Modeling Spatial Extremes Using Normal Mean-Variance Mixtures

**Zhongwei Zhang · Raphaël Huser ·
Thomas Opitz · Jennifer Wadsworth**

Abstract Classical models for multivariate or spatial extremes are mainly based upon the asymptotically justified max-stable or generalized Pareto processes. These models are suitable when asymptotic dependence is present, i.e., the joint tail decays at the same rate as the marginal tail. However, recent environmental data applications suggest that asymptotic independence is equally important and, unfortunately, existing spatial models in this setting that are both flexible and can be fitted efficiently are scarce. Here, we propose a new spatial copula model based on the generalized hyperbolic distribution, which is a specific normal mean-variance mixture and is very popular in financial modeling. The tail properties of this distribution have been studied in the literature, but with contradictory results. It turns out that the proofs from the literature contain mistakes. We here give a corrected theoretical description of its tail dependence structure and then exploit the model to analyze a simulated dataset from the inverted Brown–Resnick process, hindcast significant wave height data in the North Sea, and wind gust data in the state of Okla-

Zhongwei Zhang
CEMSE Division, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Saudi Arabia
E-mail: zhongwei.zhang@kaust.edu.sa

Raphaël Huser
CEMSE Division, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Saudi Arabia
E-mail: raphael.huser@kaust.edu.sa

Thomas Opitz
Biostatistics and Spatial Processes, INRAE, 84914 Avignon, France
E-mail: thomas.opitz@inrae.fr

Jennifer Wadsworth
Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster LA1 4YF, United Kingdom
E-mail: j.wadsworth@lancaster.ac.uk

homa, USA. We demonstrate that our proposed model is flexible enough to capture the dependence structure not only in the tail but also in the bulk.

Keywords Asymptotic independence · Copula model · Generalized hyperbolic distribution · Normal mean-variance mixtures · Spatial extremes

Mathematics Subject Classification (2020) 60G70 · 60E07 · 62H20

1 Introduction

The statistical modeling of spatial extremes has gained significant interest in recent decades due to the increasing occurrence and sizes of natural extreme events, such as heat waves, heavy rainfall, and wildfires. When modeling spatial extremes, accurate inference for the marginal distribution at each site and a precise assessment of the dependence structure of extreme events among different sites are both needed; see Davison et al. (2012); Davison and Huser (2015) and Huser and Wadsworth (2020) for an overview. In this article we focus on modeling the spatial dependence of extreme events.

Classical models for spatial extremes are mainly based upon the asymptotically justified max-stable processes (de Haan, 1984; de Haan and Pereira, 2006) or generalized Pareto processes (Rootzén and Tajvidi, 2006; Ferreira and de Haan, 2014). The limiting dependence structures that arise in these models must be either asymptotically dependent (defined in Section 2.2), i.e., the joint tail decays at the same rate as the marginal tail, or for maxima, exactly independent. However, asymptotic independence appears to be equally important, as suggested by recent environmental data applications (Wadsworth and Tawn, 2012; Le et al., 2018). Unfortunately, existing models in this setting that are both flexible and can be fitted efficiently are scarce.

Existing models from the literature that can capture asymptotic independence include the Gaussian copula model (Bortot et al., 2000), inverted max-stable models (Wadsworth and Tawn, 2012), the Huser–Wadsworth model (Huser and Wadsworth, 2019), and the conditional extremes model (Heffernan and Tawn, 2004; Wadsworth and Tawn, 2019). The Gaussian copula is the simplest of these models and has a restrictive dependence structure, i.e., its bivariate distribution can only be asymptotically independent or perfectly dependent, and its tail dependence structure is regulated by a single parameter: the correlation coefficient. Inverted max-stable models present the same difficulties for inference as max-stable models since often only likelihoods based on lower-dimensional densities are available (Padoan et al., 2010; Castruccio et al., 2016); the Huser–Wadsworth model, as well as its extension to model dependence jointly in the lower and upper tails (Gong and Huser, 2021), can capture both asymptotic dependence and independence but its distribution and (potentially censored) density functions rely on unidimensional integrals which have to be computed numerically, and this is computationally prohibitive in high dimensions, especially when computation of the multivariate normal distribution function is required; the conditional extremes model allows the change of asymptotic dependence class with distance between sites but it lacks a simple unconditional interpretation; see Huser and Wadsworth (2020) for a more detailed discussion.

Here we propose a flexible model based on an extension of scale mixtures. A random vector $\boldsymbol{X} \in \mathbb{R}^d$ is called a scale mixture if it has the representation $\boldsymbol{X} = R\boldsymbol{W}$, where R is a non-negative univariate random variable and $\boldsymbol{W} \in \mathbb{R}^d$ is a random vector. Scale mixtures provide a flexible family of distributions that can capture both asymptotic dependence and independence depending

on the specification of the tail of R and $\boldsymbol{W}$; see Huser et al. (2017), Huser and Wadsworth (2019), Engelke et al. (2019). In practice $\boldsymbol{W}$ is often taken as a Gaussian random vector and in this case $\boldsymbol{X}$ is termed a Gaussian scale mixture. Natural alternatives and extensions of Gaussian scale mixtures include the Gaussian location mixtures (Krupskii et al., 2018; Castro Camilo and Huser, 2020), and the skew-t model (Morris et al., 2017), which is a specific Gaussian location-scale mixture. Here we focus on a different form of Gaussian location-scale mixtures, for which the term *normal mean-variance mixtures* has been coined in the literature. Specifically, $\boldsymbol{X}$ is called a normal mean-variance mixture if it can be represented as

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\gamma}R + \sqrt{R}\boldsymbol{W}, \quad R \perp\!\!\!\perp \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is a location parameter vector, $\boldsymbol{\gamma} \in \mathbb{R}^d$ regulates the skewness, $R \sim F_R$ is a non-negative mixing random variable, and $\boldsymbol{W} \sim \mathcal{N}_d(\boldsymbol{0}, \Sigma)$ is a Gaussian random vector with correlation matrix Σ. This provides a very flexible family of distributions and by properly choosing the parameters and mixing distribution, many well-known multivariate distributions can be obtained. Another benefit of normal mean-variance mixtures is the simple implementation of their conditional simulation, using their property of closedness under conditioning (Jamalizadeh and Balakrishnan, 2019), which is in clear contrast to max-stable or inverted max-stable models.

A prominent example of normal mean-variance mixtures is the generalized hyperbolic (GH) distribution, when the mixing distribution F_R is taken as generalized inverse Gaussian. The skew-t model (Morris et al., 2017) can be represented as $\boldsymbol{X} = \boldsymbol{\mu} + \gamma_c \boldsymbol{1}_d \sqrt{R}|Z| + \sqrt{R}\boldsymbol{W}, \gamma_c \in \mathbb{R}$, where $\boldsymbol{1}_d$ is a d-dimensional vector of 1s, R has an inverse Gamma distribution, which is a limiting case of the generalized inverse Gaussian distribution, Z is a standard normal random variable, $\boldsymbol{W}$ remains a Gaussian random vector and $R, Z, \boldsymbol{W}$ are mutually independent. Although the representations of this skew-t model and the GH distribution look similar, a major difference between them is the extra random variable Z in the representation of the skew-t model, and their tail dependence structures are significantly different. That is, the skew-t model is asymptotically dependent, and the GH distribution has more free parameters and is asymptotically independent except in one limiting case; see Section 2.3 for more details.

The GH distribution is very popular in financial modeling (Barndorff-Nielsen, 1997; Prause, 1999) thanks to its flexible univariate distribution and infinite divisibility; more details of this distribution are given in Section 2.1. Due to the popularity of this distribution in finance, its tail properties have been studied in the literature (Nolde, 2014; von Hammerstein, 2016), but surprisingly with contradictory results. In short, Nolde (2014) claims that the GH distribution is asymptotically independent when the Gaussian correlation parameter is smaller than 1, while von Hammerstein (2016) claims that it can be asymptotically independent or asymptotically dependent, depending on the choice of the other parameters. By examining their proofs in detail, we have found that both of them contain subtle mistakes. Here, we point out the

mistakes in their proofs, which lead to the contradiction, and give a corrected description of the tail dependence of the GH distribution.

Based on this result, we propose to use the GH copula for spatial extremes, which has been less investigated compared with its flexible univariate distribution, and to our knowledge not yet exploited in environmental applications. This model provides flexible dependence structures with many subclasses and limiting models, such as the normal inverse Gaussian, hyperbolic, Gaussian and student t copulas. We apply the model using a full likelihood approach (thus, avoiding the computationally prohibitive censoring mechanism) to a simulated dataset from the inverted Brown–Resnick process, the hindcast significant wave height data considered in Wadsworth and Tawn (2012) and Huser and Wadsworth (2019), and wind gust data in the state of Oklahoma, USA, and demonstrate that our proposed model is flexible enough to capture the dependence structure not only in the tail but also in the bulk.

This paper is structured as follows. In Section 2 we introduce the multivariate generalized hyperbolic distribution and measures of tail dependence, and we then review existing results on the tail dependence properties of the GH distribution and present a corrected description. Section 3 introduces the copula-based likelihood inference, followed by two simulation studies, and Section 4 consists of two data applications. Section 5 concludes with a discussion.

2 Modeling

2.1 The Generalized Hyperbolic (GH) Distribution

The GH distribution is a specific normal mean-variance mixture with representation (1) and mixing distribution F_R specified to be generalized inverse Gaussian, denoted as $\mathrm{GIG}(\lambda, \kappa, \psi)$. The $\mathrm{GIG}(\lambda, \kappa, \psi)$ probability density function is

$$f_{\mathrm{GIG}}(x) = \left(\frac{\psi}{\kappa}\right)^{\lambda/2} \frac{x^{\lambda-1}}{2K_\lambda(\sqrt{\kappa\psi})} \exp\left\{-\frac{1}{2}\left(\frac{\kappa}{x} + \psi x\right)\right\}, \quad x > 0,$$

where K_λ is the modified Bessel function of the second kind with index λ, and, for $\boldsymbol{x} \in \mathbb{R}^d$, the probability density function of a d-dimensional GH distribution is

$$f_{\mathrm{GH}}(\boldsymbol{x}) = a_d \cdot \frac{K_{\lambda-d/2}\left[\sqrt{\{\kappa + (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})}\right] e^{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}\boldsymbol{\gamma}}}{\left[\sqrt{\{\kappa + (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})}\right]^{d/2-\lambda}},$$

where

$$a_d = \frac{\psi^{\lambda/2}(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})^{d/2-\lambda}}{(2\pi)^{d/2}|\Sigma|^{1/2}\kappa^{\lambda/2}K_\lambda(\sqrt{\kappa\psi})},$$

$\boldsymbol{\mu} \in \mathbb{R}^d$ is a location parameter vector, $\boldsymbol{\gamma} \in \mathbb{R}^d$ regulates the skewness, $\Sigma \in \mathbb{R}^{d \times d}$ is a positive definite dispersion matrix, and $\lambda, \kappa, \psi \in \mathbb{R}$ control the shape

Table 1 Some special subclasses and limiting cases of the GH distribution

Distribution	Parameter domain			
	λ	κ	ψ	γ
Hyperbolic	$(d+1)/2$	>0	>0	$\in \mathbb{R}^d$
Normal inverse Gaussian (NIG)	$-1/2$	>0	>0	$\in \mathbb{R}^d$
Student-t with df degrees of freedom	$-\mathrm{df}/2 < 0$	$\mathrm{df} > 0$	0	$\mathbf{0}$
Cauchy	$-1/2$	1	0	$\mathbf{0}$

of the mixing GIG distribution. We denote this GH distribution by $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma)$. Both in the GIG and GH distributions, the admissible parameter values are $\lambda < 0, \kappa > 0, \psi \geq 0$ or $\lambda = 0, \kappa > 0, \psi > 0$ or $\lambda > 0, \kappa \geq 0, \psi > 0$. Note that $\psi = 0$ and $\kappa = 0$ should be understood as two limiting cases.

One important property of the GH distribution is that it is closed under marginalization, conditioning and linear transformations. Specifically, if $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$, then $\boldsymbol{Y} \sim \mathrm{GH}_k(\lambda, \kappa, \psi, B\boldsymbol{\gamma}, B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top)$. This distribution is a rich family with many special subclasses and limiting cases; see Table 1 for some examples and the corresponding admissible parameter domains. Note that, for the GH, hyperbolic, and NIG distributions the dispersion matrix Σ needs to be a correlation matrix for identifiability reasons. This distribution was first introduced in Barndorff-Nielsen (1977) to model sand sizes and has become very popular in modeling turbulence and returns of financial assets; see Barndorff-Nielsen (1997) and Prause (1999). It is also worthwhile to mention that the GH distribution is elliptical if and only if $\boldsymbol{\gamma} = \mathbf{0}$; see Corollary 3 in von Hammerstein (2016). More details on its properties can be found in Blaesild and Jensen (1981), Prause (1999) and McNeil et al. (2005).

There are several equivalent parametrizations for the GH distribution. The most common one is to parametrize it as $\mathrm{GH}(\lambda, \alpha, \boldsymbol{\mu}, \Delta, \delta, \boldsymbol{\beta})$, where we let $\Delta = |\Sigma|^{-1/d}\Sigma$, $\boldsymbol{\beta} = \Sigma^{-1}\boldsymbol{\gamma}$, $\delta = \sqrt{|\Sigma|^{1/d}\kappa}$, and $\alpha = \sqrt{|\Sigma|^{-1/d}(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})}$. This parametrization was used in Barndorff-Nielsen (1977), but it does not give the nice property that the important parameters α and δ are invariant under linear transformations. Therefore, we here instead adopt the parametrization $\mathrm{GH}_d(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma)$.

2.2 Measures of Tail Dependence

Coles et al. (1999) proposed the tail dependence coefficient χ, providing it exists, to measure the extremal dependence between random variables X_1 and X_2. It may be defined through the limit

$$\chi_u := \frac{\Pr\{F_1(X_1) > u, F_2(X_2) > u\}}{\Pr\{F_1(X_1) > u\}} \to \chi \text{ as } u \to 1, \qquad (2)$$

where F_1 and F_2 are the marginal distribution functions of X_1 and X_2, respectively, and are assumed to be continuous without loss of generality. The

lower tail dependence coefficient can be obtained via reflection. Throughout this work we focus on the upper tail dependence. The vector $(X_1, X_2)^\top$ is termed asymptotically independent if $\chi = 0$ and asymptotically dependent if $\chi > 0$. . Notice that when $\chi > 0$, (2) implies that the joint survival function $\Pr\{F_1(X_1) > u, F_2(X_2) > u\}$ decays at the same rate as the marginal survival function $\Pr\{F_1(X_1) > u\} = 1 - u$, whereas when $\chi = 0$, the joint survival function decays at a faster rate than its marginal counterpart.

When asymptotic independence is present, one useful quantity to measure the residual dependence at pre-asymptotic levels is the residual tail dependence coefficient η (Ledford and Tawn, 1996), which is defined by the following asymptotic expansion of the joint tail (providing it exists):

$$\Pr\{F_1(X_1) > u, F_2(X_2) > u\} \sim \mathcal{L}\{(1-u)^{-1}\}(1-u)^{1/\eta} \text{ as } u \to 1, \quad (3)$$

where $\mathcal{L}$ is slowly varying at ∞, i.e., $\mathcal{L}(tx)/\mathcal{L}(x) \to 1$ as $x \to \infty$ for all $t > 0$, and $f(x) \sim g(x)$ as $x \to x_0$ means that, for g non-zero in a neighbourhood of x_0, $f(x)/g(x) \to 1$ as $x \to x_0$. The coefficient $\eta \in (0, 1]$ determines the joint tail decay rate. A reformulation of (3) implies that η can be alternatively defined through the limit

$$\eta_u := \frac{\log \Pr\{F_1(X_1) > u\}}{\log \Pr\{F_1(X_1) > u, F_2(X_2) > u\}} \to \eta \text{ as } u \to 1, \quad (4)$$

and this definition allows one to estimate η with η_u for u close to 1.

The relationship between these two above coefficients is $\eta < 1 \implies \chi = 0$ and $\chi > 0 \implies \eta = 1$. The converse of this relationship does not necessarily hold; see Proposition 4 in Manner and Segers (2011), and an example with $\eta = 1, \chi = 0$ in Huser and Wadsworth (2019). We say that $(X_1, X_2)^\top$ is (a) positively associated if $1/2 < \eta \le 1$; (b) near-independent if $\eta = 1/2$; (c) negatively associated if $0 < \eta < 1/2$. The bivariate measures χ and η can be easily extended to dimension $d > 2$ by replacing $\Pr\{F_1(X_1) > u, F_2(X_2) > u\}$ in (2) and (4) with $\Pr\{F_1(X_1) > u, \ldots, F_d(X_d) > u\}$. As a result, the pair of coefficients (χ, η) is able to measure the tail dependence strength across asymptotic dependence and independence classes, and it is often used for assessing model fit.

2.3 Tail Dependence of the GH Distribution

Recall that the GH distribution is elliptical if and only if $\boldsymbol{\gamma} = \mathbf{0}$. Schlueter and Fischer (2012) have proved that the elliptical GH distribution with dispersion matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, 0 \le \rho < 1$, excluding the limiting case $\psi = 0$, is asymptotically independent and has residual tail dependence coefficient $\eta = \sqrt{(1 + \rho)/2}$; see also Huser et al. (2017). Nolde (2014) considered a more general class and claimed that the whole GH distribution family with dispersion matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, |\rho| \neq 1$, has residual tail dependence coefficient $\sqrt{(1 + \rho)/2}$, which implies that the skewness parameter $\boldsymbol{\gamma}$ has no influence on the residual tail

dependence coefficient. However, von Hammerstein (2016) used a different approach and claimed that the GH distribution can be asymptotically dependent or independent depending on the choice of the parameters, although without giving the residual tail dependence coefficient when asymptotic independence is present.

After carefully investigating where this contradiction comes from, we found that both of their proofs have mistakes which lead to incorrect conclusions. We detail their mistakes and a corrected version of von Hammerstein's proof in the Supplementary Material. We now give a corrected description of the tail dependence property of the GH distribution, using a similar approach to Nolde (2014). In contrast with the often complicated calculation of the residual tail dependence coefficient by its definition, Nolde (2014) provided a geometric interpretation of this coefficient, which leads to simple and intuitive computations of this coefficient for a variety of distributions. This is particularly the case when joint densities are easier to compute than joint distribution or survival functions. We first recall some definitions and theorems from Nolde (2014) which will be used in our proof.

Consider a sequence of independent and identically distributed random vectors $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ on $\mathbb{R}^2$. Let $N_n := \{\boldsymbol{Z}_1/r_n, \ldots, \boldsymbol{Z}_n/r_n\}$ denote an n-point sample cloud with scaling constants $r_n > 0$, $r_n \to \infty$ as $n \to \infty$ and let $N_n(A)$ be the number of points of N_n contained in the Borel set $A \in \mathbb{R}^d$, i.e. $N_n(A) = \sum_{i=1}^n 1_A(\boldsymbol{Z}_i/r_n)$. Let D be a compact set in $\mathbb{R}^2$. Then D is called a *limit set* of the sample cloud N_n as $n \to \infty$ if (i) $\Pr\{N_n(U^c) > 0\} \to 0$ for open sets U containing D, and U^c denotes the complement of U, and (ii) $\Pr\{N_n(\boldsymbol{p} + \varepsilon B) > m\} \to 1$ for all $m \geq 1, \varepsilon > 0, \boldsymbol{p} \in D$, where B denotes the Euclidean unit ball. If every ray from the origin intersects the boundary of a given set D in a single point, then D can be characterized by a continuous *gauge function* n_D. That is, $n_D : \mathbb{R}^2 \to [0, \infty)$ is a homogeneous function of degree one, i.e. $n_D(t\boldsymbol{x}) = t n_D(\boldsymbol{x})$ for all $t > 0$, and $D = \{\boldsymbol{x} \in \mathbb{R}^2 : n_D(\boldsymbol{x}) < 1\}$. A set D is called *star-shaped* if $\boldsymbol{x} \in D$ implies $t\boldsymbol{x} \in D$ for all $t \in (0, 1)$.

A measurable function $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ is *regularly varying* at infinity with exponent $a \in \mathbb{R}$ if for $x > 0$, $\lim_{t \to \infty} \Psi(tx)/\Psi(x) = x^a$. Roughly speaking, regularly varying functions behave asymptotically like power functions. When $a = 0$, the function Ψ is slowly varying.

Lemma 1 (Theorem 2.1 in Nolde (2014)) *Let $\{(X_i, Y_i)^\top, i \geq 1\}$ be independent random vectors in $\mathbb{R}^2$ from a distribution G with marginal survival functions $1 - G_i(s) = e^{-\Psi_i(s)}, s > 0$, where Ψ_1 and Ψ_2 are regularly varying at infinity with the same exponent $a > 0$. Suppose there exists a scaling sequence $r_n > 0$ such that the n-point sample cloud*

$$N_n = \{(X_1/r_n, Y_1/r_n)^\top, \ldots, (X_n/r_n, Y_n/r_n)^\top\}$$

converges onto set D whose interior is bounded, open, and star-shaped as $n \to \infty$. Let $\boldsymbol{q} = (q_1, q_2)^\top = \sup D$ denote the coordinate-wise supremum of D. Then D is a star-shaped subset of $(-\infty, \boldsymbol{q}]$. Define

$$r_D = \min\{r \geq 0 : D \cap ((rq_1, \infty) \times (rq_2, \infty)) = \emptyset\}.$$

If $\boldsymbol{q} \notin D$, then $r_D < 1$, X_1 and Y_1 are asymptotically independent and the residual dependence coefficient is $\eta = r_D^a$.

The definition of r_D implies that, if D is convex, $r_D \boldsymbol{q}$ lies on the boundary of D and thus $n_D(r_D \boldsymbol{q}) = 1$. Using the homogeneity of n_D and Lemma 1 one can get $\eta = n_D(\boldsymbol{q})^{-a}$; see also Nolde and Wadsworth (2021) for further details. In practice, a random vector is often described by its multivariate probability density function. The relation between the density function and its limit set is given in the following lemma and thereby a convenient way to calculate the residual tail dependence coefficient η is given.

Lemma 2 (Proposition 3.1 in Nolde (2014)) *Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ be independent and identically distributed random vectors with a continuous positive density g on $\mathbb{R}^d$. Let $\gamma := -\log g$. Suppose there exists a function ℓ on $\mathbb{R}^d$ which is positive outside a bounded set, and a nonzero vector $\boldsymbol{v}$ such that*

$$\frac{\gamma(r_n \boldsymbol{u}_n)}{\gamma(r_n \boldsymbol{v})} \to \ell(\boldsymbol{u}), \ \text{for some } r_n \to \infty, \ \text{and any } \boldsymbol{u}_n \to \boldsymbol{u} \in \mathbb{R}^d.$$

Then the sequence of sample clouds $N_n = \{\boldsymbol{Z}_1/r_n, \ldots, \boldsymbol{Z}_n/r_n\}$ converges as $n \to \infty$ onto a limit set D whose gauge function is given by $n_D = \ell^\theta$ for some positive constant θ.

For details about the proof of Lemma 1 and Lemma 2 and further background knowledge, we refer to Nolde (2014) and references therein. We are now ready to use these two results to show that the GH distribution, except in the limiting case $\psi = 0$, is asymptotically independent and we give its residual dependence coefficient.

Proposition 1 *Let $(X_1, X_2)^\top \sim GH_2(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, and $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right), |\rho| \neq 1$. If $\psi > 0$, then X_1, X_2 are asymptotically independent and the residual dependence coefficient is*

$$\eta = \left\{ \sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma}) \boldsymbol{m}^\top \Sigma^{-1} \boldsymbol{m}} - \boldsymbol{m}^\top \Sigma^{-1} \boldsymbol{\gamma} \right\}^{-1},$$

where $\boldsymbol{m} = (m_1, m_2)^\top$ and

$$m_i = \frac{\gamma_i + \sqrt{\psi + \gamma_i^2}}{\psi}, \quad i = 1, 2.$$

Proof Note that both the tail dependence coefficient and the residual dependence coefficient are copula properties, i.e., they are invariant under strictly increasing marginal transformations. Hence, for simplicity we assume $\boldsymbol{\mu} = \boldsymbol{0}$ and study the tail dependence of $(X_1, X_2)^\top \sim \text{GH}_2(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{0}, \Sigma)$. Let f be the probability density function of $(X_1, X_2)^\top$. Using the asymptotic property of the Bessel function (see formula (9.7.2) in Abramowitz and Stegun (1972)),

we know that $K_\lambda(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) + o(1/\sqrt{x})$ as $x \to \infty$. Then, for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ and $s \to \infty$, we have

$$\frac{-\log\{f(s\boldsymbol{u})\}}{-\log\{f(s\boldsymbol{v})\}} \sim \frac{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})(\kappa + s^2 \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u})} - s\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma}}{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})(\kappa + s^2 \boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{v})} - s\boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{\gamma}}$$
$$\sim \frac{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u}} - \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma}}{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{v}} - \boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{\gamma}}.$$

For fixed $\boldsymbol{v} \neq \boldsymbol{0}$, let $h := \sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{v}} - \boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{\gamma}$. Then, consider the set D:

$$D = \{\boldsymbol{u} \in \mathbb{R}^2 : \left(\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u}} - \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma}\right)/h \leq 1\}.$$

Note that $\left(\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u}} - \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma}\right)/h \leq 1$ is equivalent to

$$(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u} \leq (h + \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma})^2,$$

which can be simplified to

$$(\psi + \gamma_2^2)u_1^2 + (\psi + \gamma_1^2)u_2^2 - 2(\rho\psi + \gamma_1\gamma_2)u_1 u_2 - 2h\{(\gamma_1 - \rho\gamma_2)u_1 + (\gamma_2 - \rho\gamma_1)u_2\}$$
$$\leq h^2(1 - \rho^2). \tag{5}$$

Since $\psi > 0, \rho \neq 1$, we have

$$\Delta = 4(\rho\psi + \gamma_1\gamma_2)^2 - 4(\psi + \gamma_2^2)(\psi + \gamma_1^2)$$
$$= -4(1 - \rho^2)\psi^2 - 4\psi(\gamma_1^2 + \gamma_2^2 - 2\rho\gamma_1\gamma_2)$$
$$< 0.$$

This implies that D is an ellipse, and thus convex. By Lemma 2, D is the limit set and its associated gauge function is $n_D(\boldsymbol{u}) = \{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1} \boldsymbol{\gamma})\boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{u}} - \boldsymbol{u}^\top \Sigma^{-1} \boldsymbol{\gamma}\}/h$. Furthermore, note that $n_D(\boldsymbol{0}) \leq 1$, which implies that $\boldsymbol{0}$ is contained in D. Hence, D is bounded and star shaped. Now it remains to find the coordinatewise supremum of D.

Note that the inequality (5) can be written as

$$(\psi + \gamma_2^2)u_1^2 - 2[(\rho\psi + \gamma_1\gamma_2)u_2 + h(\gamma_1 - \rho\gamma_2)]u_1 + (\psi + \gamma_1^2)u_2^2 - 2h(\gamma_2 - \rho\gamma_1)u_2$$
$$\leq h^2(1 - \rho^2).$$

Denote by $\boldsymbol{q} = (u_1^*, u_2^*)^\top$ the coordinatewise supremum of D, then we have

$$(\psi + \gamma_1^2)u_2^{*2} - 2h(\gamma_2 - \rho\gamma_1)u_2^* = h^2(1 - \rho^2) + \frac{\{(\rho\psi + \gamma_1\gamma_2)u_2^* + h(\gamma_1 - \rho\gamma_2)\}^2}{\psi + \gamma_2^2},$$

which can be simplified to

$$\psi u_2^{*2} - 2h\gamma_2 u_2^* - h^2 = 0.$$

Hence,

$$
\begin{aligned}
u_2^* &= \max\left(\frac{2h\gamma_2 + \sqrt{4h^2\gamma_2^2 + 4h^2\psi}}{2\psi}, \frac{2h\gamma_2 - \sqrt{4h^2\gamma_2^2 + 4h^2\psi}}{2\psi}\right) \\
&= \frac{h(\gamma_2 + \sqrt{h^2\gamma_2^2 + h^2\psi})}{\psi}.
\end{aligned}
$$

Similarly, we get $u_1^* = h(\gamma_1 + \sqrt{h^2\gamma_1^2 + h^2\psi})/\psi$. Since the GH distribution is closed under linear transformation, we know that $X_i \sim \mathrm{GH}_2(\lambda, \kappa, \psi, \gamma_i, 0, 1), i = 1, 2$. Denote the marginal density function and distribution function of $X_i, i = 1, 2$ as f_{X_i} and F_{X_i}, respectively, and write $1 - F_{X_i}(x) = e^{-\Psi_i(x)}$. Using the asymptotic relation $K_\lambda(x) \sim \sqrt{\pi/(2x)}e^{-x}, x \to \infty$ we get

$$
f_{X_i}(x) \sim c_i x^{\lambda-1}\exp\left\{-\left(\sqrt{\psi + \gamma_i^2} - \gamma_i\right)x\right\} \text{ as } x \to \infty,
$$

where c_i is a constant. As $\psi > 0$ we know that $\sqrt{\psi + \gamma_i^2} - \gamma_i > 0$. By Proposition 2 in von Hammerstein (2016), we have

$$
1 - F_{X_i}(x) \sim \frac{c_i x^{\lambda-1}}{\sqrt{\psi + \gamma_i^2} - \gamma_i}\exp\left\{-\left(\sqrt{\psi + \gamma_i^2} - \gamma_i\right)x\right\} \text{ as } x \to \infty.
$$

Hence, the functions $\Psi_i, i = 1, 2$ are both regularly varying at infinity with the same exponent $a = 1$.

As the supremum point $\boldsymbol{q}$ of an ellipse D satisfies $\boldsymbol{q} \notin D$, by Lemma 1 and Lemma 2, X_1 and X_2 are asymptotically independent and the residual dependence coefficient is

$$
\begin{aligned}
\eta &= n_D(\boldsymbol{q})^{-a} \\
&= \left[\left\{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})\boldsymbol{q}^\top \Sigma^{-1}\boldsymbol{q}} - \boldsymbol{q}^\top \Sigma^{-1}\boldsymbol{\gamma}\right\}/h\right]^{-1} \\
&= \left\{\sqrt{(\psi + \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma})\boldsymbol{m}^\top \Sigma^{-1}\boldsymbol{m}} - \boldsymbol{m}^\top \Sigma^{-1}\boldsymbol{\gamma}\right\}^{-1},
\end{aligned}
$$

where $\boldsymbol{m} = (m_1, m_2)^\top$ and

$$
m_i = \frac{\gamma_i + \sqrt{\psi + \gamma_i^2}}{\psi}, \quad i = 1, 2.
$$

When $\rho = 0$, which corresponds to independence between the components of the Gaussian random vector $\boldsymbol{W}$ and implies that the dependence between X_1 and X_2 is fully specified by the mixing variable R, we have $\eta = \{(\psi + \gamma_1^2 + \gamma_2^2)^{1/2}(m_1^2 + m_2^2)^{1/2} - m_1\gamma_1 - m_2\gamma_2\}^{-1}$, where m_1, m_2 are the same as above. One further observation is that when $\psi \to \infty$, which means the mixing variable R has a very light tail, we have $\eta = \sqrt{(1+\rho)/2}$. In this case, we obtain the same η as the elliptical GH distribution, i.e., when $\boldsymbol{\gamma} = \boldsymbol{0}$, and it equals the square root of the residual dependence coefficient of a bivariate Gaussian random vector with correlation ρ.
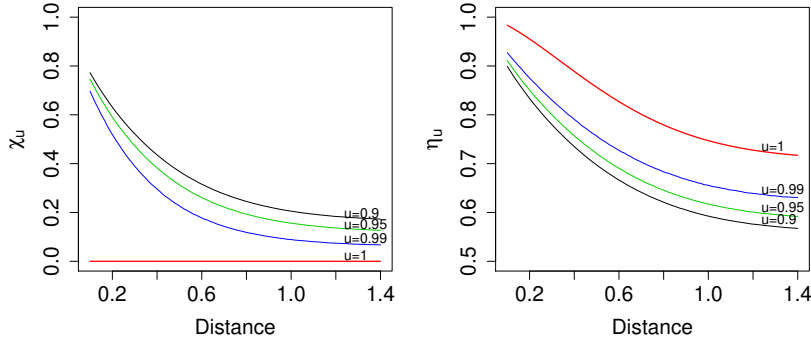
Fig. 1 Bivariate χ_u, η_u for $u = 0.9, 0.95, 0.99$ and their limits as $u \to 1$ of the GH distribution for $\lambda = -0.5, \kappa = \psi = 1, \boldsymbol{\gamma} = \mathbf{0}$, and Σ being a correlation matrix determined by a powered exponential correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\{-(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/0.6)^{3/2}\}$, where $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ is the Euclidean distance between sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. For $u < 1$, χ_u and η_u are calculated numerically using their definition (2) and (4), and their limits are calculated using the results in Proposition 1.

It is also important to note that when $\psi = 0, \boldsymbol{\gamma} \neq \mathbf{0}$, the sample cloud does not converge onto a bounded set and thus the method we used above fails. When $\psi = 0, \boldsymbol{\gamma} = \mathbf{0}$, the GH distribution reduces to the (asymptotically dependent) Student-t distribution and the tail dependence coefficient is known from the literature; see Embrechts et al. (2001).

Figure 1 illustrates the flexibility in extremal dependence structures of the GH distribution, by plotting the bivariate χ_u, η_u and their limits as a function of distance between the pairs, for a range of values of $u \in [0.9, 1)$, and $\lambda = -0.5, \kappa = \psi = 1, \boldsymbol{\gamma} = \mathbf{0}$, Σ being a correlation matrix determined by a powered exponential correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\{-(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/0.6)^{3/2}\}$, where $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ is the Euclidean distance between sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. The results indicate the slow convergence of χ_u and η_u to their limits, and at any observable levels, including $u = 0.99$, nonnegligible dependence may exist (χ_u could be as large as 0.4 or 0.5 for pairs at short distances), but the pairs are asymptotically independent. This emphasizes again the need for asymptotically independent models or models that can capture both asymptotic independence and dependence.

3 Copula-Based Inference

3.1 Copula Model and Likelihood Inference

As here we focus on modeling the spatial dependence, we take a copula approach. A copula is a multivariate distribution function with standard uniform margins. Thanks to Sklar's theorem (Sklar, 1959), for any continuous multidimensional distribution function there is a unique copula associated with it. This is the case for the multivariate GH distribution and we call the asso-

ciated copula the GH copula. Specifically, suppose $(X_1, \ldots, X_d)^\top$ follows a d-dimensional GH distribution F, which implies that the marginal distribution functions $F_1, \ldots, F_d$ are univariate GH distributions, then the GH copula is defined by

$$C(\boldsymbol{u}) = \Pr\{F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d\} = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\},$$
$$\boldsymbol{u} \in [0,1]^d,$$

where $F_i^{-1}, i = 1, \ldots, d$, is the inverse of F_i. Its density function thus can be derived easily as

$$c(\boldsymbol{u}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(\boldsymbol{u}) = \frac{f\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}}{f_1\{F_1^{-1}(u_1)\} \cdots f_d\{F_d^{-1}(u_d)\}},$$

where f is the joint density function and $f_1, \ldots, f_d$ are the marginal GH density functions of $X_1, \ldots, X_d$, respectively.

Due to the closed-form density of the multivariate GH distribution, inference based on the full likelihood is feasible. Here both in the simulation study and data applications, we fit the GH copula model to the whole dataset, rather than using a censored likelihood approach as Wadsworth and Tawn (2012). By doing so, we gain two major benefits: (i) we can avoid the computational issues associated with the censored likelihood, which requires calculations of potentially high-dimensional distribution functions and that often have to be evaluated numerically; (ii) we can model dependence both in the bulk and in the tail. However, we also point out that an obvious disadvantage of taking such a full likelihood approach is that the extremal dependence structure would not be as well captured as using a censored likelihood approach. Here we choose the full likelihood approach because the GH copula model is highly flexible, more so than most other spatial models. This means that it could be a reasonable approach, and our simulation study in Section 3.2 supports this.

We leverage the R (R Core Team, 2020) package *Rcpp* to evaluate the GH density function with low-level C++ language and the package *ghyp* to compute the quantiles of the GH distribution, which uses splines to interpolate the univariate GH distribution function and then finds the root efficiently with the uniroot function in R. These "tricks" have improved the efficiency of the algorithm and, as an illustration, the computational time for fitting the GH copula to a dataset with 100 sites and 2000 observations at each site is less than three hours on a laptop with 8 GB memory and 2.3 GHz Intel Core i5 processor.

3.2 Simulation Study

Simulation from the multivariate GH distribution is straightforward due to its stochastic representation (1). Specifically, independently sampling $R \sim \text{GIG}(\lambda, \kappa, \psi)$ and $\boldsymbol{W} \sim \mathcal{N}_d(\boldsymbol{0}, \Sigma)$ and plugging them into formula (1) yields exact samples from the distribution $\text{GH}_d(\lambda, \kappa, \psi, \boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma)$. Sampling from a
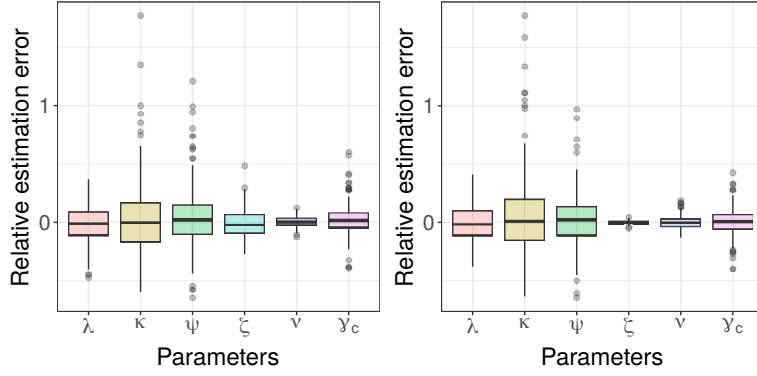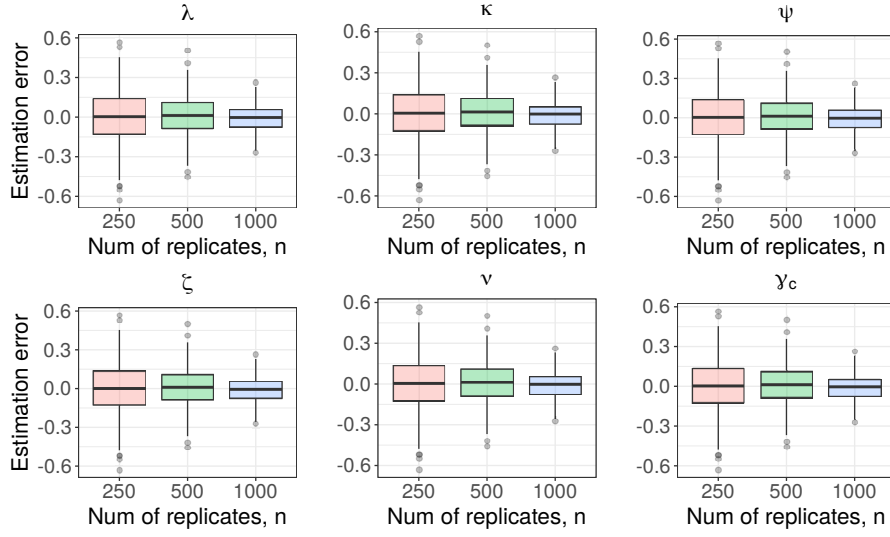
Fig. 2 Boxplots of relative estimation errors of all the parameters for different scenarios in the first simulation study. Left panel: scenario $\zeta = 0.7, \nu = 0.5, n = 500$; right panel: scenario $\zeta = 0.2, \nu = 1.5, n = 500$.

multivariate normal distribution is easy and sampling from a generalized inverse Gaussian distribution is feasible using the algorithm in Dagpunar (1989), which is implemented in the R package *ghyp*. Once a sample is generated from the multivariate GH distribution, one can obtain a sample for the GH copula by transforming the marginals to the standard uniform distribution, using the probability integral transform.

To reduce the number of parameters in the GH copula model, in both the simulation study and data application, we here assume that $\boldsymbol{\gamma} = \gamma_c \mathbf{1}_d, \gamma_c \in \mathbb{R}$, and Σ is a correlation matrix determined by a powered exponential correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\{-(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/\zeta)^\nu\}$, where $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ is the Euclidean distance between sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, $\zeta > 0$ is the range parameter, and $\nu \in (0, 2]$ controls the smoothness of the realized random fields.

In the first simulation study we consider 25 sites on the uniform grid $\{0, 0.25, 0.5, 0.75, 1\}^2$, and set $\lambda = \kappa = \psi = \gamma_c = 1$, $\zeta = 0.2$ or 0.7 (short or long range dependence), $\nu = 0.5$ or 1.5 (rough or smooth random field). Then we generate $n = 250, 500$ or 1000 replicates from the GH copula at these sites and fit the GH copula to the simulated dataset of each scenario. We run the experiment for 300 times and Figure 2 depicts boxplots of relative estimation errors (i.e., differences between parameter estimates and the true parameter values, divided by the true parameter values) for all of the six parameters when $\zeta = 0.7, \nu = 0.5, n = 500$ and $\zeta = 0.2, \nu = 1.5, n = 500$. The results indicate that the spatial range parameter ζ, the smoothness parameter ν, and the skewness parameter γ_c are relatively easier to estimate than the parameters λ, κ, ψ, which determine the shape of the mixing GIG distribution. We also observe that changing the values of ζ and ν does not affect the estimation of other parameters, but the larger value of ζ, which corresponds to stronger spatial dependence between two sites, leads to smaller variability of the estimate of ν, and larger value of ν also yields smaller variability of the estimate of ζ. One further observation is that when we increase the number of replicates n,

Fig. 3 Boxplots of estimation errors of all the parameters for the scenario $\zeta = 0.7, \nu = 1.5$ with different number of replicates n in the first simulation study.

the variability of the estimates of all parameters decreases, roughly at a rate of $\sqrt{n}$ as expected, as in shown in Figure 3.

In the second simulation study we illustrate the flexibility of the GH distribution in a misspecified setting. We consider 100 randomly sampled irregularly-spaced locations by randomly jittering gridded locations, in order to keep a relatively even spatial coverage. Specifically, for each location of the regular grid $\{0.1, 0.2, \ldots, 1\}^2$, we randomly generate two numbers from the uniform distribution on $(0, 0.05)$, and then subtract these two numbers from each coordinate of the location. The new locations are shown in the left panel of Figure 4. We then simulate $n = 1000$ replicates from the inverted Brown–Resnick process (Kabluchko et al., 2009; Wadsworth and Tawn, 2012) using the extremal functions approach of Dombry et al. (2016), with the variogram of the Brown–Resnick process defined as $\gamma(\boldsymbol{h}) = 2\|\boldsymbol{h}\|/\zeta$, where $\zeta = 0.3$. This model is known to be asymptotically independent with bivariate residual tail dependence coefficient function $\eta(\boldsymbol{h}) = 1/[2\Phi\{\sqrt{\gamma(\boldsymbol{h})}/2\}]$, where $\Phi(\cdot)$ is the standard normal distribution function. We now fit the GH copula and the normal inverse Gaussian (NIG) copula, which is a subclass of the GH copula when $\lambda = -1/2$, to the whole simulated dataset. The right panel of Figure 4 displays the true residual tail dependence coefficient of the inverted Brown–Resnick process, its counterpart from the fitted GH copula and NIG copula model, and empirical estimates of η_u with $u = 0.95$, as defined in (4), based on the simulated dataset and on the fitted GH model. The results show that, despite fitting the GH copula to the entire dataset without censoring low non-extreme values, the GH copula captures the residual tail dependence quite well for the simulated dataset. Furthermore, the shape of the fitted residual tail
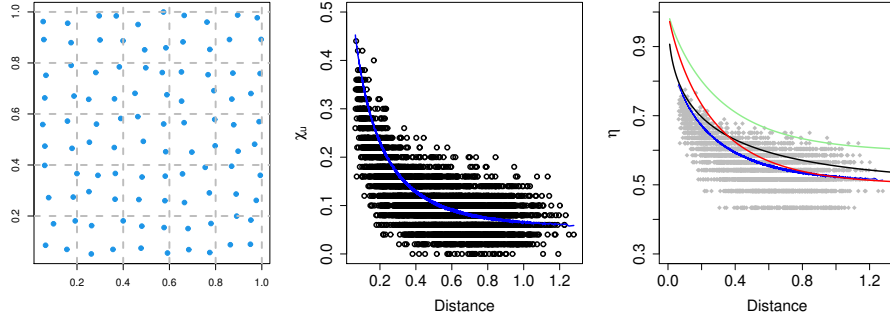
Fig. 4 Left panel: randomly generated sites in the second simulation study. Middle panel: empirical estimate of χ_u with $u = 0.95$ (black points) and its counterpart from the fitted GH copula model (blue line), plotted as a function of distance. Right panel: true η of the inverted Brown–Resnick process (black line), theoretical η of the fitted normal inverse Gaussian copula (green line), theoretical η of the fitted GH copula (red line), empirical estimate of η_u with $u = 0.95$ (grey points) and its counterpart from the fitted GH copula model (blue line), plotted as a function of distance.

dependence coefficient function, with respect to distance, resembles its true function very closely, although a slight bias exists as expected. The middle panel of Figure 4 displays the empirical bivariate tail dependence coefficient χ_u with $u = 0.95$ of the simulated data from the inverted Brown-Resnick process and its counterpart based on 10^6 replicates simulated from the fitted GH copula model, plotted as a function of distance. This confirms that the extremal dependence can be captured well by the GH copula, even in a misspecified setting.

4 Environmental Applications

4.1 Wave Height Example

We first consider the hindcast dataset of significant wave heights analyzed by Wadsworth and Tawn (2012) and Huser and Wadsworth (2019). This dataset contains eight observations per day over a period of 31 years, at 50 spatial locations in the North Sea. Similarly to Huser and Wadsworth (2019), only the first observation on each day in the winter months (December, January, and February) at a subset of 20 sites is used to reduce temporal dependence and ease the computational burden. The locations of the 20 sites are shown in the left panel of Figure 5. Note that here distance is measured in units of latitude (one unit $\approx$ 111 km).

The margins at each site are transformed to standard uniform following the semiparametric approach of Coles and Tawn (1991). Then, three different models are fitted to the standardized dataset, namely the Huser–Wadsworth model, the Gaussian copula model and the GH copula model. To illustrate the flexibility of our proposed GH model to capture the upper tail dependence in a
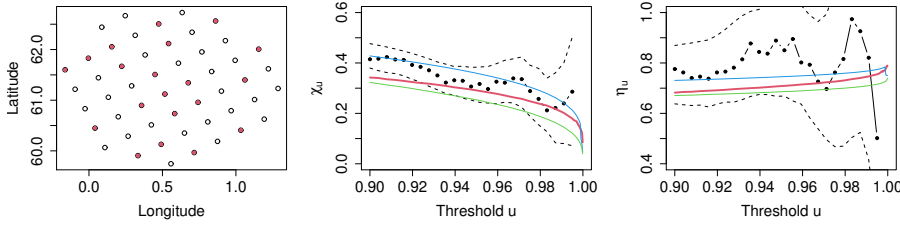
Fig. 5 Left panel: locations of the 20 sites which are randomly selected in the hindcast wave height data application. Middle and right panels: estimates of 20-variate χ_u (middle panel) and η_u (right panel) for the hindcast wave height data, as a function of the threshold u. Central black dots: empirical estimates of χ_u and η_u; dashed line: approximate pointwise 95% confidence intervals based on a stationary bootstrap procedure; thick solid red line: fit from the GH copula; thin solid blue line: fit from the Huser–Wadsworth model; thin solid green line: fit from the Gaussian model.

disadvantageous situation, we first use a censored likelihood scheme to fit the Huser–Wadsworth model and the Gaussian copula model, which helps them better capture the extremal dependence structure; see equation (18) in Huser and Wadsworth (2019) for the censoring scheme. Then we fit the GH copula to the data without censoring low observations.

The middle and right panels of Figure 5 display the fitted values of 20-variate χ_u and η_u, as defined in Section 2.2, for the three models. The uncertainty measures are based on the same stationary bootstrap procedure as in Huser and Wadsworth (2019), i.e., 200 bootstrap samples are generated using the stationary bootstrap which sample blocks of geometric length. Strikingly, the results show that the GH copula can capture the extremal dependence better than the Gaussian copula, even though a censored likelihood has been used for the Gaussian copula. The performance of the GH copula is slightly worse than the Huser–Wadsworth copula, but the differences are fairly minor overall and the GH model still provides a very good fit in the upper tail despite its full uncensored likelihood estimation approach. We stress again that the Huser–Wadsworth copula is specially designed to model the upper tail and a censored likelihood has been used for this purpose, while our proposed GH model is here used to model the entire distribution, from low to high quantiles.

To further investigate the fitted model, we plot the estimated bivariate survival probability isolines for two pairs of locations at different distances in Figure 6. More precisely, we consider 2500 points $(u_1, u_2)^\top$ on the grid $\{0.01, 0.03, \ldots, 0.99\}^2$ and estimate the survival probabilities $\Pr\{F_1(X_1) > u_1, F_2(X_2) > u_2\}$ based on the fitted GH copula evaluated at these points by Monte Carlo. Then an estimate of the survival probabilities at other points in the unit square is obtained by interpolation. Note that here we only considered isolines with survival probabilities $0.2, 0.1$ and 0.05 since obtaining accurate estimates of lower survival probabilities may require far more Monte Carlo simulations and this is computationally too demanding. Note that, in practice, such survival probability estimates are commonly used as risk measures (Di Bernardino et al., 2015; Zscheischler et al., 2017). Figure 6 also shows
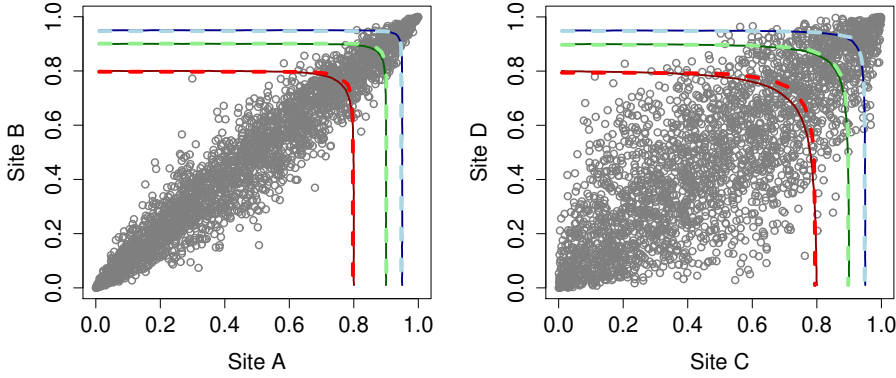
Fig. 6 Plot of the wave height data at two pairs of locations with different distances and estimated survival function isolines, after standardization to the uniform scale. In each panel, isolines produced using the fitted GH copula model are dark-colored thin solid lines, while isolines produced using the empirical survival copula are light-colored thick dashed lines. The distance between site A and B (left panel) is around 98 km and the one between site C and site D (right panel) is around 217 km. Estimated isolines correspond to survival probabilities $\Pr\{F_1(X_1) > u_1, F_2(X_2) > u_2\} = 0.2, 0.1, 0.05$ (red, green, blue lines, respectively).

the estimated bivariate survival probability isolines for the same pairs of locations produced using the empirical survival copula. One can see that the isolines obtained using these two different (parametric and non-parametric) methods almost coincide with each other, and this confirms that the GH copula fits quite well, especially in the joint upper tail. If one wishes to further assess the accuracy of the estimated isolines, the diagnostic tools in Murphy-Barltrop et al. (2021); Cooley et al. (2019) might be useful. We also fit the Huser–Wadsworth model and the Gaussian model without censoring the low observations. The log-likelihood of GH copula is around 2000 units larger than that of the Huser–Wadsworth model and 6000 units larger than that of the Gaussian copula model, which indicates that the GH copula is clearly more flexible than the other two if all observations are considered.

4.2 Wind Gust Example

To illustrate the computational benefits of our proposed model, we now consider a higher-dimensional dataset of daily maximum wind gusts from the state of Oklahoma, USA, which can be freely downloaded from mesonet.org. This dataset contains daily measurements of highest 3-second wind speed from January 1, 1995 to December 31, 2020, at 120 observational stations across the state of Oklahoma. To induce approximate stationarity, we only consider the summer observations, i.e., observations in July, August and September, which are often the highest in a year. As some observational stations have many missing values, we delete the ones with more than 100 missing values, resulting in 95 sites with 2392 observations at each site. The locations of these
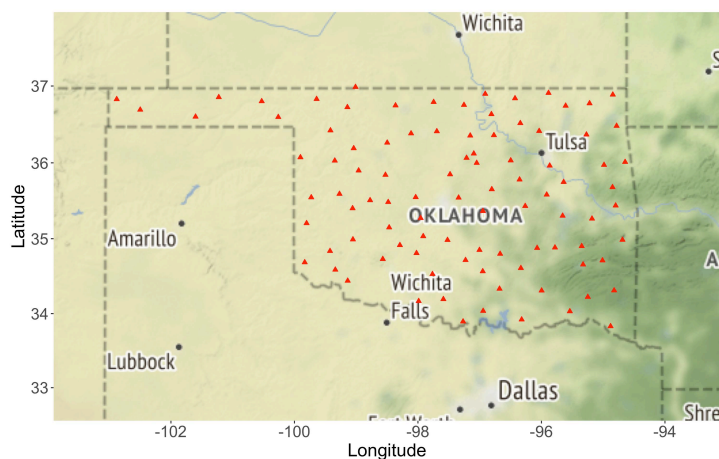
Fig. 7 Locations of the 95 observational sites in the state of Oklahoma, USA, in the wind gust data application.
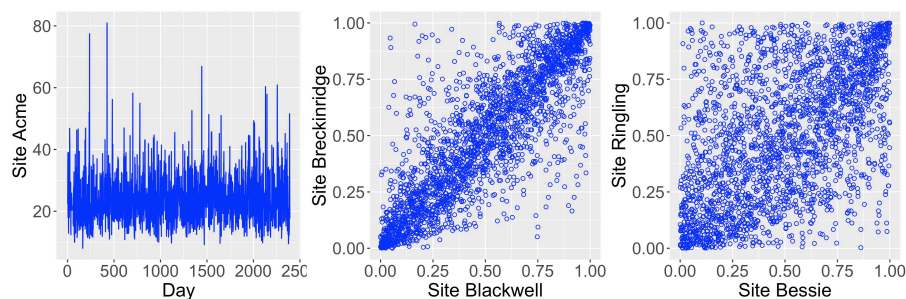
Fig. 8 Exploratory data analysis of the wind gust data. Left panel: time series at site Acme; middle and right panels: scatter plots of observations at site Blackwell and Breckinridge (distance 55 km), and at site Bessie and Ringling (distance 190 km).

95 sites are shown in Figure 7. The minimum distance between two sites is around 12 km and the maximum distance is around 800 km. The time series at each site appears approximately stationary (see the time series at a randomly selected site in the left panel of Figure 8), hence we work directly with this dataset and transform all the marginals to standard uniform using the nonparametric approach based on ranks. Alternatively, we could use the semiparametric approach of Coles and Tawn (1991) as in the wave height example. On the middle and right panels, we show the scatter plots of observations at the Blackwell and Breckinridge sites (distance 55 km), and at the Bessie and Ringling sites (distance 190 km), on the uniform scale. One can observe that as the distance between the pair of sites increases, the dependence decreases, but it is hard to tell by the naked eye whether the pairs are asymptotically independent or dependent.

Table 2 Number of parameters and AIC values for the different models fitted in our wind gust data application

Model	# of parameters	AIC
Gaussian copula	2	-298898
t copula	3	-324069
Hyperbolic copula	5	-316636
NIG copula	5	-327035
GH copula	6	$\mathbf{-328141}$

We then fit the GH copula model to the standardized dataset, as well as four of its subclasses and limiting models, namely the Gaussian copula, the t copula model, the hyperbolic copula and the NIG copula model, based on a full likelihood approach. The computational time for fitting the GH copula is less than three hours on a laptop, which is a significant speed-up compared to the censored Huser–Wadsworth model in such high dimensions. It takes less than ten minutes to fit the Gaussian copula and t copula due to the efficient computation of their quantile functions in R. Table 2 shows the number of parameters in each model and the Akaike information criterion (AIC) of the fitted models. The results indicate that the general GH copula has the best performance as expected, and the NIG submodel has the second best performance. The large difference between the AIC values of the GH copula model and its limiting cases, the widely used Gaussian copula model and the t copula model, implies that there is a clear advantage of using the GH copula model for this dataset despite its higher model complexity.

To assess how well our model captures the dependence both in the bulk and the tail of the data, we consider Spearman's rank correlation S_ρ and the bivariate extremal dependence measures χ_u and η_u described in Section 2.2. We first calculate the empirical estimates of S_ρ, χ_u and η_u with $u = 0.95$ from the data, and their model-based counterparts from the estimated model. Then we compute the bivariate kernel density estimators for pairs of observed empirical S_ρ and fitted S_ρ, pairs of observed empirical χ_u and fitted χ_u, and pairs of observed empirical η_u and fitted η_u, using the R package *ks*. Figure 9 depicts the contours at level $25\%, 50\%$ and 75% of these kernel density estimators. The results show that our model can capture strong dependencies (at short and moderate distances) both in the bulk and the tail very well, but it overestimates weak dependencies (at large distances). This is not very surprising as there are fewer pairs of sites that are weakly dependent (at long distances), so that strongly dependent pairs influence model estimation to a higher degree. In particular, the spatially-constant mixing variable R is constrained to provide a good fit at short distances, which simultaneously induces stronger dependencies across the entire spatial domain. This can also be seen from the discussion on the values of η for special cases of the GH distribution in Section 2.3. When the mixing variable R has very light tails ($\psi \to \infty$), as the distance between two sites tends to infinity and their correlation $\rho \to 0$,
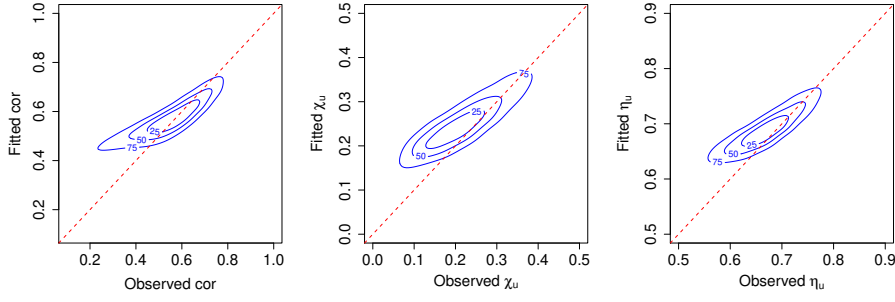
Fig. 9 Contours at level 25%, 50% and 75% of the bivariate kernel density estimators for pairs of observed empirical S_ρ and fitted S_ρ (left), pairs of observed empirical χ_u and fitted χ_u (middle), and pairs of observed empirical η_u and fitted η_u (right). A good fit should have density contours concentrated around the red dashed diagonal line.

the residual dependence coefficient η converges to $1/\sqrt{2}$, which is much larger than $1/2$ and indicates that a strong dependence is still present in this case.

5 Discussion

The GH distribution has a long history, and found popularity in financial modeling. In this work we have re-investigated its tail dependence properties, pointed out the mistakes in its derivations in the literature, and gave a corrected description. Based on this result, we propose to use the GH copula for spatial extremes when asymptotic independence is present, which contributes to the spatial extremes literature, as existing models in this setting that are both flexible and can be fitted efficiently are scarce. We demonstrate the flexibility of this model both by a simulation study and two environmental data applications.

Unlike the asymptotic dependence case, where the scientific community broadly agrees that max-stable models should be fitted to block maxima, and generalized Pareto models should be fitted to threshold exceedances, there is no such consensus in the asymptotic independence case yet. Wadsworth and Tawn (2012) argue that it is often more natural, especially in the presence of asymptotic independence, to model the extremes of original events rather than site-wise maxima, and they suggest using a censored likelihood approach to better capture the extremal dependence. Here, we did not censor the low observations when using the GH copula, because, as shown in the wave height data application, the non-censored GH copula can capture the tail dependence even better than the censored Gaussian copula and the high flexibility of the GH copula allows this full likelihood approach. Moreover, in this case we can gain significant computational efficiency and also capture dependence in the bulk, as shown in the wind gust data application.

One interesting research direction is to investigate the tail properties in the limiting case of the GH distribution when $\psi = 0$, which is an interesting

distribution in its own as its marginal distribution has asymmetric tails, i.e., one has a power decay and the other has an exponential decay, and it reduces to the t distribution when the skewness parameter γ tends to zero. Unfortunately, the two different methods that we used in this work to derive the tail properties both fail in the limiting case when $\psi = 0$ and the skewness parameter is not zero. Hence, this remains an open problem. Another future research direction is to investigate the tail dependence properties of more general normal mean-variance mixtures when the mixing distribution is specified only through its tail behavior. Finally, as discussed in the wind gust data application, a weakness of the GH copula model is that independence cannot be captured as distance tends to infinity. This is a typical weakness for all random scale (and location) models. Hence, another future research direction is to try and overcome this issue and improve the performance of the GH copula model to better capture long-range weak dependence structures. One possibility could be to use the random partitioning approach of Morris et al. (2017).

References

Abramowitz M, Stegun IA (eds) (1972) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th edn. National Bureau of Standards, United States of America

Barndorff-Nielsen OE (1977) Exponentially decreasing distributions for the logarithm of particle size. Proceedings of the Royal Society of London, Series A 353:401–419

Barndorff-Nielsen OE (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. Scandinavian Journal of Statistics 24(1):1–13

Blaesild P, Jensen JL (1981) Multivariate distributions of hyperbolic type. Statistical Distributions in Scientific Work 4:45–66

Bortot P, Coles S, Tawn JA (2000) The multivariate gaussian tail model: an application to oceanographic data. Journal of the Royal Statistical Society (Series C) 49(1):31–49

Castro Camilo D, Huser R (2020) Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. Journal of the American Statistical Association 115:1037–1054

Castruccio S, Huser R, Genton MG (2016) Higher-order composite likelihood inference for max-stable distributions an processes. Journal of Computational and Graphical Statistics 25(4):1212–1229

Coles S, Tawn JA (1991) Modelling extreme multivariate events. Journal of the Royal Statistics Society (Series B) 53:377–392

Coles S, Heffernan J, Tawn JA (1999) Dependence measures for extreme value analyses. Extremes 2(4):339–365

Cooley D, Thibaud E, Castillo F, Wehner MF (2019) A nonparametric method for producing isolines of bivariate exceedance probabilities. Extremes 22:373–390

Dagpunar JS (1989) An easily implemented generalized inverse Gaussian generator. Communications in Statistics - Simulation and Computation 18:703–710

Davison AC, Huser R (2015) Statistics of extremes. The annual review of statistics and its applications 2:203–235

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Statistical Science 27(2):161–186

de Haan L (1984) A spectral representation for max-stable processes. The Annals of Probability 12(4):1194–1204

de Haan L, Pereira TT (2006) Spatial extremes: models for the stationary case. The Annals of Statistics 34(1):146–168

Di Bernardino E, Fernández-Ponce JM, Palacios-Rodríguez F, nolo MRRG (2015) On multivariate extensions of the conditional value-at-risk measure. Insurance: Mathematics and Economics 61:1–16

Dombry C, Engelke S, Oesting M (2016) Exact simulation of max-stble process. Biometrika 103(2):303–317

Embrechts P, McNeil A, Straumann D (2001) Correlation and dependence in risk management: properties and pitfalls. In: Dempster M, Moffatt H (eds) Risk Management: Value at Risk and Beyond, Cambrige University Press, pp 176–223

Engelke S, Opitz T, Wadsworth JL (2019) Extremal dependence of random scale constructions. Extremes 22:623–666

Ferreira A, de Haan L (2014) The generalized Pareto process; with a view towards application and simulation. Bernoulli 20(4):1717–1737

Gong Y, Huser R (2021) Asymmetric tail dependence modeling, with application to cryptocurrency market data. Annals of Applied Statistics To appear

Heffernan JE, Tawn JA (2004) A conditional approach to multivariate extreme values. Journal of the Royal Statistics Society (Series B) 66:497–546

Huser R, Wadsworth JL (2019) Modeling spatial processes with unknown extremal dependence class. Journal of the American Statistical Association 114(525):434–444

Huser R, Wadsworth JL (2020) Advances in statistical modeling of spatial extremes. Wiley Interdisciplinary Reviews: Computational Statistics e1537

Huser R, Opitz T, Thibaud E (2017) Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. Spatial Statistics 21:166–186

Jamalizadeh A, Balakrishnan N (2019) Conditional distributions of multivariate normal mean-variance mixtures. Statistics and Probability Letters 145:312–316

Kabluchko Z, Schlather M, de Haan L (2009) Stationary max-stable fields associated to negative definite functions. The Annals of Probability 37(5):2042–2065

Krupskii P, Huser R, Genton MG (2018) Factor copula models for replicated spatial data. Journal of the American Statistical Association 113(521):467–479

Le PD, Davison AC, Engelke S, Leonard M, Westra S (2018) Dependence properties of spatial rainfall extremes and areal reduction factors. Journal of Hydrology 565:711–719

Ledford AW, Tawn JA (1996) Statistics for near independence in multivariate extreme values. Biometrika 83(1):169–187

Manner H, Segers J (2011) Tails of correlation mixtures of elliptical copulas. Insurance: Mathematics and Economics 48:153–160

McNeil AJ, Frey R, Embrechts P (2005) Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press

Morris SA, Reich BJ, Thibaud E, Cooley D (2017) A space-time skew-t model for threshold exceedances. Biometrics 73:749–758

Murphy-Barltrop CJR, Wadsworth JL, Eastoe EF (2021) On the estimation of bivariate return curves for extreme values, available from https://arxiv.org/abs/2107.01942

Nolde N (2014) Geometric interpretation of the residual dependence coefficient. Journal of Multivariate Analysis 123:85–95

Nolde N, Wadsworth JL (2021) Linking representations for multivariate extremes via a limit set. Advances in Applied Probability, to appear

Padoan SA, Ribatet M, Sisson SA (2010) Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105(489):263–277

Prause K (1999) The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. PhD thesis, Albert-Ludwigs-Universität Freiburg

R Core Team (2020) R: A language and environment for statistical computing, r Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rootzén H, Tajvidi N (2006) Multivariate generalized Pareto distributions. Bernoulli 12(5):917–930

Schlueter S, Fischer M (2012) The weak tail dependence coefficient of the elliptical generalized hyperbolic distribution. Extremes 15:159–174

Sklar A (1959) Functions de répartition à n dimensions et leurs marges. Publications de l'institut de Statistique de l'Université de Paris 8:229–231

von Hammerstein EA (2016) Tail behavior and tail dependence of generalized hyperbolic distributions. In: Kallsen J, Papapantoleon A (eds) Advanced Modelling in Mathematical Finance, Springer, Cham, pp 3–40

Wadsworth JL, Tawn JA (2012) Dependence modelling for spatial extremes. Biometrika 99(2):253–272

Wadsworth JL, Tawn JA (2019) Higher-dimensional spatial extremes via single-site conditioning, available from https://arxiv.org/abs/1912.06560

Zscheischler J, Orth R, Seneviratne SI (2017) Bivariate return periods of temperature and precipitation explain a large fraction of european crop yields. Biogeosciences 14:3309–3320