Tukey Max-Stable Processes for Spatial Extremes

Ganggang Xu[1] and Marc G. Genton[2]

September 11, 2016

Abstract

We propose a new type of max-stable process that we call the Tukey max-stable process for spatial extremes. It brings additional flexibility to modeling dependence structures among spatial extremes. The statistical properties of the Tukey max-stable process are demonstrated theoretically and numerically. Simulation studies and an application to Swiss rainfall data indicate the effectiveness of the proposed process.

Some key words: Brown–Resnick process; Composite likelihood; Extremal coefficient; Extremal-t process; Geometric Gaussian process, Max-stable process; Spatial extreme.

Short title: Tukey max-stable processes

[1]Department of Mathematical Sciences, Binghamton University, State University of New York, Binghamton, New York 13902, USA.
E-mail: gang@math.binghamton.edu

[2]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
E-mail: marc.genton@kaust.edu.sa

1 Introduction

Max-stable processes are widely used to model spatial extremes because they naturally arise as the limits of pointwise maxima of rescaled stochastic processes (de Haan & Ferreira, 2006) and therefore extend generalized extreme-value distributions to that setting; see the reviews of recent advances in the statistical modeling of spatial extremes by Davison et al. (2012), Cooley et al. (2012), Ribatet (2013) and Davison & Huser (2015). A useful spectral characterisation of simple max-stable processes for which the margins are standardized to unit Fréchet distributions was proposed by de Haan (1984), Penrose (1992) and Schlather (2002). Specifically, let $\{R_j\}_{j=1}^{\infty}$ be the points of a Poisson process on $(0, \infty)$ with intensity $r^{-2}dr$, and let $\{W_j(s)\}_{j=1}^{\infty}$ be independent replicates of a nonnegative stochastic process, $W(s)$, on $\mathbb{R}^d$ with continuous sample paths satisfying $\mathrm{E}\{W(s)\} = 1$ for all $s \in \mathbb{R}^d$. Then,

$$Z(s) = \max_{j \geq 1}\{R_j W_j(s)\} \tag{1}$$

is a max-stable process on $\mathbb{R}^d$ with unit Fréchet marginal distributions; that is, $\mathrm{pr}\{Z(s) < z\} = \exp(-1/z)$ for all $z > 0$ and $s \in \mathbb{R}^d$. The finite-dimensional distributions of simple max-stable processes follow from (1) for any set of D locations, $s_1, \ldots, s_D$:

$$\mathrm{pr}\{Z(s_1) \leq z_1, \ldots, Z(s_D) \leq z_D\} = \exp\{-V_{s_1,\ldots,s_D}(z_1, \ldots, z_D)\}, \quad z_1, \ldots, z_D > 0, \tag{2}$$

where the exponent measure

$$V_{s_1,\ldots,s_D}(z_1, \ldots, z_D) = \mathrm{E}\left[\max_{i=1,\ldots,D}\left\{\frac{W(s_i)}{z_i}\right\}\right]$$

captures all the extremal dependence information. The latter can be summarized by the extremal coefficient function (Schlather & Tawn, 2003):

$$\theta_D(s_1, \ldots, s_D) = V_{s_1,\ldots,s_D}(1, \ldots, 1) = \mathrm{E}[\max_{i=1,\ldots,D}\{W(s_i)\}] \in [1, D],$$

where the values 1 and D for θ_D correspond to complete dependence and independence, respectively. For a pair of locations (s_1, s_2), under the stationarity assumption of the latent Gaussian random field $W(s)$, the bivariate extremal coefficient function $\theta_2(s_1, s_2) \equiv \theta(h)$ with $h = \|s_1 - s_2\|$.

Different choices for $W(s)$ in (1) lead to various models for spatial max-stable processes proposed in recent years. In a seminal unpublished University of Surrey 1990 technical report, R. L. Smith proposed a Gaussian extreme-value process based on $W(s) = \phi_d(s - U; \Sigma)$, where $\phi_d(\cdot; \Sigma)$ denotes the d-dimensional Gaussian probability density function with covariance matrix Σ and U is a unit rate Poisson process on $\mathbb{R}^d$. Schlather (2002) considered the extremal-Gaussian process obtained with $W(s) = (2\pi)^{1/2} \max\{0, \varepsilon(s)\}$ where $\varepsilon(s)$ is a stationary standard Gaussian process with correlation function $\rho_\Psi(h)$ and parameters Ψ. A generalization called the extremal-t process (Nikoloulopoulos et al., 2009; Opitz, 2013) was defined by taking $W(s) = \pi^{1/2} 2^{1-\nu/2} [\Gamma\{(\nu+1)/2\}]^{-1} \max\{0, \varepsilon(s)\}^\nu$ where $\Gamma(\cdot)$ is the gamma function and $\nu > 0$ the degrees of freedom. The case $\nu = 1$ reduces to the extremal-Gaussian process. The pairwise extremal coefficient for the extremal-t process is

$$\theta_t(h) = 2T_{\nu+1}\left\{(\nu+1)^{1/2}\sqrt{\frac{1-\rho_\Psi(h)}{1+\rho_\Psi(h)}}\right\},$$

where T_ν denotes the cumulative distribution function of a standard Student-t random variable with $\nu > 0$ degrees of freedom. It is trivial to see that $\theta_t(h) \leq \theta_t^0$ with $\theta_t^0 = 2T_{\nu+1}\{(\nu+1)^{1/2}\}$ and that as $\rho_\Psi(h) \to 0$, $\theta_t(h) \to \theta_t^0$. The θ_t^0 can only approach 2, which corresponds to independence, when $\nu \to \infty$, although this can be resolved by incorporating a compact random set element (Schlather, 2002; Davison & Gholamrezaee, 2012; Huser & Davison, 2014). Modelling of extreme values with asymptotically independent processes has been discussed by de Haan & Zhou (2011), Wadsworth & Tawn (2012) and Padoan (2013).

Another popular spatial max-stable model is the so-called geometric Gaussian process for which $W(s) = \exp\{\sigma\varepsilon(s) - \sigma^2/2\}$ with some $\sigma > 0$. The resulting extremal coefficient is

$$\theta_{geo}(h) = 2\Phi\left\{\sigma\sqrt{\frac{1-\rho_\Psi(h)}{2}}\right\}, \tag{3}$$

where Φ denotes the cumulative distribution function of a standard Gaussian random variable. Similar to the extremal-t process, $\theta_{geo}(h)$ is bounded above by $\theta_{geo}^0 = 2\Phi\left(\sigma/\sqrt{2}\right) < 2$ and the upper-bound $\theta_{geo}^0 \to 2$ only if $\sigma \to \infty$. The geometric Gaussian process can be viewed as a special case of the well-known Brown–Resnick process (Brown & Resnick, 1977; Kabluchko et al., 2009), where the latent process is defined as $W(s) = \exp\{\tilde{\varepsilon}(s) - \gamma(s)\}$ with $\tilde{\varepsilon}(s)$ being an intrinsically stationary Gaussian process with semivariogram $\gamma(\cdot)$ and a starting location at $s = 0$ such that $\tilde{\varepsilon}(0) = 0$ almost surely. One simple example of such $\tilde{\varepsilon}(s)$ is the Brownian motion.

By observing that $W(s)$ is a log-Gaussian process for the geometric Gaussian process suggests new possibilities. Indeed, Tukey (1977) introduced a generalisation of the log-Gaussian distribution called the g-and-h distribution and defined as the distribution of

$$Y = \frac{\exp(\tilde{g}Z) - 1}{\tilde{g}} \exp\left(\frac{\tilde{h}Z^2}{2}\right)$$

where Z is a standard Gaussian random variable. Here, $\tilde{g}$ can take any real value (positive or negative) and controls skewness whereas $\tilde{h} > 0$ regulates the heaviness of the tails of the distribution. Tukey's g-and-h distribution is endowed with various appealing properties and has been shown to be useful in many applications. For more detailed studies of the Tukey g-and-h distribution, see, for example, Martinez & Iglewicz (1984), MacGillivray (1992), and Xu & Genton (2015, 2016).

Following Tukey's idea, we propose an extension of the geometric Gaussian process based on a new spectral process $W(s)$ defined as

$$W(s) = \exp\left\{\frac{a}{2}\varepsilon^2(s) + b\varepsilon(s) + \frac{1}{2}\log(1-a) - \frac{b^2}{2(1-a)}\right\}, \quad a < 1, \quad b \geq 0, \tag{4}$$

which satisfies $\mathrm{E}\{W(s)\} = 1$ and where $\varepsilon(s)$ is a stationary standard Gaussian process with correlation function $\rho_\Psi(h)$. The resulting max-stable process, $Z(s)$, generated from (1) with (4), is referred to as the Tukey max-stable process. An additional parameter, a, is introduced here to increase the flexibility of modeling the dependence of spatial extremes. The parameter b is restricted to $b \geq 0$ to ensure identifiability. When $a = 0$ and $b = \sigma$, the Tukey max-stable process reduces to a classical geometric Gaussian process. When $0 < a < 1$, the spectral process $W(s)$

is uniformly bounded from below as

$$W(s) \geq \exp\left\{\frac{1}{2}\log(1-a) - \frac{b^2}{2a(1-a)}\right\}, \quad 0 < a < 1, \quad s \in \mathbb{R}^d, \tag{5}$$

while in the case when $a < 0$, $W(s)$ is uniformly bounded from above by

$$W(s) \leq \exp\left\{\frac{1}{2}\log(1-a) - \frac{b^2}{2a(1-a)}\right\}, \quad a < 0, \quad s \in \mathbb{R}^d. \tag{6}$$

Note that $W(s)$ in (4) is unbounded for $a = 0$ and is not well defined when $a \geq 1$. Bounds (5) and (6) may be of particular interests for different applications. The upper bound (6) is particularly attractive in the sense that the simulation of the resulting max-stable process $Z(s)$ defined in (1) can be exact by using a simple algorithm (Schlather, 2002). To the best of our knowledge, most, if not all, existing max-stable processes with representation (1) do not have such a nice property. We defer a more detailed discussion to Subsection 2.4.

The remainder of the paper is organized as follows. In Section 2, we derive the finite dimensional distributions of the Tukey max-stable process for $D = 2$ as well as its bivariate extremal coefficient. We also describe a composite likelihood procedure for fitting purpose and an exact simulation method. In Section 3, we report the results of Monte Carlo simulations. In Section 4, we fit our Tukey max-stable model to rainfall data from Switzerland and compare it with other models. We end the paper with a discussion in Section 5. The proofs of the theoretical results are provided in an Appendix.

2 Properties of the Tukey max-stable process

2.1 Bivariate distribution function

Although the finite-dimensional distributions of a simple max-stable process are given in (2), the explicit form of $V_{s_1,\ldots,s_D}(z_1,\ldots,z_D)$ is generally difficult to derive for $D>2$ except in some specific cases (Genton et al., 2011; Huser & Davison, 2013). The following proposition gives the bivariate cumulative distribution function of the proposed Tukey max-stable process. We write ρ for $\rho_\Psi(h)$ whenever there is no ambiguity.

Proposition 1 *Let ρ be the correlation coefficient between $\varepsilon(s_1)$ and $\varepsilon(s_2)$ and $\Theta=(a,b,\Psi^{\mathrm{T}})^{\mathrm{T}}$ be the parameter vector. Then, the bivariate cumulative distribution function of the Tukey max-stable process is (2) with $D=2$ and*

$$V_{s_1,s_2}(z_1,z_2;\Theta)=\frac{1}{z_1}\left[1-G_\Theta\left\{\log\left(\frac{z_1}{z_2}\right)\right\}\right]+\frac{1}{z_2}\left[1-G_\Theta\left\{\log\left(\frac{z_2}{z_1}\right)\right\}\right], \quad (7)$$

where

$$G_\Theta(y)=\int_{-\infty}^{\infty}f_X(t)\Phi\left(\frac{y-t^2}{\sqrt{2}|t|}\right)dt, \quad (8)$$

with $f_X(\cdot)$ the probability density function of $X\sim N\left(\mu_1(a,b,\rho),\varphi^2(a,\rho)\right)$, $\varphi^2(a,\rho)=a^2(1-\rho^2)/\{2(1-a)\}$ and $\mu_1(a,b,\rho)=b(1-a)^{-1}\{(1-\rho)(2-a+\rho a)/2\}^{1/2}$.

The proof is presented in the Appendix.

Remark 1 *A special case of $G_\Theta(y)$ in (8) is when $a\to 0$. We then have that $\varphi(a,\rho)\to 0$. Hence, X degenerates to a random variable with a point mass at $X=\mu_1(0,b,\rho)=b\{1-\rho\}^{1/2}$.*

Thus, by letting $F_X(\cdot)$ be the cumulative distribution function of X, we have as $a \to 0$ that (7) reduces to the bivariate exponent measure of the geometric Gaussian max-stable process since

$$G_\Theta(y) = \int_{-\infty}^{\infty} \Phi\left(\frac{y - t^2}{\sqrt{2}|t|}\right) dF_X(t) \to \Phi\left(\frac{y - b^2(1-\rho)}{b\sqrt{2}(1-\rho)^{1/2}}\right).$$

Remark 2 *In principle, one can generalize $W(s)$ defined in (4) by using a higher-order polynomial function of $\varepsilon(s)$. However, it would be very difficult, if not impossible, to derive a closed form pairwise cumulative distribution function, which makes such a choice of no practical interest. Indeed, the appeal of using a quadratic function of $\varepsilon(s)$ in (4) is that it easily combines with the form of the Gaussian probability density function. In this sense, the proposed $W(s)$ in (4) is an interesting extension of the geometric Gaussian max-stable process with tractable statistical properties.*

2.2 The bivariate extremal coefficient

The second proposition gives an explicit form of the bivariate extremal coefficient of the Tukey max-stable process.

Proposition 2 *The bivariate extremal coefficient of the Tukey max-stable process is*

$$\theta_{TK}(h) = 2\{1 - G_\Theta(0)\} = 2 - 2\left[\Phi(\rho_a\delta_1) + \Phi(\delta_1) - 2\Phi_2\{(\delta_1, \rho_a\delta_1); \rho_a\}\right], \tag{9}$$

where $\delta_1 = a^{-1}b\{(1-\rho)/(1+\rho) + 1/(1-a)\}^{1/2}$, $\rho_a = a[(1-\rho^2)/\{(2-a)^2 - \rho^2a^2\}]^{1/2}$ and $\Phi_2\{(x_1, x_2); \varrho\}$ is the bivariate standard normal cumulative distribution function with correlation coefficient ϱ evaluated at (x_1, x_2).

The proof is presented in the Appendix.

As $\rho_\Psi(h) \to 1$ and hence $\rho_a \to 0$, it follows that $\theta_{TK}(h) \to 2\{1-\Phi(0)\} = 1$. Thus the proposed Tukey max-stable process can capture total dependence. On the other hand, as $\rho_\Psi(h) \to 0$,

$$\theta_{TK}(h) \to \theta_{TK}^0 \equiv 2 - 2\left[\Phi(\rho_a^0\delta_1^0) + \Phi(\delta_1^0) - 2\Phi_2\{(\delta_1^0, \rho_a^0\delta_1^0); \rho_a^0\}\right],$$

where $\delta_1^0 = a^{-1}b\{(2-a)/(1-a)\}^{1/2}$ and $\rho_a^0 = a/(2-a)$. To capture complete independence of the induced max-stable process, we need to ensure that $\theta_{TK}^0 = 2$. A sufficient condition for $\theta_{TK}^0 = 2$ is to ensure that $\delta_1^0 \to \infty$ by letting $b/\{1-a\}^{1/2} \to \infty$. For the geometric Gaussian process, we have to let $\sigma \to \infty$, which is equivalent to the parameter b of the Tukey max-stable process, so that θ_{geo}^0, the upper-bound of $\theta_{geo}(h)$ defined in (3), approaches 2. Thus, the newly introduced parameter, a, "speeds up" (when $a > 0$) or "slows down" (when $a < 0$) the rate at which $\theta_{geo}^0 \to 2$ as $\sigma \to \infty$, as illustrated in the left panel of Figure 1. In Figure 1, we can clearly see that for the Tukey max-stable process, both a and b impact the shape and range of the extremal coefficient. Therefore, the introduction of the parameter a brings in additional flexibility to model complicated dependence structures of spatial extremes.

2.3 Composite likelihood fitting

As for many max-stable processes, given a set of independent data, $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$, with each $\mathcal{Z}_i$ consisting of observations from a spatial domain, $\mathcal{S}$, the joint distribution of each $\mathcal{Z}_i$ is not available. The literature on fitting max-stable processes has therefore largely focused on the composite likelihood approach (Lindsay, 1998; Cox & Reid, 2004; Padoan et al., 2010; Varin et al., 2011; Genton et al., 2011; Huser & Davison, 2013; Sang & Genton, 2014), particularly the pairwise likelihood approach. Specifically, the model parameter set Θ can be estimated by maximizing the composite log-likelihood function defined as follows:

$$l_n(\Theta) = \sum_{i=1}^{n} \sum_{\{j<k: z_j, z_k \in \mathcal{Z}_i\}} \log f(z_j, z_k; \Theta), \tag{10}$$

where $f(z_j, z_k; \Theta)$ is the bivariate density function of $\mathcal{Z}_i(s_j)$ and $\mathcal{Z}_i(s_k)$ for $s_j, s_k \in \mathcal{S}$. We first define two functions

$$g'_\Theta(t; y) = \frac{1}{\sqrt{2}|t|} f_X(t) \phi\left(\frac{y - t^2}{\sqrt{2}|t|}\right), \quad G'_\Theta(y) = \int_{-\infty}^{\infty} g'_\Theta(t; y)\, dt,$$

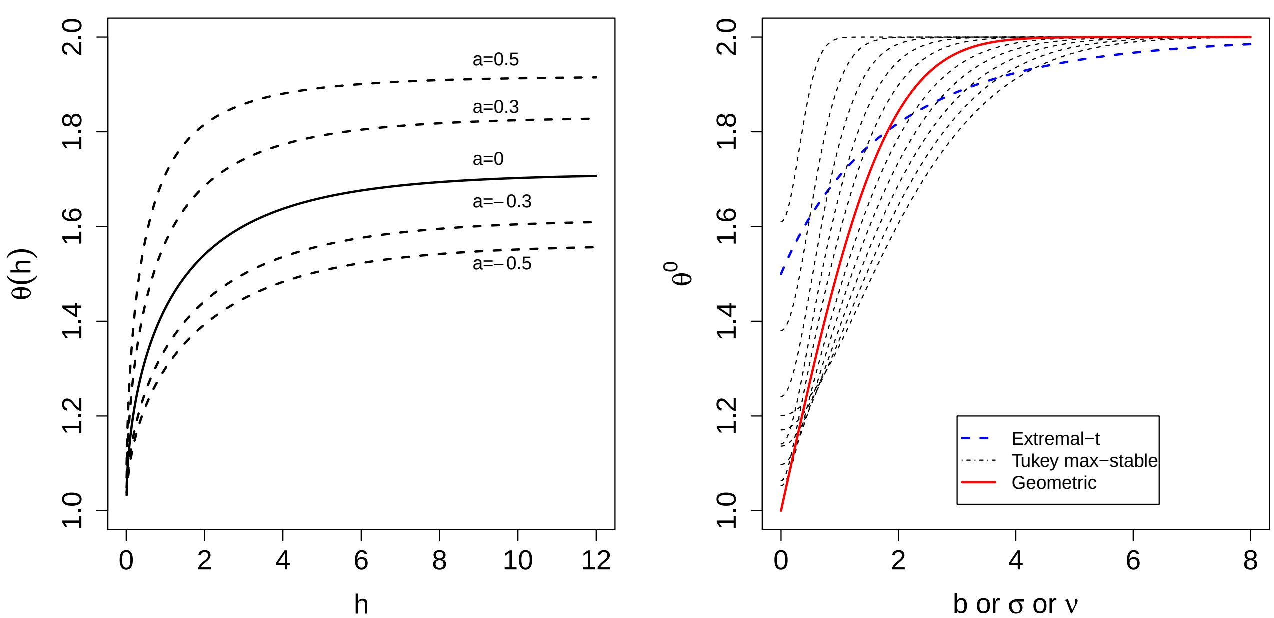

Figure 1: Bivariate extremal coefficients of various max-stable processes with a Matérn correlation function (12) ($\phi = 3$, $\nu = 0.5$). Left: $\theta_{TK}(h)$ for various a's while fixing b =1.5. (Note: $a = 0$ corresponds to the geometric Gaussian process with $\sigma = 1.5$.) Right: θ_t^0 as a function of ν; θ_{TK}^0 for 11 values of a equispaced on $[-0.9, 0.9]$ as a function of b; and θ_{geo}^0 as a function of σ.

where $\phi(\cdot)$ is the standard normal probability density function. Then, using Proposition 1, some straightforward algebra leads to the bivariate density function of the Tukey max-stable process:

$$f(z_1, z_2; \Theta) = \exp\{-V_{s_1,s_2}(z_1, z_2; \Theta)\} \left[\frac{1}{z_1^2 z_2^2}\{1 - G_\Theta(v)\}\{1 - G_\Theta(-v)\} + \frac{1}{z_1^2 z_2} G'_\Theta(v) \right], \qquad (11)$$

where $v = \log(z_1/z_2)$ and $V_{s_1,s_2}(z_1, z_2; \Theta)$ is given by (7).

Although numerical integration has to be used to compute $G_\Theta(y)$ and $G'_\Theta(y)$, the computational cost of maximizing the composite likelihood function (10) is not high. The integration bounds in (11) are from $-\infty$ to ∞, but a closer examination of the integrands $g_\Theta(t; y)$ and $g'_\Theta(t; y)$ reveals that it suffices to set the integration bounds to $\mu_1(a, b, \rho) \pm 10 \times \varphi(a, \rho)$, which can greatly reduce the computational cost due to numerical integration. The reason is that when t is outside this interval, the value of $f_X(t)$ is so small that the integration outside this interval is practically 0. In our simulation studies, this approximation works extremely well. Another potential computational issue is that when $y = 0$, $g'_\Theta(t; y)$ may not be integrable, depending on the values of $\mu_1(a, b, \rho)$ and $\varphi(a, \rho)$. For this reason, to ensure maximal numerical stability when computing $l_n(\Theta)$, we recommend removing all pairs (z_j, z_k) with $|\log z_j - \log z_k| < \tau_0$ for some $\tau_0 > 0$. Since the max-stable process $Z(s)$ defined in (1) has a unit Fréchet marginal distribution, which is a continuous distribution, one can expect that as long as there is no very strong spatial dependence in the max-stable random field, the probability of having a pair with $|\log z(s_1) - \log z(s_2)| > \tau_0$ is negligible for a sufficiently small τ_0. In our simulations, we choose $\tau_0 = 10^{-6}$, which seems to work well. With such a choice, in almost all cases, less than 1% of pairs were removed. In the Swiss rainfall data, no pairs of observations were removed with

$\tau_0 = 10^{-6}$. However, when the spatial dependence is very strong, one needs to be more cautious about the choice of τ_0.

2.4 Simulation of Tukey max-stable process

Computer simulation has been an important tool to study mathematically intractable properties of max-stable processes (Oesting et al., 2015). The representation given in (1) provides a convenient way to simulate a max-stable process by taking the maximum over an infinite number of replications from the processes R_j and $W_j(s)$. However, in practice, only a finite number of replications can be taken and as a result the simulated max-stable process may not be exact (Ribatet, 2013; Oesting et al., 2012, 2015). Schlather (2002) proposed a simple approach to obtain exact simulations of a max-stable process defined in (1) provided that the latent process $W(s)$ is uniformly bounded by some constant $C > 0$, that is $\sup_{s\in\mathbb{R}^d} W(s) \leq C$ almost surely. We adopt a simpler approach outlined in Oesting et al. (2015) (Proposition 1.2.1) to simulate exact samples from such a max-stable process over a set $\mathcal{S} \subset \mathbb{R}^d$ as follows:

Algorithm I: Simulate a max-stable process $Z(s)$ in (1) with $\sup_{s\in\mathbb{R}^d} W(s) \leq C$.

1. Step I: Set $Z(s) = 0$ for all $s \in \mathcal{S}$, $E = 0$ and $i = 1$;
2. Step II: Generate a random number $\xi_i \sim$ Exponential(1) and a random realization of $\{W_i(s), s \in \mathcal{S}\}$. Then update $E = E + \xi_i$ and $Z(s)$ by $\max\{Z(s), W_i(s)/E\}$;
3. Repeat Step II until $\inf_{s\in\mathcal{S}} Z(s) \leq C/E$.

Most popular max-stable models constructed using representation (1), including the geometric Gaussian process, the extremal-Gaussian process, the extremal-t process and the Brown–Resnick process, do not meet the condition $\sup_{s\in\mathbb{R}^d} W(s) \leq C$. As a remedy, Schlather (2002) proposed to use a large enough constant C^* such that $\mathrm{pr}\left\{\sup_{s\in\mathbb{R}^d} W(s) > C^*\right\}$ is sufficiently small. Although there has been some alternative proposals for obtaining exact samples from such max-stable processes (Oesting et al., 2012, 2015; Dombry et al., 2016), Algorithm I remains attractive due to its simplicity.

A distinctive feature of the proposed Tukey max-stable process is that when $a < 0$, the spectral process $W(s)$ as defined in (4) is uniformly bounded above by $C = \exp\left\{\frac{1}{2}\log(1-a) - \frac{b^2}{2a(1-a)}\right\}$. Therefore, the simulation of a Tukey max-stable process with $a < 0$ using Algorithm I can be exact based on the theoretical results of Schlather (2002). On the other hand, if we restrict $0 < a < 1$, we can still use Algorithm I with a large value C^* to obtain approximate samples. In all our numerical examples, we choose $C^* = \exp\left\{\frac{5^2}{2}a + 5b + \frac{1}{2}\log(1-a) - \frac{b^2}{2(1-a)}\right\}$ obtained by plugging $\varepsilon(s) = 5$ in (4), which works sufficiently well in almost all cases.

3 Monte Carlo simulations

We conduct a simulation study to evaluate the estimation accuracy of the pairwise likelihood estimator obtained by maximizing (11). In each simulation run, $D = 40$ spatial locations were randomly chosen from 225 grid points located on a 15×15 regular grid over the region $[0, u]\times[0, u]$, where u is chosen as the smallest h such that $\theta_{TK}(h) \geq \theta^0_{TK} - 0.01$ for a given combination of (a, b). The spatial domain is chosen in this way to ensure that the $D = 40$ spatial locations are spread out over the region with a varying extremal coefficient function $\theta_{TK}(h)$. The latent standard Gaussian process, $\varepsilon(s)$, in (4) has the Matérn correlation function

$$\rho_\Psi(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{\phi}\right)^{\nu} K_\nu\left(\frac{h}{\phi}\right), \tag{12}$$

where ν is the smoothness parameter, ϕ is the range parameter, $\Gamma(\cdot)$ is the gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. In this study, we set $\phi = 3$, $\nu = 0.3$ and experiment with multiple values of a and b when generating data from the Tukey max-stable process. The number of independent replicates of the max-stable process for the same spatial locations varies from $n = 20$ to $n = 100$. The smoothness parameter, $\nu = 0.3$, is assumed to be known, a common practice, and the other parameters, a, b, ϕ, must be estimated. To find the global maximum of (11), we used multiple starting points for the maximization routine. The simulation of the Tukey max-stable process is similar to the default method for generating max-stable processes in the R package `SpatialExtremes` (Ribatet et al., 2015), following Algorithm I in Subsection 2.4.

The empirical bias and root mean-squared error based on 200 simulation runs are summarized in Table 1, where we can see that when n increases from 20 to 100, the root mean-squared error consistently decreases, indicating the effectiveness of the pairwise likelihood approach. However, when n is small, we can observe a certain amount of bias for estimators of the range parameter, ϕ, which decreases as n increases. Furthermore, the magnitude of the bias increases as u decreases, which is expected since the data are concentrated in a smaller region and contain less information about the range parameter, ϕ.

A second observation is that when a and b are large, there is also some appreciable biases in $\hat{b}$.

One possible explanation is that in such cases, as we can see from Figure 2, the extremal coefficient function $\theta_{TK}(h)$ quickly reaches its upper bound θ^0_{TK}. For example, in the case of $a = 0.5$ and $b = 1.5$, $\theta_{TK}(h)$ stabilizes at the distance around $h = 2$. Recall that in this case, the 40 spatial locations are evenly distributed in the square $[0, 5.62] \times [0, 5.62]$. It indicates that most pairs of locations do not carry effective information about the variation of $\theta_{TK}(h)$, which is controlled by a and b. As a result, the pairwise likelihood function, $l_n(\Theta)$, may be quite flat in the neighborhood of the true value of Θ. However, even in such cases, the pairwise dependence structure can still be correctly captured even though the estimate $\widehat{\Theta}$ may be different from the true underlying value.

Table 1: Estimation accuracies of pairwise likelihoods for the Tukey max-stable process.

				$\hat{a}$	$\hat{b}$	$\hat{\phi}$
b	a	u	n	Bias (RMSE)	Bias (RMSE)	Bias (RMSE)
1	−0.5	7.42	20	−0.007 (0.143)	0.063 (0.257)	0.803 (2.651)
			50	−0.001 (0.074)	0.013 (0.122)	0.163 (0.957)
			100	−0.000 (0.055)	0.013 (0.088)	0.157 (0.659)
	−0.3	7.52	20	0.023 (0.108)	0.021 (0.179)	0.585 (2.329)
			50	0.015 (0.068)	0.007 (0.099)	0.210 (1.025)
			100	0.003 (0.045)	0.008 (0.068)	0.113 (0.616)
	0.3	7.22	20	−0.058 (0.154)	0.092 (0.245)	0.746 (2.906)
			50	−0.026 (0.102)	0.049 (0.150)	0.527 (2.323)
			100	−0.016 (0.064)	0.018 (0.076)	0.084 (0.802)
	0.5	6.62	20	−0.085 (0.204)	0.203 (0.563)	0.945 (3.030)
			50	−0.029 (0.113)	0.073 (0.178)	0.772 (2.338)
			100	−0.016 (0.078)	0.045 (0.129)	0.494 (1.844)
1.5	−0.5	8.12	20	0.011 (0.370)	-0.024 (0.310)	0.188 (1.454)
			50	0.018 (0.115)	−0.004 (0.208)	0.100 (0.972)
			100	0.011 (0.100)	−0.002 (0.135)	0.048 (0.611)
	−0.3	8.02	20	0.022 (0.205)	0.035 (0.376)	0.712 (2.649)
			50	0.023 (0.133)	0.006 (0.185)	0.324 (1.070)
			100	0.004 (0.093)	0.017 (0.139)	0.215 (0.716)
	0.3	6.92	20	−0.074 (0.215)	0.165 (0.462)	0.978 (3.360)
			50	−0.061 (0.187)	0.124 (0.373)	0.474 (2.002)
			100	−0.046 (0.153)	0.093 (0.265)	0.345 (1.799)
	0.5	5.62	20	−0.055 (0.287)	0.428 (0.900)	1.253 (4.161)
			50	−0.078 (0.236)	0.241 (0.660)	1.233 (3.874)
			100	−0.031 (0.177)	0.135 (0.469)	1.281 (3.229)

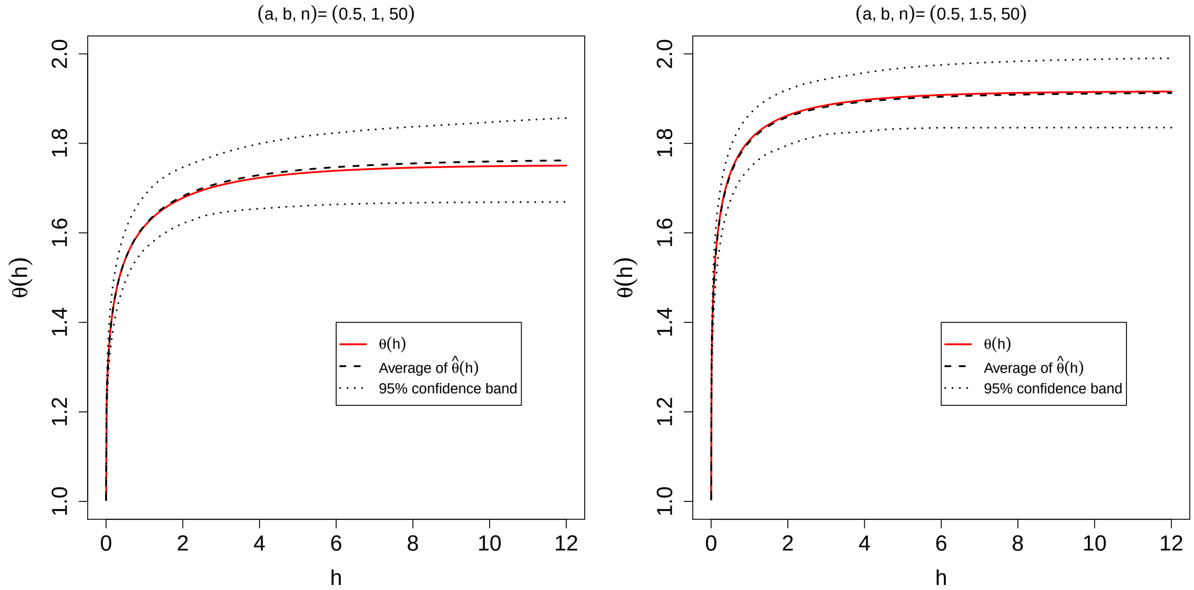

Figure 2: Estimation accuracy of the extremal coefficient function for the Tukey max-stable process for $n = 50$.

In Figure 2, we can see that the estimated extremal coefficient, $\hat{\theta}_{TK}(h)$, is unbiased for $\theta_{TK}(h)$ in the cases of $(a, b) = (0.5, 1)$ and $(a, b) = (0.5, 1.5)$ when $n = 50$, and it holds more generally for the other combinations of a and b. This problem may be resolved by using a triplewise likelihood of the Tukey max-stable process. For example, it was suggested by Castruccio et al. (2016) that the bias of estimators obtained through maximizing pairwise likelihood for the Brown–Resnick process can be eliminated by using a triplewise likelihood. Considering the similarity between (4) and the Brown–Resnick process, whether this bias-correction phenomenon also applies to the Tukey max-stable process is left for future research.

Finally, we want to point out that the simulation results in Table 1 assume the knowledge of whether $a < 0$ or $a > 0$. That is, if the true value $a < 0$, we restrict a to be negative in our optimization routine. We empirically observe that for a parameter vector Θ with an $a < 0$, there often exist another Θ' with an $a' > 0$ that gives a pairwise likelihood $\ell(\Theta') \approx \ell(\Theta)$. Therefore, considering the benefits for simulation described in Section 2.4, we would recommend using $a < 0$ in practice.

4 Swiss rainfall data

We apply the Tukey max-stable process to summer maximum daily rainfall data available from the R package `SpatialExtremes` (Ribatet et al., 2015) for the years 1962-2008 at 79 weather stations in the Plateau region of Switzerland. Since our focus is on modelling the dependence of the maxima, we first fit the raw data with the generalized extreme value distribution model as suggested by Davison et al. (2012), assuming that all data are independent and then we transform the raw data to the unit Fréchet scale using the estimated parameters. The R code can be found in the Appendix. We fit three max-stable models to the transformed data: the Tukey, geometric Gaussian and extremal-t max-stable processes. For the latent standard Gaussian process, we used the Matérn correlation function (12) with a fixed range parameter, $\phi = 700$, following suggestions by Davison et al. (2012). We used the composite likelihood information criterion (Varin & Vidoni, 2005, CLIC) to compare the three fitted models. For comparison, we also fit the Brown–Resnick process based on a fractional variogram (Brown & Resnick, 1977; Kabluchko et al., 2009). The geometric Gaussian, extremal-t and Brown–Resnick processes were fitted using *fitmaxstab()* function and the reported CLIC values were computed by the *TIC()* function of the `SpatialExtremes` package. The CLIC values for the Tukey max-stable process were computed following suggestions of Ribatet (2009), Section 4.2.2 and Section 5.1. For numerical stability, we used $\tau = 10^{-6}$ as suggested in Section 2.3 but no pairs of observations were removed.

From Table 2, we can see that the Tukey max-stable process with an $a < 0$ gives the smallest CLIC value and the Tukey max-stable process with an $a > 0$ gives the second smallest value. Both of them seem to outperform the geometric Gaussian and Brown–Resnick processes, which is not surprising. Compared to the extremal-t process, the differences in CLIC values are much smaller and it is hard to tell which one is better. Figure 3(a) shows the fitted extremal coefficient functions for the five models, which are all similar to each other and approximate the empirical extremal coefficients (Schlather & Tawn, 2003) reasonably well. For comparison, we also plot the local polynomial regression fitting (Loess) of the empirical extremal coefficients, which seems to be quite off.

Table 2: Estimation results for five max-stable models on the Swiss rainfall data.

Model	(a, b)	DoF	σ^2	ϕ (km)	ν	$l_n(\Theta)$	CLIC
Tukey max-stable ($a > 0$)	(0.73, 1.07)	n/a	n/a	700	0.39	−600,626.5	1,201,947
Tukey max-stable ($a < 0$)	(−2.22, 10.20)	n/a	n/a	700	0.34	−600,595.8	1,201,884
Geometric Gaussian	n/a	n/a	10.89	700	0.33	−601,663.5	1,203,943
Extremal-t	n/a	6.43	n/a	700	0.28	−600,689.5	1,201,997
Brown–Resnick	n/a	n/a	n/a	20.65	0.66	−601,660.3	1,203,938

Although the extremal coefficients appear to be very similar for all five fitted models, they may have different properties. Figure 4 depicts realizations of four fitted max-stable processes. To make the four realizations comparable, Figure 4 was generated in the following way: 1) Generate a realization from the Tukey max-stable process with $(a, b) = (-2.22, 10.20)$ on a 50×50 fine grid; 2) Choose an evenly spaced 17×17 sub-grid from the 50×50 fine grid; 3) Simulate $1,000$ realizations from the other three fitted max-stable models and find the realization that gives the smallest average absolute difference on the chosen 17×17 to the realization obtained in step 1). The average absolute differences between Figure 4(b)-(d) and Figure 1(a) on these 17×17

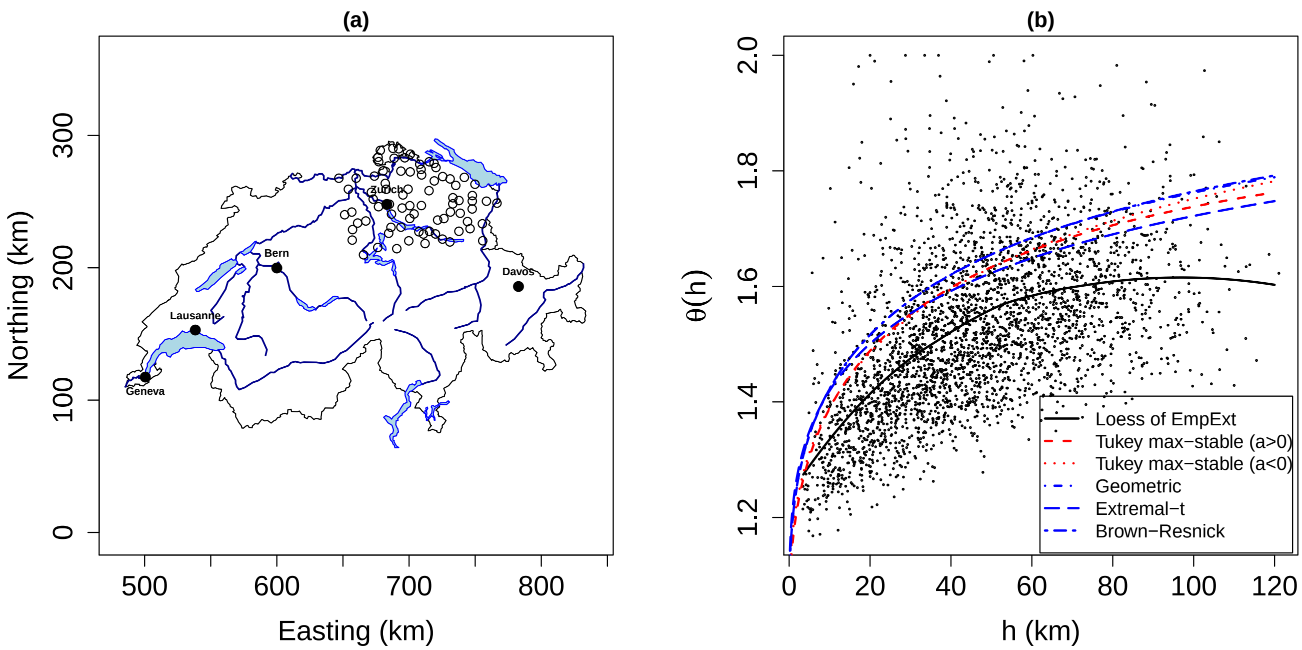

Figure 3: Swiss rainfall data: (a) Spatial locations of 79 weather stations; (b) Estimated extremal coefficients.

locations are 6.75, 6.99, and 7.00, respectively. As we can see, realizations in Figure 4 have some similar overall patterns but also with appreciable differences in sample paths.

5 Discussion

In this paper, we have proposed a new max-stable process, coined Tukey max-stable process, that brings additional flexibility for modeling spatial extremes. We have derived its bivariate distribution and bivariate extremal coefficient. We also proposed pairwise composite likelihood to fit this new model. Results of Monte Carlo simulations have demonstrated the effectiveness of the fitting procedure. We have illustrated our new model on rainfall data from Switzerland.

Despite its flexibility, this new Tukey max-stable process does have some limitations that need to be addressed in the future. The first challenge is how to reduce the computational cost due to the numerical integration. This is especially important if one wishes to use high-order composite

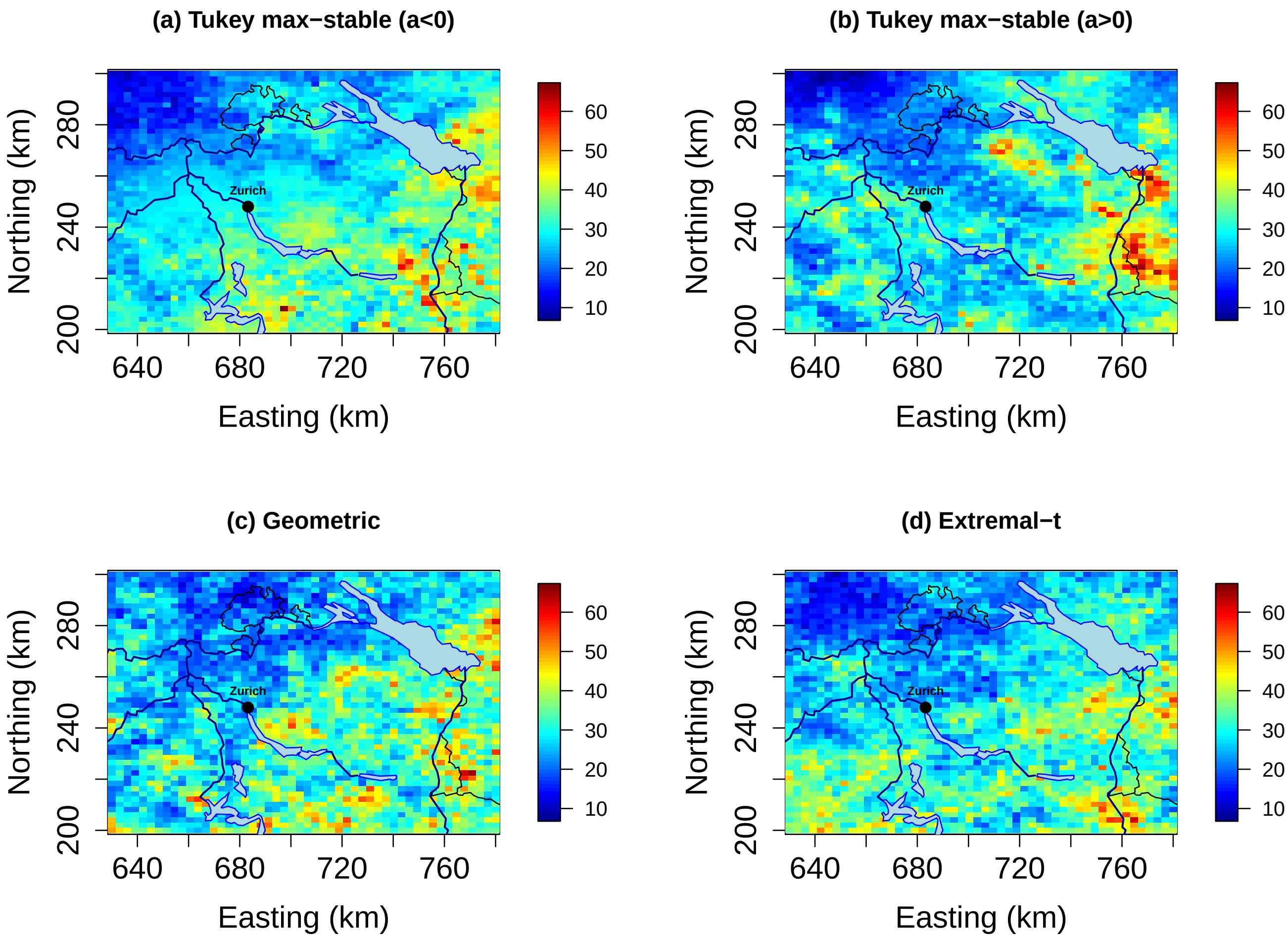

Figure 4: Realizations of four fitted max-stable processes to the Swiss rainfall data.

likelihood to obtain more efficient estimator as in Huser & Davison (2013) and Castruccio et al. (2016). A second challenge is to develop conditional simulation techniques for the Tukey max-stable process. Conditional simulation is of particular importance if one wants to make accurate predictions for spatial extremes at locations without observations, which has already been investigated for some max-stable processes (Dombry et al., 2013). Another challenge is to incorporate non-stationarity structures into the Tukey max-stable process. One possible route is to combine our model with covariates along the lines proposed by Huser & Genton (2016).

Appendix

Proof of Proposition 1: From the representation (1) and equation (2), we have

$$V_{s_1,s_2}(z_1,z_2) = \frac{1}{z_1}\mathrm{E}\left[W(s_1)I\left\{\frac{W(s_1)}{z_1} > \frac{W(s_2)}{z_2}\right\}\right] + \frac{1}{z_2}\mathrm{E}\left[W(s_2)I\left\{\frac{W(s_2)}{z_2} > \frac{W(s_1)}{z_1}\right\}\right].$$

Using the definition of $W(s)$ in (4), for a pair of locations s_1, s_2, denote $\xi_j = \varepsilon(s_j)$ with $\varepsilon(s_j)$ as defined in $W(s_j)$, $j = 1, 2$. Some straightforward calculus gives that

$$\mathrm{E}\left[W(s_1)I\left\{\frac{W(s_1)}{z_1} > \frac{W(s_2)}{z_2}\right\}\right] = \frac{(1-a)^{1/2}}{2\pi(1-\rho^2)^{1/2}}\int_A \exp\left\{-\frac{1}{2}(\xi-\mu)^{\mathrm{T}}\Sigma^{-1}(\xi-\mu)\right\}d\xi_1 d\xi_2,$$

where the integration region $A = \left\{(\xi_1,\xi_2) : \left[\frac{a}{2}(\xi_1+\xi_2)+b\right](\xi_1-\xi_2) > \log\frac{z_1}{z_2}\right\}$ and

$$\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \frac{b}{1-a}\begin{pmatrix}1\\ \rho\end{pmatrix}, \quad \Sigma = \frac{1}{1-a}\begin{pmatrix}1 & \rho\\ \rho & 1-a+\rho^2 a\end{pmatrix}.$$

Notice that $|\Sigma| = (1-\rho^2)/(1-a)$, we have that

$$\mathrm{E}\left[W(s_1)I\left\{\frac{W(s_1)}{z_1} > \frac{W(s_2)}{z_2}\right\}\right] = \mathrm{pr}\left[\left\{\frac{a}{2}(\xi_1+\xi_2)+b\right\}(\xi_1-\xi_2) > \log\frac{z_1}{z_2}\right],$$

with $\xi = (\xi_1,\xi_2)^{\mathrm{T}} \sim N(\mu,\Sigma)$.

Case I: $(a = 0)$. When $a = 0$,

$$\mathrm{E}\left[W(s_1)I\left\{\frac{W(s_1)}{z_1} > \frac{W(s_2)}{z_2}\right\}\right] = \mathrm{pr}\left\{\xi_1-\xi_2 > \frac{1}{b}\log\frac{z_1}{z_2}\right\} = 1-\Phi\left\{\frac{\log\frac{z_1}{z_2}-(1-\rho)b^2}{b\sqrt{2}(1-\rho)^{1/2}}\right\}.$$

Case II: $a \neq 0$ and $a < 1$. Let $U_1 = a(\xi_1+\xi_2)/2+b$ and $U_2 = \xi_1-\xi_2$ with

$$\nu_1 = \mathrm{E}(U_1) = \frac{2-a+\rho a}{2(1-a)}b, \quad \sigma_1^2 = \mathrm{var}(U_1) = \frac{(1+\rho)a^2}{4(1-a)}(2-a+\rho a),$$

$$\nu_2 = \mathrm{E}(U_2) = \frac{1-\rho}{1-a}b, \quad \sigma_2^2 = \mathrm{var}(U_2) = \frac{1-\rho}{1-a}(2-a-\rho a),$$

$$\mathrm{cov}(U_1, U_2) = \frac{a}{2}\{\mathrm{var}(\xi_1) - \mathrm{var}(\xi_2)\} = \frac{a^2(1-\rho^2)}{2(1-a)} \neq 0, \quad \rho_a = \frac{|a|(1-\rho^2)^{1/2}}{\{(2-a)^2 - \rho^2 a^2\}^{1/2}},$$

$$\delta_1 = \frac{\nu_1}{\sigma_1} = \frac{b}{|a|}\left\{\frac{2-a+\rho a}{(1+\rho)(1-a)}\right\}^{1/2}, \quad \delta_2 = \frac{\nu_2}{\sigma_2} = b\left\{\frac{1-\rho}{(1-a)(2-a-\rho a)}\right\}^{1/2}.$$

The cumulative distribution of $Y = U_1U_2/(\sigma_1\sigma_2)$ is derived in Meeker & Escobar (2007) as

$$G_\Theta^*(y) = \int_{-\infty}^{\infty} \phi(t-\delta_1)\Phi\left\{\mathrm{sign}(t)\frac{y/t - \delta_2 - \rho_a(t-\delta_1)}{(1-\rho_a^2)^{1/2}}\right\} dt,$$

which can be simplified with the equality $\delta_2 = \rho_a\delta_1$. Define a new function

$$G_\Theta(y) = G_\Theta^*\left(\frac{y}{\sigma_1\sigma_2}\right) = \int_{-\infty}^{\infty} \phi(t-\delta_1)\Phi\left\{\mathrm{sign}(t)\frac{y/t - \varphi^2(a,\rho)t}{\sqrt{2}\varphi(a,\rho)}\right\} dt = \int_{-\infty}^{\infty} f_X(t)\Phi\left(\frac{y-t^2}{\sqrt{2}|t|}\right) dt,$$

where $\varphi(a,\rho) = |a|\sqrt{\frac{1-\rho^2}{2(1-a)}}$. Following the customary definitions $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$ and $\mathrm{sign}(0) = 1$, the above integrand is continuous for any $a < 0$ and $0 < a < 1$. Then, we have

$$\mathrm{E}\left[W(s_1)I\left\{\frac{W(s_1)}{z_1} > \frac{W(s_2)}{z_2}\right\}\right] = \mathrm{pr}\left(Y > \sigma_1\sigma_2 \log\frac{z_1}{z_2}\right) = 1 - G_\Theta\left(\log\frac{z_1}{z_2}\right). \tag{13}$$

Similarly, we can derive the second part of $V_{s_1,s_2}(z_1, z_2)$ which, combined with (13), gives (7). □

Proof of Proposition 2: By the definition of $\theta_{TK}(h) = 2\{1 - G_\Theta(0)\}$, we have that

$$G_\Theta(0) = G_\Theta^*(0) = \int_{-\infty}^{0} \phi(t-\delta_1)\Phi\left\{\frac{\rho_a t}{(1-\rho_a^2)^{1/2}}\right\} dt + \int_{-\infty}^{0} \phi(t+\delta_1)\Phi\left\{\frac{\rho_a t}{(1-\rho_a^2)^{1/2}}\right\} dt$$
$$= \Phi_2\{(-\delta_1, \rho_a\delta_1); -\rho_a\} + \Phi_2\{(\delta_1, -\rho_a\delta_1); -\rho_a\},$$

where the last equation follows from the cumulative distribution function of the extended skew-t distribution with $v \to \infty$ (Arellano-Valle & Genton, 2010). From the properties of $\Phi_2\{(\cdot,\cdot);\cdot\}$, we have $G_\Theta(0) = \Phi(\rho_a\delta_1) + \Phi(\delta_1) - 2\Phi_2\{(\delta_1, \rho_a\delta_1); \rho_a\}$, which leads to (9). □

Swiss rainfall data: R code

In this subsection, we only provide R code for max-stable processes other than the Tukey max-stable process. R code to fit the Tukey max-stable model is available from the authors.

```
###Warning: SpatialExtremes_2.0-0 and SpatialExtreme_2.0-2 produce
### slightly different CLIC values
library("SpatialExtremes")
data(rainfall); data <- rain;
##Transform the data into unit frechet distribution##
form.loc <- loc ~ lat + lon
form.scale <- scale ~ lat + lon
form.shape <- shape ~ 1
obj.fit <- fitspatgev(data, coord[,1:2], form.loc, form.scale, form.shape)
fit.para <- predict(obj.fit)
K <- dim(data)[2]; n <- dim(data)[1]; fret<-data*0
for(k in 1:K )
fret[,k] <- gev2frech(data[,k], loc=fit.para[k,3], scale=fit.para[k,4],
 shape=fit.para[k,5], emp = FALSE)
#################fitting the Geometric process#####
obj.geo <- fitmaxstab(fret,coord[,1:2],cov.mod="gwhitmat",nugget=0,range=700)
TIC(obj.geo)
#################fitting the extremal-t process###
obj.ext <- fitmaxstab(fret,coord[,1:2],cov.mod="twhitmat",nugget=0,range=700)
TIC(obj.ext)
##Fit the Brown-Resnick process ###
obj.br <- fitmaxstab(fret,coord[,1:2],cov.mod="brown")
TIC(obj.br)
```

References

Arellano-Valle, R. B. & Genton, M. G. (2010). Multivariate extended skew-t distributions and related families. *Metron* **68**, 201–34.

Brown, B. M. & Resnick, S. I. (1977). A spectral representation for max-stable processes. *J. Appl. Prob.* **14**, 732–9.

Castruccio, S., Huser, R. & Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Stat.* to appear.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo B. O. & Sun, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *REVSTAT Statist. J.* **10**, 135–65.

Cox, D. R. & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–37.

Dombry, C., Éyi-Minko, F. & Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100**, 111–124 .

Davison, A. C. & Gholamrezaee, M. (2012). Geostatistics of extremes. *P. Roy. Soc. A-Math. Phy.* **468**, 581–608.

Davison, A. C. & Huser, R. (2015). Statistics of Extremes. *Annual Rev. Statist. Appl.* **2**, 203–35.

Davison, A. C., Padoan, S. A. & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27**, 161–86.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–204.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* New York: Springer.

de Haan, L. & Zhou, C. (2011). Extreme residual dependence for random vectors and processes. *Adv. Appl. Prob.* **43**, 217–42.

Dombry, C., Engelke, S. & Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103**, 303–17.

Genton, M. G., Ma, Y. & Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–8.

Huser, R. & Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**, 511–8.

Huser, R. & Davison, A. C. (2014). Space-time modelling of extreme events. *J. R. Statist. Soc.* B **76**, 439–61.

Huser, R. & Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *J. Agric. Biol. Env. Stat.* **21**, 470–91.

Kabluchko, Z., Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Prob.* **37**, 2042–65.

Lindsay, B. G. (1998). Composite likelihood methods. *Contemp. Math.* **80**, 221–39.

MacGillivray, H. L. (1992). Shape properties of the g-and-h and Johnson families. *Commun. Stat. Theo. & Meth.* **21**, 1233–50.

Martinez, J. & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Commun. Stat. Theo. & Meth.* **13**, 353–69.

Meeker, W. Q. & Escobar, L. A. (2007). An algorithm to compute the cdf of the product of two normal random variables. *Commun. Stat. B-Simul.* **23**, 271–80.

Nikoloulopoulos, A. K., Joe, H. & Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12**, 129–48.

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122**, 409–13.

Oesting, M., Ribatet, M. & Dombry, C. (2015). Simulation of Max-Stable Processes. In *Extreme Value Modeling and Risk Analysis*, Dipak K. Dey and Jun Yan (eds) Chapman and Hall/CRC, Chapter 10: 195–214.

Oesting, M., Kabluchk, Z. & Schlather, M. (2012). Simulation of Brown–Resnick processes. *Extremes* **15**, 89–107.

Padoan, S. A. (2013). Extreme dependence models based on event magnitude. *J. Multivariate Anal.* **122**, 1–19.

Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Am. Statist. Assoc.* **105**, 263–77.

Penrose, M. D. (1992). Semi-min-stable processes. *Ann. Prob.* **20**, 1450–63.

Ribatet, M. (2009). *A User's Guide to the SpatialExtremes Package.* EPFL, Lausanne, Switzerland.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *J. Soc. Française Stat.* **154**, 156–77.

Ribatet, M., Singleton, R., & R Core Team (2015). SpatialExtremes: Modelling Spatial Extremes. *URL https://CRAN.R-project.org/package=SpatialExtremes.* R package version 2.0-2.

Sang, H. & Genton, M. G. (2014). Tapered composite likelihood for spatial max-stable models. *Spat. Stat.* **8**, 86–103.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**, 33–44.

Schlather, M. & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**, 139–54.

Tukey, J. W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, MA.

Varin, C., Reid, N. & Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5–42.

Varin, C. & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–28.

Wadsworth, J. L. & Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika* **99**, 253–72.

Xu, G. & Genton, M. G. (2015). Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution. *Comput. Stat. Data Anal.* **91**, 78–91.

Xu, G. & Genton, M. G. (2016). Tukey g-and-h random fields. *J. Am. Statist. Assoc.*, in press.