{\alpha}-1} (1-(w_2+w_3))^{\boldsymbol{\alpha}(p-2)-p+2} \frac{\Gamma(2\beta_{2,3})}{\Gamma^2(\beta_{2,3})} \left(\frac{w_2}{w_2+w_3}\right)^{\beta_{2,3}-1} \left(\frac{w_3}{w_2+w_3}\right)^{\beta_{2,3}-1} \\
&\quad \times \frac{(1-(w_2+w_3))^{p-2}}{(p-2)!} \mathrm{d}w_3 \mathrm{d}w_2 \\
&= \frac{\sqrt{p}}{(p-2)!} \int_{\theta=0}^{1} \frac{\Gamma(2\beta_{2,3})}{\Gamma^2(\beta_{2,3})} \theta^{\beta_{2,3}-1} (1-\theta)^{\beta_{2,3}-1} \mathrm{d}\theta \int_{r=0}^{1} r^{2\boldsymbol{\alpha}} (1-r)^{\boldsymbol{\alpha}(p-2)} \mathrm{d}r \\
&= \frac{\sqrt{p}}{(p-2)!} \frac{\Gamma(2\alpha+1)\Gamma(\alpha(p-2)+1)}{\Gamma(\alpha p+2)}.
\end{aligned} \tag{16}$$

Putting (15) and (16) into (14) and substituting for $K_p(\alpha)$, we obtain

$$\begin{aligned}
\int_{S_{p-1}} w_1 h(\boldsymbol{w}; \alpha, \boldsymbol{\beta}) \mathrm{d}\boldsymbol{w} &= \frac{2(p-3)!}{p(p-1)\sqrt{p}} \frac{\Gamma(\alpha p+1)}{\Gamma(2\alpha+1)\Gamma(\alpha(p-2))} \left[ (p-1) \frac{\sqrt{p}}{2(p-3)!} \frac{\Gamma(2\alpha+2)\Gamma(\alpha(p-2))}{\Gamma(\alpha p+2)} \right. \\
&\quad \left. + \left( \binom{p}{2} - (p-1) \right) \frac{\sqrt{p}}{(p-2)!} \frac{\Gamma(2\alpha+1)\Gamma(\alpha(p-2)+1)}{\Gamma(\alpha p+2)} \right]
\end{aligned}$$

$$= \frac{1}{p}\left(\frac{2\alpha+1}{\alpha p+1} + \frac{\alpha(p-2)}{\alpha p+1}\right)$$
$$= 1/p. \tag{17}$$

**Theorem 2.** *Let $\boldsymbol{Z}$ be a max-stable random vector with angular measure described by the pairwise beta model with parameters $(\alpha, \boldsymbol{\beta})$, and let $\boldsymbol{Z}^*$ be another pairwise beta max-stable random vector with parameters $(\alpha, \boldsymbol{\beta}^*)$, where $\beta^*_{i,j} > \beta_{i,j}$ and $\beta^*_{k,l} = \beta_{k,l}$ for $(k, l) \neq (i, j)$. Then $\phi_{i,j}(\boldsymbol{Z}^*) < \phi_{i,j}(\boldsymbol{Z})$, implying that the bivariate dependence between the ith and jth elements is greater for $\boldsymbol{Z}^*$ than for $\boldsymbol{Z}$.*

Without loss of generality, we consider $\phi_{1,2}(\boldsymbol{Z})$.

$$\begin{aligned}
\phi_{1,2}(\boldsymbol{Z}) &= \lim_{z\to\infty} V(1, 1, z, \ldots, z) = p\int_{S_{p-1}} \max(w_1, w_2)h(\boldsymbol{w}; \alpha, \boldsymbol{\beta})\mathrm{d}\boldsymbol{w} \\
&= p\int_{S_{p-1}} \left(\max(w_1, w_2)K_p(\alpha)h_{1,2}(\boldsymbol{w}; \alpha, \beta_{1,2}) + \sum_{\substack{1\le i<j\le p \\ (i,j)\neq(1,2)}} \max(w_1, w_2)K_p(\alpha)h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j})\right)\mathrm{d}\boldsymbol{w} \\
&= p\int_{S_{p-1}} \max(w_1, w_2)K_p(\alpha)h_{1,2}(\boldsymbol{w}; \alpha, \beta_{1,2})\mathrm{d}\boldsymbol{w} + C\left(\alpha, \boldsymbol{\beta}_{-(1,2)}\right),
\end{aligned}$$

where $\boldsymbol{\beta}_{-(1,2)}$ denotes the vector $\boldsymbol{\beta}$ with the $(1, 2)$ element removed and $C\left(\alpha, \boldsymbol{\beta}_{-(1,2)}\right)$ is constant by the assumptions. Following in a similar manner to Theorem 1, one obtains

$$\begin{aligned}
\phi_{1,2}(\boldsymbol{Z}) &= \frac{2}{p-1}\int_{w_1=0}^{1}\int_{w_2=0}^{1-w_1} \max(w_1, w_2) \\
&\quad\times \frac{\Gamma(\alpha p+1)}{\Gamma(2\alpha+1)\Gamma(\alpha(p-2))}(w_1+w_2)^{2\alpha-1}(1-(w_1+w_2))^{\alpha(p-2)-1} \\
&\quad\times \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\left(\frac{w_1}{w_1+w_2}\right)^{\beta_{1,2}-1}\left(\frac{w_2}{w_1+w_2}\right)^{\beta_{1,2}-1}\mathrm{d}w_2\mathrm{d}w_1 + C\left(\alpha, \boldsymbol{\beta}_{-(1,2)}\right) \\
&= \frac{2}{p-1}\int_{\theta=0}^{1}\int_{r=0}^{1} r\max(\theta, 1-\theta)\frac{\Gamma(\alpha p+1)}{\Gamma(2\alpha+1)\Gamma(\alpha(p-2))}r^{2\boldsymbol{\alpha}-1}(1-r)^{\boldsymbol{\alpha}(p-2)-1} \\
&\quad\times \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\theta^{\beta_{1,2}-1}(1-\theta)^{\beta_{1,2}-1}r\mathrm{d}r\mathrm{d}\theta + C\left(\alpha, \boldsymbol{\beta}_{-(1,2)}\right) \\
&= \frac{2}{p-1}\frac{2\alpha+1}{\alpha p+1}\left\{\int_{\theta=0}^{1/2}(1-\theta)\frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\theta^{\beta_{1,2}-1}(1-\theta)^{\beta_{1,2}-1}\mathrm{d}\theta + \int_{\theta=0}^{1/2}\theta\frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\theta^{\beta_{1,2}-1}(1-\theta)^{\beta_{1,2}-1}\mathrm{d}\theta\right\} \\
&= \frac{4}{p-1}\frac{2\alpha+1}{\alpha p+1}\frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\int_{\theta=0}^{1/2}(1-\theta)\theta^{\beta_{1,2}-1}(1-\theta)^{\beta_{1,2}-1}\mathrm{d}\theta.
\end{aligned} \tag{18}$$

One recognizes the above integral as an incomplete beta function. Unfortunately, the extra $(1-\theta)$ makes the integrand asymmetric and the integral cannot be evaluated directly.

To proceed, let $h_{1,2}(\theta, \beta_{1,2}) = \frac{\Gamma(2\beta_{1,2})}{\Gamma^2(\beta_{1,2})}\theta^{\beta_{1,2}-1}(1-\theta)^{\beta_{1,2}-1}$; that is, the symmetric beta density function with parameter $\beta_{1,2}$. By symmetry,

$$\int_0^{1/2} h_{1,2}(\theta, \beta_{1,2})\mathrm{d}\theta = 1/2\int_0^1 h_{1,2}(\theta, \beta_{1,2})\mathrm{d}\theta = 1/2 \tag{19}$$

for all $\beta_{1,2}$. Let $\beta^*_{1,2} > \beta_{1,2}$. Consider the continuous function $h_{1,2}(\theta, \beta^*_{1,2}) - h_{1,2}(\theta, \beta_{1,2})$ on $\theta \in (0, 1/2)$. It is straightforward to show (see Lemma 1) that there exists a single $\theta_* \in (0, 1/2)$ such that:

(1) $h_{1,2}(\theta_*, \beta^*_{1,2}) - h_{1,2}(\theta_*, \beta_{1,2}) = 0$,
(2) $h_{1,2}(\theta, \beta^*_{1,2}) - h_{1,2}(\theta, \beta_{1,2}) < 0$ for all $\theta \in (0, \theta_*)$,
(3) $h_{1,2}(\theta, \beta^*_{1,2}) - h_{1,2}(\theta, \beta_{1,2}) > 0$ for all $\theta \in (\theta_*, 1/2)$.

From (19),

$$\int_0^{1/2}\left(h_{1,2}(\theta, \beta^*_{1,2}) - h_{1,2}(\theta, \beta_{1,2})\right)\mathrm{d}\theta = 0 \tag{20}$$

$$\int_{\theta_*}^{1/2} \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta = -\int_0^{\theta_*} \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta$$

$$\int_{\theta_*}^{1/2} (1-\theta_*) \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta = -\int_0^{\theta_*} (1-\theta_*) \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta$$

$$\int_{\theta_*}^{1/2} (1-\theta) \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta < -\int_0^{\theta_*} (1-\theta) \left(h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2})\right) \mathrm{d}\theta$$

$$\int_0^{1/2} (1-\theta) h_{1,2}(\theta, \beta_{1,2}^*) \mathrm{d}\theta < \int_0^{1/2} (1-\theta) h_{1,2}(\theta, \beta_{1,2}) \mathrm{d}\theta. \tag{21}$$

Putting together (18) and (21) yields the result.

**Lemma 1.** *Given $\beta_{1,2}^* > \beta_{1,2}$, there exists a single $\theta_*$ such that:*

(1) $h_{1,2}(\theta_*, \beta_{1,2}^*) - h_{1,2}(\theta_*, \beta_{1,2}) = 0$,
(2) $h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2}) < 0$ *for all* $\theta \in (0, \theta_*)$, *and*
(3) $h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2}) > 0$ *for all* $\theta \in (\theta_*, 1/2)$.

From (20) and by the mean value theorem, there exists a $\theta \in (0, 1/2)$ such that $h_{1,2}(\theta, \beta_{1,2}^*) - h_{1,2}(\theta, \beta_{1,2}) = 0$ or equivalently, $f(\theta) = 1$ where $f(\theta) = h_{1,2}(\theta, \beta_{1,2}^*)/h_{1,2}(\theta, \beta_{1,2})$. Since $f'(\theta) > 0$ for $\theta \in (0, 1/2)$ there exists a unique solution to $f(\theta) = 1$ which we denote by $\theta_*$.

It follows by $f$ being a strictly increasing function that $f(\theta) < f(\theta_*) = 1$ if $\theta \in (0, \theta_*)$ and $f(\theta) > f(\theta_*) = 1$ if $\theta \in (\theta_*, 1/2)$ which proves (2) and (3).

## References


[1] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, D.D. Waal, C. Ferro, Statistics of Extremes: Theory and Applications, Wiley, New York, 2004.
[2] M.-O. Boldi, A.C. Davison, A mixture model for multivariate extremes, Journal of the Royal Statistical Society, Series B 69 (2007) 217–229.
[3] P. Capéraà, A.-L. Fougéres, Estimation of a bivariate extreme value distribution, Extremes 3 (2000) 311–329.
[4] S. Coles, Regional modelling of extreme storms via max-stable processes, Journal of the Royal Statistical Society, Series B 1993 (1993) 797–816.
[5] S. Coles, J. Heffernan, J. Tawn, Dependence measures for extreme value analysis, Extremes 2 (1999) 339–365.
[6] S. Coles, J. Tawn, Modeling multivariate extreme events, Journal of the Royal Statistical Society, Series B 53 (1991) 377–392.
[7] S. Coles, J. Tawn, Statistical methods for multivariate extemes: an application to structural design, Applied Statistics 43 (1994) 1–48.
[8] D. Cooley, P. Naveau, P. Poncet, Variograms for spatial max-stable random fields, in: P. Bertail, P. Doukhan, P. Soulier (Eds.), Dependence in Probability and Statistics, in: Springer Lecture Notes in Statistics, Springer, New York, 2006.
[9] R. Davis, S. Resnick, Prediction of stationary max-stable processes, Annals of Applied Probability 3 (1993) 497–525.
[10] L. de Haan, A. Ferreira, Extreme Value Theory, in: Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
[11] L. de Haan, T.T. Pereira, Spatial extremes: models for the stationary case, The Annals of Statistics 34 (2006) 146–168.
[12] P. Deheuvels, On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions, Statistics and Probability Letters 12 (1991) 429–439.
[13] G. Draisma, H. Drees, A. Ferreira, L. deHaan, Bivariate tail estimation: dependence in asymptotic independence, Bernoulli 10 (2004) 251–280.
[14] J. Einmahl, L. de Haan, V. Piterbarg, Nonparametric estimation of the spectral measure of an extreme value distribution, Annals of Statistics 29 (2001) 1401–1423.
[15] J.H. Einmahl, J. Segers, Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution. Technical Report 2008-42, Tilburg University, 2008.
[16] E. Gumbel, Distributions des valeurs extrêmes en plusieurs dimensions, Publ. Inst. Statis. Univ. Paris 9 (1960) 171–173.
[17] P. Hall, N. Tajvidi, Distribution and dependence function estimation for bivariate extreme-value distributions, Bernoulli 6 (2000) 835–844.
[18] J.E. Heffernan, J.A. Tawn, A conditional approach for multivariate extreme values, Journal of the Royal Statistical Society, Series B 66 (2004) 497–546.
[19] J. Hüsler, R. Reiss, Maxima of normal random vectors: between independence and complete dependence, Statistics and Probability Letters 7 (1989) 283–286.
[20] H. Joe, Families of min-stable multivariate exponential and multivariate extreme value distributions, Statistics and Probability Letters 9 (1990) 75–81.
[21] H. Joe, R. Smith, I. Weissmann, Bivariate threshold methods for extremes, Journal of the Royal Statisitcal Society, Series B 54 (1992) 171–183.
[22] S. Kotz, S. Nadarajah, Extreme Value Distributions, Imperial College Press, London, 2000.
[23] A.W. Ledford, J.A. Tawn, Diagnostics for dependence within time series extremes, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 65 (2) (2003) 521–543.
[24] A. Ledford, J. Tawn, Modelling dependence within joint tail regions, Journal of the Royal Statistical Society, Series B B (1997) 475–499.
[25] A.W. Ledford, J.A. Tawn, Statistics for near independence in multivariate extreme values, Biometrika 83 (1996) 169–187.
[26] S. Nadarajah, A polynomial model for bivariate extreme value distributions, Probability Letters 42 (1999) 15–25.
[27] P. Naveau, A. Guillou, D. Cooley, J. Diebolt, Modeling pairwise dependence of maxima in space, Biometrika 96 (2009) 1–17.
[28] L. Peng, Estimation of the coefficient of tail dependence in bivariate extremes, Statistics and Probability Letters 43 (1999) 399–409.
[29] A. Ramos, A. Ledford, A new class of models for bivariate joint tails, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71 (1) (2009) 219–241.
[30] S. Resnick, Extreme Values, Regular Variation, and Point Processes, Springer-Verlag, New York, 1987.
[31] S. Resnick, Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, in: Springer Series in Operations Research and Financial Engineering, Springer, New York, 2007.
[32] M. Schlather, Models for stationary max-stable random fields, Extremes 5 (1) (2002) 33–44.
[33] M. Schlather, J. Tawn, Inequalities for the extremal coefficients of multivariate extreme value distributions, Extremes 5 (1) (2002) 87–102.
[34] M. Schlather, J. Tawn, A dependence measure for multivariate and spatial extreme values: Properties and inference, Biometrika 90 (2003) 139–156.
[35] R. Smith, Max-stable processes and spatial extremes Unpublished manusript (1990).
[36] R.L. Smith, J.A. Tawn, S.G. Coles, Markov chain models for threshold exceedances, Biometrika 84 (1997) 249–268.
[37] J. Tawn, Modeling multivariate extreme value distributions, Biometrika 75 (1990) 245–253.