Non-Stationary Dependence Structures for Spatial Extremes

Raphaël Huser\textsuperscript{1}, Marc G. Genton\textsuperscript{1}

November 13, 2014

Abstract

Max-stable processes are natural models for spatial extremes because they provide suitable asymptotic approximations to the distribution of maxima of random fields. In the recent past, several parametric families of stationary max-stable models have been developed, and fitted to various types of data. However, a recurrent problem is the modeling of non-stationarity. In this paper, we develop non-stationary max-stable dependence structures in which covariates can be easily incorporated. Inference is performed using pairwise likelihoods, and its performance is assessed by an extensive simulation study based on a non-stationary locally isotropic extremal $t$ model. Evidence that unknown parameters are well estimated is provided, and estimation of spatial return level curves is discussed. The methodology is demonstrated with temperature maxima recorded over a complex topography. Models are shown to satisfactorily capture extremal dependence.

Keywords: anisotropy; covariate; extremal $t$ model; extreme event; max-stable process; non-stationarity; pairwise likelihood; Smith model.

\textsuperscript{1}CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mails: raphael.huser@kaust.edu.sa, marc.genton@kaust.edu.sa
1 Introduction

Max-stable processes have drawn attention in the recent past, by providing an asymptotically justified framework for modelling spatial extremes, and allowing extrapolation beyond observed data (see, e.g., Davison et al. 2012). Although max-stable processes cannot be characterized by a parametric family, the canonical approach is to fit flexible parametric max-stable models. However, in practice, strong constraints are usually imposed: the max-stable models applied up to now are usually stationary (i.e., shift-invariant) and isotropic (i.e., rotation-invariant). While it is relatively straightforward to construct non-stationary models for marginal distributions, e.g., by letting the underlying parameters depend on covariates or splines (Chavez-Demoulin and Davison 2005; Cooley et al. 2007; Northrop and Jonathan 2011; Davison and Gholamrezaei 2012), it is more difficult to model non-stationarity in the dependence structure. Furthermore, even if a suitable family of non-stationary models can be identified, performing inference may be awkward if the dataset is not spatially rich enough. Since rare events are scarce by nature, it is even more tricky to detect non-stationary patterns at extreme levels, and there have been very few attempts to tackle this important issue so far. A related problem is the incorporation of substantive knowledge, e.g., from physical processes, into max-stable processes. In particular, information might be gained by including meaningful covariates in the dependence structure.

In an analysis of extreme snow depths, Blanchet and Davison (2011) proposed splitting the region of study into distinct homogeneous climatic zones to which stationary models were fitted separately, and where anisotropy was dealt with simple geometric deformations of the space. Although their approach simplifies the problem at first sight, it yields a physically unrealistic description of extreme events at the boundary between zones, while the number of parameters also increases dramatically. Another solution advocated by Cooley et al. (2007) is to map the original latitude-longitude space to an alternative “climate space” in which stationarity may be a reasonable assumption, but this might lead to unrealistic realizations and conclusions in the original space. Alternatively, Smith and Stephenson (2009) and Reich
and Shaby (2012) proposed Bayesian non-stationary max-stable models. These models are, however, intrinsically linked to the stationary Smith (1990) model, which is built from very smooth storm profiles and therefore lacks flexibility. In practice, more flexible models are needed.

In the classical geostatistics literature, several non-stationary models have been suggested. Paciorek and Schervish (2006) proposed a large family of non-stationary correlation functions based on Gaussian kernel convolutions, which can be constructed from known stationary isotropic models. Nychka et al. (2002) built flexible non-stationary covariance functions using multi-resolution wavelets. Fuentes (2001) and Reich et al. (2011) created non-stationary models by mixing stationary covariance functions and letting the weights depend on covariates. Jun and Stein (2007, 2008), Castruccio and Stein (2013) and Castruccio and Genton (2014) advocated a spectral approach that provides flexible non-stationary covariance models on the sphere. Alternatively, Sampson and Guttorp (1992), Perrin and Monestiez (1999), Schmidt and O'Hagan (2003) and Anderes and Stein (2008) created non-stationary processes by smooth deformations of isotropic random fields. Lindgren et al. (2011) developed non-stationary models for Gaussian random fields and Gaussian Markov random fields based on stochastic partial differential equations (SPDEs). Another popular approach to induce spatial non-stationarity is to use latent processes or random effects; see Paciorek and Schervish (2006) for non-stationary kriging of annual precipitation data, and Cooley et al. (2007), Sang and Gelfand (2009) and Sang and Gelfand (2010) for applications to extreme value theory. These hierarchical models are usually embedded into a Bayesian framework and fitted using Markov chain Monte Carlo techniques, but they are difficult to use when max-stable processes are involved (Ribatet et al. 2012).

The present paper aims at merging ideas from extreme-value theory and classical geostatistics by providing frequentist models able to capture non-stationary patterns in spatial extremes through covariates. To this end, a flexible approach based on max-stable processes and Paciorek and Schervish’s correlation model is advocated. Loosely speaking, the new
models proposed here are formed by a first layer justified for extremes, within which non-
stationarity is handled with locally elliptical kernels, and by a second layer, where these
kernels are further described using covariates. As will be explained below, these models
can also be seen locally as smoothly deformed isotropic max-stable random fields. Use of
mixtures is advocated to capture different smoothness behaviors in distinct regions of the
space.

The full likelihood for max-stable processes is intractable when the number of spatial lo-
cations is large (typically greater than 10), and for some models, the joint density can only be
computed for dimension $D = 2$. This explains why pairwise likelihoods (Lindsay 1988 Varin
et al. 2011) have become the standard tool for inference in this context (Padoan et al. 2010
Thibaud et al. 2013 Huser and Davison 2014), although more efficient approaches based
on the point process characterization of extremes have recently been proposed (Wadsworth
and Tawn 2014 Thibaud and Opitz 2014 Engelke et al. 2014).

In Section 2, max-stable processes are introduced and some properties and limitations
of the Smith–Stephenson model are discussed. In Section 3 we propose new non-stationary
max-stable models and explain their link with the Smith–Stephenson model. In Section 4 we
discuss inference based on pairwise likelihoods and conduct a simulation study to investigate
the ability of the estimators to capture nonstationarity in the dependence structure. We also
investigate the effect of ignoring non-stationarity on the estimation of spatial return levels.
In Section 5 we illustrate the methods on temperature annual maxima recorded in Colorado
over the period 1895-1997, and we conclude with a discussion in Section 6.

## 2 Max-stable processes

### 2.1 Theoretical foundations

Suppose that $X_1, X_2, \ldots$, are independent and identically distributed random variables with
distribution $F(x)$. If there exist normalizing sequences $a_n > 0$ and $b_n$ such that the renormal-
ized maximum $a_n^{-1}\{\max(X_1, \ldots, X_n) - b_n\}$ converges to a non-degenerate variable $Z \sim G(z)$,
as \( n \to \infty \), then the limiting distribution \( G(z) \) must be generalized extreme-value (GEV), or equivalently max-stable, i.e.,

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

where \( a_+ = \max(0, a) \), and \( \mu, \sigma > 0, \xi \) are the location, scale and shape parameters, respectively (Coles 2001). This result provides strong support for fitting (1) to block maxima, though the asymptotic approximation applied to finite \( n \) entails several practical difficulties; see, e.g., the review by Davison and Huser (2014). Furthermore, by standardizing the individual variables as \( Y_i = 1/[1 - F(X_i)] \) \((i = 1, 2, \ldots)\), we can show that \( n^{-1} \max(Y_1, \ldots, Y_n) \) converges to a GEV distribution with \( \mu = \sigma = \xi = 1 \) (unit Fréchet distribution).

Likewise, suppose that \( X(s), X_1(s), X_2(s), \ldots \), are independent and identically distributed random processes with continuous sample paths on \( S \subset \mathbb{R}^d \), and that there exist sequences of continuous functions \( a_n(s) > 0 \) and \( b_n(s) \) such that the renormalized process of componentwise maxima \( a_n(s)^{-1}[\max\{X_1(s), \ldots, X_n(s)\} - b_n(s)] \) converges weakly to a process \( Z(s) \) with non-degenerate margins, as \( n \to \infty \). Then, \( Z(s) \) must be max-stable, i.e., for any positive integer \( k \), the finite-dimensional distributions of \( Z(s) \) and \( \max\{Z_1(s), \ldots, Z_k(s)\} \), where \( Z_1(s), \ldots, Z_k(s) \) denote independent replicates of \( Z(s) \), differ only through location and scale coefficients. In particular, margins are GEV with spatially-varying parameters \( \mu(s), \sigma(s) > 0, \xi(s) \). Furthermore, defining standardized processes as \( Y_i(s) = 1/[1 - F_s(X_i(s))] \) \((i = 1, 2, \ldots)\) with \( F_s(x) \) the marginal distribution of \( X(s) \) at location \( s \), the limiting distribution of \( n^{-1} \max\{Y_1(s), \ldots, Y_n(s)\} \) is max-stable with unit Fréchet margins. Such a limiting process is called a simple max-stable process. Standardization allows the treatment of the margins to be separated from the dependence structure.

Simple max-stable processes have been characterized by de Haan (1984); see also Schlather (2002) and de Haan and Ferreira (2006, §9.4). Given points \( \{P_i; i = 1, 2, \ldots\} \) of a Poisson process with intensity \( p^{-2} \) \((p > 0)\) and independent replicates \( \{W_i(s); i = 1, 2, \ldots\} \)
of a positive process \(W(s) (s \in S \subset \mathbb{R}^d)\) with unit mean, the process created as

\[
Z(s) = \sup_{i=1,2,\ldots} P_i W_i(s),
\]

provided it is well defined, is a simple max-stable process. Since the supremum on the right-hand side of (2) could make \(Z(s)\) be degenerate at infinity, further technical conditions need to be imposed on \(W(s)\) (de Haan and Ferreira 2006, §9.4). Conversely, under mild conditions, each continuous simple max-stable process can be decomposed as in (2). Finite-dimensional distributions for simple max-stable processes are easily derived from (2). For any set of \(D\) spatial locations \(D = \{s_1, \ldots, s_D\} \subset S\), one has

\[
\Pr\{Z(s_1) \leq z_1, \ldots, Z(s_D) \leq z_D\} = \exp\{-V_D(z_1, \ldots, z_D)\},
\]

where the so-called exponent measure \(V_D\) contains all the information about extremal dependence. It may be expressed as \(V_D(z_1, \ldots, z_D) = \text{E}\left[\max\{W(s_1)/z_1, \ldots, W(s_D)/z_D\}\right]\); in particular, when \(D = \{s_1, s_2\}\), one has

\[
V_D(z_1, z_2) = \text{E}\left[\max\left\{\frac{W(s_1)}{z_1}, \frac{W(s_2)}{z_2}\right\}\right].
\]

The exponent measure has a closed-form formula for specific choices of \(W(s)\); see, e.g., Schlather (2002), Genton et al. (2011), Huser and Davison (2013), and Opitz (2013). A useful related quantity is the so-called extremal coefficient \(\theta(s_1, s_2) = V_D(1, 1) \in [1, 2]\), which gives a measure of dependence between variables \(Z(s_1)\) and \(Z(s_2)\), or equivalently, extremal dependence between variables \(Y(s_1)\) and \(Y(s_2)\): \(\theta(s_1, s_2) = 1\) corresponds to perfect dependence and \(\theta(s_1, s_2) = 2\) to independence.

Mixing properties of max-stable processes are characterized by the value of the extremal coefficient \(\theta(s_1, s_2)\) as the lag distance \(||h|| = ||s_2 - s_1||\) tends to infinity (Kabluchko and Schlather 2010). A stationary max-stable process is called mixing if \(\theta(s_1, s_2) \equiv \theta(||h||) \to 2\), as \(||h|| \to \infty\), which loosely means that extreme events become independent as they become infinitely distant apart. Similarly, a sequence \(s_n \in S\) is called an admissible trajectory for location \(s_0 \in S\) if \(||s_0 - s_n|| \to \infty\) as \(n \to \infty\), and a non-stationary max-stable process is
called *locally mixing at location* $s_0 \in S$ along an admissible trajectory $s_n$ if $\theta(s_0, s_n) \to 2$ as $n \to \infty$. We say that a max-stable process is *mixing everywhere* if it is locally mixing at every location $s_0 \in S$ along any admissible trajectory. Of course, in the stationary case, all these notions coincide. In Appendix A, we provide an example of a max-stable process that is locally mixing in some region along some trajectories, but not mixing everywhere.

For more details about univariate and multivariate extremes, see [Beirlant et al. (2004)](#), and for an account of spatial extremes, see the review papers by [Davison et al. (2012)](#), Cooley et al. (2012) and Davison et al. (2013). See also the book by [de Haan and Ferreira (2006)](#), which explains the technicalities in depth.

### 2.2 The celebrated Smith model and its non-stationary extension

The first stationary max-stable model proposed in the literature is the Smith model ([Smith, 1990](#)), which assumes in [2] that $W_i(s) = \phi_d(s - U_i; \Omega)$, where the $U_i$s are the points of a unit rate Poisson process on $S$ and $\phi_d(\cdot; \Omega)$ denotes the $d$-dimensional Gaussian density function with covariance matrix $\Omega$. Although finite-dimensional distributions are known in arbitrary dimensions ([Genton et al., 2011](#)), they are always degenerate for $D > d + 1$, which raises the question of the suitability of the Smith model in practice. The non-stationary extension proposed by [Smith and Stephenson, 2009](#) considers spatially varying covariance matrices $\Omega_s$, capturing the small-scale dependence structure around location $s \in S$. The generalized storm profiles are of the form

$$W_i(s) = \phi_d(s - U_i; \Omega_{U_i}).$$

(5)

This model is locally elliptic, in the sense that infinitesimal contours of the extremal coefficient form ellipses everywhere, see Figure 1. Several special cases may be of interest in practice: if contours are locally circular with $\Omega_s = \omega^2(s)I_d$, where $\omega(s) > 0$ and $I_d$ is the $d$-by-$d$ identity matrix, the model is locally isotropic (top right panel of Figure 1), and when $\omega(s) = \omega > 0$ for all $s \in S$, (5) reduces to the stationary isotropic case, i.e., the classical Smith model (top left panel of Figure 1). When $\Omega_s = \omega^2(s)R$ for some fixed $d$-by-$d$ corre-
Figure 1: Simulations of Model (5) for locations \( s = (s_x, s_y) \in [0, 1]^2 \). *Top left:* a stationary isotropic case with \( \Omega_s = 0.1^2 I_2 \); *top right:* a non-stationary locally isotropic case with \( \Omega_s = 0.4^2 2^{-8|s_x|} I_2 \); *bottom left:* a non-stationary homogeneously anisotropic case with \( \Omega_s = 0.4^2 2^{-8|s_x|} R, R \in \mathbb{R}^{2 \times 2} \) being a correlation matrix with correlation 0.8; *bottom right:* general non-stationary case with \((\Omega_s)_{11} = 0.4^2 2^{-8|s_x|}, (\Omega_s)_{22} = 0.4^2 2^{-8|1-s_x|}\) and \((\Omega_s)_{12} = (\Omega_s)_{21} = \{(\Omega_s)_{11}(\Omega_s)_{22}\}^{1/2}\{e^{h(s)} - 1\}/\{e^{h(s)} + 1\}\), where \( h(s) = 2\log(3)e^{-30|s_x-0.5|^2} \). Realizations are based on the same random seed. The contours correspond to \( \theta(s_1, s_2) = 1.2, 1.5, 1.8 \) (narrow to wide), where \( s_1 \) is the center location (cross). The color scale indicates quantile probabilities.

As the isotropic correlation matrix \( R \), the model is not isotropic, but the anisotropy is homogeneous over space; see the bottom left panel of Figure 1. If \( \omega(s) = \omega > 0 \) for all \( s \in S \), this model reduces to the stationary anisotropic case, illustrated by Blanchet and Davison (2011). Smith and Stephenson (2009) provide bivariate margins in the homogeneously anisotropic case only (and sub-cases); in Appendix B calculations are performed in full generality for \( d = 2 \).
The stationary Smith model is mixing, which implies that complete independence can be captured at infinity. In Appendix C we show that the Smith–Stephenson model with \( \Omega_s = \omega(s)^2 I_d \) (the locally isotropic case) is mixing everywhere if and only if \( \omega(s) = o(\|s\|) \), and by a simple extension, this is also true when \( \Omega_s = \omega(s)^2 R \) with \( R \) a correlation matrix (the homogeneously anisotropic case). This result makes sense because if one has \( \omega(s) = O(\|s\|) \), the extent of a storm centered at \( s \) increases at the same rate as the distance separating \( s \) from any fixed other point \( s_0 \), such that the storm contributes to the supremum \( \{ \] at location \( s_0 \) with positive probability, no matter how far it is from \( s_0 \). If \( \omega(s) \) is uniformly bounded over \( S \), the corresponding max-stable process is mixing everywhere. In particular, examples displayed in Figure 1 are all mixing everywhere.

Although the Smith–Stephenson model is easily interpretable and closely linked to the existing literature, it has several limitations. First, finite-dimensional distributions are known for \( d = 2 \) only. Second, pairwise densities involve the cumulative distribution and density function of quadratic forms of normal variables, which may be computationally intensive to compute (see Appendix B). Finally, as illustrated in Figure 1, this process is very smooth. Realizations are infinitely differentiable in neighborhoods of all points that do not lie on the border between distinct storms, and this appears too strong an assumption in most environmental applications. In fact, the storm profiles are almost deterministic; randomness is solely created by the storm locations \( U_i \). More flexible non-stationary max-stable models with stochastic storm profiles, generalizing (5), are proposed in Section 3.

3 Flexible non-stationary dependence structures

3.1 The non-stationary extremal \( t \) model

The extremal \( t \) model (Nikoloulopoulos et al., 2009; Opitz, 2013; Thibaud and Opitz, 2014) is defined by taking

\[
W(s) = \frac{2^{-df/2} \pi^{1/2}}{\Gamma\left\{ (df + 1)/2 \right\}} \max\{0, \varepsilon(s)\}^{df}
\]

(6)
in (2), where \( \text{df} > 0 \), \( \varepsilon(s) \) is a Gaussian process with zero mean, unit variance and correlation function \( \rho(s_1, s_2) \), and \( \Gamma(\cdot) \) is the gamma function. The extremal \( t \) model is not mixing unless \( \text{df} \to \infty \) (Davison et al. 2012), but this issue may be resolved by incorporating a random set element (Davison and Gholamrezaee 2012; Huser and Davison 2014), though the inference is more tricky and might be subject to parameter identifiability problems. The stationary version of model (6), and its sub-models, have been applied extensively. In particular, when \( \text{df} = 1 \), (6) reduces to the Schlather (2002) model, which has been fitted in numerous applications (Davison and Gholamrezaee 2012; Davison et al. 2012; Ribatet 2013; Thibaud et al. 2013). Another interesting stationary sub-model is the Brown–Resnick process (Brown and Resnick 1977; Kabluchko et al. 2009), which arises as a limiting case of (6) as \( \text{df} \to \infty \) (Davison et al. 2012); its storm profiles may be expressed as \( W(s) = \exp\{\varepsilon(s) - \gamma(s)\} \), where \( \varepsilon(s) \) is a Gaussian random field with semi-variogram \( \gamma(h) \) such that \( \varepsilon(0) = 0 \) almost surely. The Brown–Resnick process extends the Smith model (Huser and Davison 2013), and it can also be viewed as the generalization of the Hüsler and Reiss (1989) multivariate extreme-value distribution to the spatial framework. In practice, Brown–Resnick processes have proven to be quite flexible compared to the Smith and Schlather alternatives (Davison et al. 2012; Jeon and Smith 2012). Model (6) not only generalizes all aforementioned stationary max-stable models, but it is also the max-attractor for the broad class of all suitably rescaled elliptical processes (Demarta and McNeil 2005; Opitz 2013), which provides strong support for its use in practice. The bivariate exponent measure for (6) may be expressed as

\[
V_D(z_1, z_2) = \frac{1}{z_1} T_{\text{df}+1} \left[ \frac{(\text{df} + 1)^{1/2} \left\{ \frac{(z_2/z_1)^{1/\text{df}} - \rho(s_1, s_2)}{1 - \rho(s_1, s_2)^2} \right\}^{1/2}}{1 - \rho(s_1, s_2)^2} \right] + \frac{1}{z_2} T_{\text{df}+1} \left[ \frac{(\text{df} + 1)^{1/2} \left\{ \frac{(z_1/z_2)^{1/\text{df}} - \rho(s_1, s_2)}{1 - \rho(s_1, s_2)^2} \right\}^{1/2}}{1 - \rho(s_1, s_2)^2} \right],
\]

where \( T_{\text{df}}(\cdot) \) is the Student \( t \) cumulative distribution function with \( \text{df} \) degrees of freedom. Explicit expressions in dimension \( D \) are also available (see Thibaud and Opitz 2014).

Our approach to modeling non-stationarity in spatial extremes consists of combining
the extremal $t$ model \((6)\) with non-stationary correlation functions $\rho(s_1, s_2)$ proposed in the classical spatial statistics literature. As mentioned above, there exist several ways to construct non-stationary correlation functions, spanning from space deformations to SPDEs, and including wavelets, spectral methods, mixtures of stationary correlations or kernel convolutions. Hence, our methodology to tackle non-stationarity in extremes is very general and can potentially yield a large variety of models. In the present paper, we focus on the kernel convolution approach advocated by Paciorek and Schervish (2006). The latter have proposed a very general construction of non-stationary correlation functions that are based on known isotropic correlation models. Specifically, let $\Omega_s$ denote a (smoothly) spatially varying $d$-by-$d$ covariance matrix, and for any two locations $s_1, s_2 \in S$ with separation vector $h = s_2 - s_1$, define the quadratic form $Q_{s_1; s_2}$ as

$$Q_{s_1; s_2} = h^T \left( \frac{\Omega_{s_1} + \Omega_{s_2}}{2} \right)^{-1} h. $$

(8)

Paciorek and Schervish (2006) show that for any isotropic correlation function $R(\|h\|)$ valid on $\mathbb{R}^d$ ($d = 1, 2, \ldots$), the function

$$\rho(s_1, s_2) = \left| \Omega_{s_1} \right|^{1/4} \left| \Omega_{s_2} \right|^{1/4} \left| \frac{\Omega_{s_1} + \Omega_{s_2}}{2} \right|^{-1/2} R \left( Q_{s_1; s_2}^{1/2} \right) $$

(9)

provides a valid non-stationary correlation function on $\mathbb{R}^d$ ($d = 1, 2, \ldots$). Furthermore, to avoid parametrization redundancy, the function $R(\|h\|)$ can be assumed to have unit range. In particular, choosing

$$R(\|h\|) = \frac{\|h\|^\nu}{\Gamma(\nu)2^{\nu-1}K_\nu(\|h\|)}, $$

(10)

where $K_\nu$ is the modified Bessel function of second kind and $\nu > 0$, model \((9)\) yields a flexible non-stationary correlation based on the widely used Matérn model (Handcock and Stein, 1993; Stein, 1999; Guttorp and Gneiting, 2006), where the degree of differentiability of the associated Gaussian random field can be controlled through the smoothness parameter $\nu$ (Gneiting and Guttorp, 2010). An alternative possibility is to consider the powered exponential family, i.e.,

$$R(\|h\|) = \exp \left( -\|h\|^\alpha \right), $$

(11)
Figure 2: Simulation of the extremal $t$ model (6), with df = 5 and non-stationary correlation function (9), combined with (11), in $[0,1]^2$. Columns correspond to different smoothness scenarios, with $\alpha = 0.5, 1.0, 1.5$ (left to right). Locally isotropic (top) and general non-stationary (bottom) cases are displayed. The underlying spatially-varying matrices are $\Omega_s = (2 \text{df})^{2/\alpha} \times \Omega_s^{BR}$, where $\Omega_s^{BR}$ is given in Figure 1's caption. Realizations are from the same random seed. Contour curves correspond to $\theta(s_1, s_2) = 1.2, 1.5, 1.8$ (narrow to wide), where $s_1$ is the center location (cross). The color scale indicates quantile probabilities.

where $\alpha \in (0, 2]$ is a smoothness parameter, and the exponential and squared exponential models correspond to $\alpha = 1$ and $\alpha = 2$, respectively. This correlation family covers scenarios from very rough random fields (with $\alpha \to 0$) to random fields with analytical sample paths (with $\alpha = 2$). Hence, a great deal of flexibility can be attained by combining (9) with (10) or (11). Since the max-stable model in (6) inherits its sample path differentiability properties from the underlying Gaussian process $\varepsilon(s)$, the parameters $\nu$ in (10) and $\alpha$ in (11) have a direct relationship with the smoothness of the resulting max-stable process. To illustrate this, typical realizations from the non-stationary extremal $t$ model with df = 5 combined with (9) and (11) are displayed in Figure 2. It can be noticed from the extremal coefficient
contours that rougher random fields are less dependent at short distances.

Like the non-stationary Smith model, the correlation function (9) is locally elliptic, and this geometric property is therefore shared with the resulting non-stationary max-stable random field. This implies that the latter can be seen locally as a smoothly deformed isotropic max-stable process. To see this, fix $s_0 \in S$ and let $s_1, s_2 \in N(s_0) \subset S$ be two locations within some small neighborhood $N(s_0)$ of $s_0$. By continuity of the map $s \mapsto \Omega_s$, one has that $\Omega_{s_2} \approx \Omega_{s_1} \approx \Omega_{s_0}$ and $Q_{s_1; s_2} \approx h^T \Omega_{s_0}^{-1} h$, where $h = s_2 - s_1$ is the lag vector. Then, applying the spatial transformation $s \mapsto s^* = \Omega_{s_0}^{-1/2} (s - s_0)$ in $N(s_0)$, one can easily verify that the correlation function on the new coordinate system satisfies $\rho(s_1^*, s_2^*) \approx R(\|h^*\|)$ with $h^* = s_2^* - s_1^*$, and that it is therefore locally isotropic.

The non-stationary extremal $t$ model defined above using (9) and (11) with covariance matrices $\Omega_s = (2 df)^{2/\alpha} \times \Omega_{s}^{BR}$, converges as $df \to \infty$ to the Brown–Resnick process with variogram $2\gamma(s_1, s_2) = (Q_{s_1; s_2}^{BR})^{\alpha/2}$, where $Q_{s_1; s_2}^{BR}$ is defined in (8) using $\Omega_{s}^{BR}$. In particular, the Smith–Stephenson model (5) is recovered when $\alpha = 2$ (Huser and Davison, 2013).

### 3.2 Covariates

For simplicity, we now continue our modeling of non-stationarity for $d = 2$, although our approach could be applied in higher dimensions. In order to incorporate covariate effects in the dependence structure at extreme levels, we propose further modeling the covariance matrices $\Omega_s \ (s \in S)$ as follows: let

$$\Omega_s = \begin{pmatrix} \omega^2_x(s) & \omega_x(s)\omega_y(s)\delta(s) \\ \omega_x(s)\omega_y(s)\delta(s) & \omega^2_y(s) \end{pmatrix},$$

with, for example,

$$\log\{\omega_x(s)\} = X^{T}_{\omega_x}(s)\beta_{\omega_x}, \quad \log\{\omega_y(s)\} = X^{T}_{\omega_y}(s)\beta_{\omega_y}, \quad \logit\{\delta(s) + 1\}/2 = X^{T}_{\delta}(s)\beta_{\delta},$$

where $X_{\omega_x}(s), X_{\omega_y}(s)$ and $X_{\delta}(s)$ denote vectors of covariates corresponding to location $s$, and $\beta_{\omega_x}, \beta_{\omega_y}$ and $\beta_{\delta}$ are the associated vectors of parameters measuring importance of
covariates. The link functions in (13) ensure that $\omega_x(s) > 0$, $\omega_y(s) > 0$ and $\delta(s) \in (-1, 1)$, but they could in principle be replaced by other functions that satisfy these conditions. The construction (12) guarantees the positive definiteness of $\Omega_s$. In higher dimensions, spatially varying covariance matrices can be created similarly. The local correlation range at station $s$ with respect to the $x$ (respectively $y$) axis is measured by the functions $\omega_x(s)$ (respectively $\omega_y(s)$), whereas $\delta(s)$ captures the local anisotropy level: if $\delta(s) = 0$ and $\omega_x(s) = \omega_y(s)$, the resulting process is locally isotropic, i.e., infinitesimal contours are circular everywhere, whereas if $\delta(s) \neq 0$, contours take the form of slanted ellipses; see Figure 2.

### 3.3 Mixture models

Although the non-stationary model (6) appears quite flexible, one limitation is that it cannot capture different smoothness behaviors in distinct regions of the space. This issue may be overcome by using the idea of mixtures.

The first type of mixture consists of max-mixtures of max-stable models. Let $Z^1(s)$ and $Z^2(s)$ be independent max-stable processes with unit Fréchet margins defined on the same space $S$. Then for any function $0 \leq a(s) \leq 1$, the spatial process defined as $Z(s) = \max[a(s)Z^1(s), \{1 - a(s)\}Z^2(s)]$ is a simple max-stable process with exponent measure

$$V_D(z_1, \ldots, z_D) = a(s)V^1_D(z_1, \ldots, z_D) + \{1 - a(s)\}V^2_D(z_1, \ldots, z_D),$$

where $V^1_D(z_1, \ldots, z_D)$ and $V^2_D(z_1, \ldots, z_D)$ are the exponent measures of $Z^1(s)$ and $Z^2(s)$, respectively. The function $a(s)$ is a spatially varying proportion, determining which of the processes $Z^1(s)$ and $Z^2(s)$ is dominant at location $s$. Model (14) is stationary if $a(s)$ is constant over space and $Z^1(s)$ and $Z^2(s)$ are stationary, but it can be made non-stationary by allowing $a(s)$ to depend upon covariates, e.g., $\text{logit}\{a(s)\} = X^T_a(s)\beta_a$, where $X_a(s)$ is a vector of covariates for location $s$ and $\beta_a$ is the associated vector of parameters. One interesting feature of this model is that different smoothness behaviors may be captured in different regions, provided that $Z^1(s)$ and $Z^2(s)$ have different smoothness properties. Furthermore, replacing $Z^2(s)$ by an asymptotically independent process allows models to be constructed
with different asymptotic behaviors at different distances (Wadsworth and Tawn, 2012).

More complex non-stationary max-stable models $Z(s)$ may be constructed by considering a collection of independent stationary max-stable random fields $Z^1(s), \ldots, Z^k(s)$ with unit Fréchet margins and associated proportions $a^1(s), \ldots, a^k(s) \in [0, 1]$ such that $\sum_{i=1}^k a^i(s) = 1$ for each $s$, yielding the simple max-stable process $Z(s) = \max_{i=1, \ldots, k} \{a^i(s)Z^i(s)\}$. In practice, however, this model involves a large number of parameters, which implies that strong parameter constraints must be imposed. A related max-stable model, based on spatial knots, is proposed by Reich and Shaby (2012).

The second type of mixture consists of sum-mixtures of correlation functions. Specifically, let $\rho^1(s_1, s_2), \rho^2(s_1, s_2)$ be two correlation functions, and let $0 \leq a(s) \leq 1$ be a function defined on $S$. Then, a non-stationary extremal $t$ model may be obtained by using a Gaussian process $\varepsilon(s)$ in (6) with correlation function

$$
\rho(s_1, s_2) = a(s_1)a(s_2)\rho^1(s_1, s_2) + \{1 - a(s_1)a(s_2)\}\rho^2(s_1, s_2).
$$

(15)

Here, the proportion is not $a(s)$, but rather $a(s_1)a(s_2)$. Notice that model (15) is related but does not exactly correspond to the Gaussian mixture model advocated by Fuentes (2001) and Reich et al. (2011). Again, the coefficient $a(s)$ may be modeled in terms of covariates. Similarly, different smoothness behaviors over the space may be captured by the different mixture components. As above, model (15) can easily be extended to higher-dimensional mixtures, though this may lead to heavy parametrization.

4 Inference

4.1 Pairwise likelihoods

Likelihood inference for max-stable processes is not an easy task. The joint density for max-stable processes stems from the differentiation of (3) with respect to $z_1, \ldots, z_D$. In dimension $D = 2$, this equals $(V_1V_2 - V_{12}) \exp(-V)$, where $V_1 = \partial V_D(z_1, z_2)/\partial z_1$, etc., where the subscript $D$ and the arguments are dropped for clarity. However, as $D$ increases,
the size of this expression renders the full likelihood quickly intractable. To illustrate this fact, the number of terms in the full likelihood when $D = 10, 20, 50, 100$ is of the order $10^5, 10^{13}, 10^{47}, 10^{115}$, respectively. Thus, to enumerate all likelihood terms for dimension $D = 100$, it would take roughly $10^{80}$ billion years using the fastest supercomputer currently available! To get around this computational bottleneck, the use of pairwise likelihoods is now a common practice (see, e.g., Padoan et al. 2010; Huser and Davison 2014). Denoting the vector of unknown parameters by $\psi \in \Psi \subset \mathbb{R}^p$, log pairwise likelihoods for model (3) may be expressed as

$$
\ell(\psi) = \sum_{i=1}^{m} \sum_{(j_1,j_2) \in \mathcal{P}} \log \left\{ V_1(z_{j_1;i}; z_{j_2;i}) V_2(z_{j_1;i}; z_{j_2;i}) - V_{12}(z_{j_1;i}; z_{j_2;i}) \right\} - V(z_{j_1;i}; z_{j_2;i}),
$$

(16)

where $z_{j;i}$ denotes the $i$th block maximum recorded at the $j$th station, $i = 1, \ldots, m$, $j = 1, \ldots, D$, and where the non-empty set $\mathcal{P} \subset \mathcal{P}_{\text{tot}} = \{(j_1,j_2) : 1 \leq j_1 < j_2 \leq D\}$ defines the pairs of observations included in the pairwise likelihood. If $\mathcal{P} = \mathcal{P}_{\text{tot}}$, all pairs are considered in (16). Computational and statistical efficiency might however be gained by carefully selecting a much smaller number of pairs (Huser and Davison 2014; Castruccio et al. 2014). A possibility is to include a small fraction of informative pairs, i.e., typically the most dependent ones, though Huser and Davison (2014) show that further improvements may be obtained in special cases by including some weakly dependent pairs as well. For stationary isotropic processes, this might be achieved by including the closest pairs, whereas for non-stationary max-stable processes, one might consider pairs $(j_1,j_2)$ with the lowest extremal coefficients $\theta(s_{j_1}, s_{j_2})$. Since the latter are unknown in practice, the choice of pairs might be guided by pre-computed empirical extremal coefficients $\hat{\theta}(s_{j_1}, s_{j_2})$, e.g., calculated using the $F$-madogram estimator (Cooley et al. 2006); however, simulations (not shown) have revealed that this approach induces a “selection bias”. In complex non-stationary settings with large numbers of locations, Delauney triangulations may be helpful to select the pairs. Under temporal independence, the maximum pairwise likelihood estimator $\hat{\psi}$, which maximizes (16), is strongly consistent, asymptotically Gaussian, converges at $m^{1/2}$ rate, and its asymptotic variance is of the sandwich form, as is typical for mis-specified
likelihood estimators (Padoan et al., 2010). More precisely, if \( \psi_0 \in \text{int}(\Psi) \) denotes the “true” parameter vector, then under mild regularity conditions, one has the large sample approximation

\[
\hat{\psi} \sim N_p(\psi_0, J(\psi_0)^{-1}K(\psi_0)J(\psi_0)^{-1}), \quad m \to \infty,
\]

where \( J(\psi) = \mathbb{E}\{-\partial^2 \ell(\psi)/\partial \psi \partial \psi^T\} \in \mathbb{R}^{p \times p} \) and \( K(\psi) = \text{var}\{\partial \ell(\psi)/\partial \psi\} \in \mathbb{R}^{p \times p} \). Uncertainty may be assessed by plugging estimates of the matrices \( J(\psi_0) \) and \( K(\psi_0) \) into the asymptotic variance in (17); see Padoan et al. (2010). Alternatively, one can bootstrap the independent replicates \( z_i = (z_{1,i}, \ldots, z_{D,i}), \ i = 1, \ldots, m, \) and re-estimate parameters using the pseudo-samples, to assess the variability surrounding \( \hat{\psi} \). Similar asymptotic properties hold for mildly time-dependent processes (Davis et al., 2013; Huser and Davison, 2014) in which uncertainty may be assessed using block bootstrap. Model comparison is typically performed using the composite likelihood information criterion (CLIC), defined as

\[
\text{CLIC} = -2\ell(\hat{\psi}) + 2\text{tr}\{J(\hat{\psi})^{-1}K(\hat{\psi})\},
\]

which can be interpreted similarly to the Akaike information criterion. Another possibility is to use the composite Bayesian information criterion (CBIC), i.e., the counterpart of the classical Bayesian information criterion. It is defined as

\[
\text{CBIC} = -2\ell(\hat{\psi}) + \log(m)\text{tr}\{J(\hat{\psi})^{-1}K(\hat{\psi})\},
\]

and therefore penalizes model complexity more than does CLIC. The lower the CLIC or CBIC, the better the model. Theoretical properties of CLIC and CBIC have recently been investigated by Ng and Joe (2014) (in which CLIC and CBIC are called instead CLAIC and CLBIC, respectively). In particular, they show that CLIC has a tendency to select over-complicated models. For a broad survey of composite likelihood methods, see Varin et al. (2011).

4.2 Simulation study

In this simulation study, we assess the ability of the maximum pairwise likelihood estimator (17) to estimate and detect non-stationarity dependence structures in a large variety of contexts. We also study the effect of neglecting non-stationarity on spatial return levels.

Throughout this section, we focus on the locally isotropic extremal t model illustrated
in the left column of Figure 2 and consider various parameter combinations. Specifically, the extremal $t$ process with $\text{df} = 1, 2, 5, 10$ is simulated on $[0, 1]^2$, using the non-stationary correlation function $\rho(s_1, s_2)$ defined in (9) based on the powered exponential model (11) with $\alpha = 0.5, 1, 1.5, 1.9$ (rough to smooth). The underlying spatially varying covariance matrix is taken to be of the form $\Omega_s = (2 \text{df})^{2/\alpha} \times \omega(s)^2 I_2$, where $I_2$ is the 2-by-2 identity matrix and $\omega(s) = \beta_1 2^{-\beta_2 |s_x|}$, $s = (s_x, s_y)$, with range $\beta_1 > 0$ and non-stationary parameter $\beta_2 \in \mathbb{R}$.

To investigate different non-stationary scenarios, we consider $(\beta_1, \beta_2) = (0.1, 0)$ (stationary), $(0.1\sqrt{2}, 1)$ (weakly non-stationary), $(0.2, 2)$ (mildly non-stationary), and $(0.4, 4)$ (strongly non-stationary). Although these scenarios exhibit different non-stationarity patterns, the overall dependence strength is comparable in the sense that all cases satisfy $\omega(s) = 0.1$ for any $s = (0.5, s_y)$. The df = 1 case corresponds to a non-stationary Schlather process, whereas the df = 10 case is a crude approximation of a non-stationary Brown–Resnick process (with $\alpha = 1.9$ corresponding approximately to the non-stationary Smith model); recall Section 3.1.

In each case, $m = 10, 20, 50, 100$ independent replicates of these processes are simulated at $S = 10, 20, 50, 100$ fixed locations uniformly sampled in the unit square. Simulations are repeated 300 times to compute empirical diagnostics.

**Estimation and detection of non-stationarity.** We first investigate the performance of the maximum pairwise likelihood estimator (17) to recover the true parameters under the correct model. To this end, we estimate model parameters $\psi = (\beta_1, \beta_2, \text{df}, \alpha)^T$ with (17) using the 10% closest pairs; then we derive the empirical root mean squared errors (RMSE) from the 300 independent experiments. Results are reported in Table 1.

The range parameter $\beta_1$ is quite well identified overall. The corresponding RMSE is less than 0.02, 0.04, 0.06 and 0.12 for $\beta_1 = 0.1$, $0.1\sqrt{2}$, 0.2, and 0.4, respectively, and it decreases as the smoothness parameter $\alpha$ increases, and as the degrees of freedom (df) increase. Furthermore, the higher $\beta_1$, the larger its RMSE, as expected. The RMSE for the non-stationary parameter $\beta_2$ follows a similar pattern, though large values of $\beta_2$ seem
Table 1: Root mean squared error (×100) of the maximum pairwise likelihood estimator \([17]\), using the 10% closest pairs, for the locally isotropic extremal \(t\) with various parameter combinations. These diagnostics are computed for parameters \(\beta_1/\beta_2/df/\alpha\) from 300 independent simulations of \(m = 100\) independent max-stable processes simulated at \(S = 100\) fixed locations in \([0, 1]^2\).

| \((\beta_1, \beta_2)\) | df  | 0.5    | 1.0    | 1.5    | 1.9    |
|-----------------------|-----|--------|--------|--------|--------|
| \((0.1, 0)\)          | 1   | 2/42/12/3 | 1/21/11/4 | 1/14/10/5 | 1/11/10/5 |
|                       | 2   | 2/41/30/3 | 1/21/27/3 | 1/13/23/5 | 0/10/22/4 |
|                       | 5   | 2/32/113/3 | 1/17/92/3 | 0/11/92/3 | 0/10/100/3 |
|                       | 10  | 1/27/434/2 | 1/15/392/3 | 0/11/277/3 | 0/8/417/3 |
| \((0.1\sqrt{2}, 1)\)  | 1   | 4/47/12/3 | 2/22/10/4 | 1/15/11/5 | 1/11/10/6 |
|                       | 2   | 3/43/30/3 | 1/20/26/3 | 1/14/22/4 | 1/10/23/4 |
|                       | 5   | 3/36/111/3 | 1/19/93/3 | 1/12/88/3 | 1/10/93/3 |
|                       | 10  | 2/28/285/3 | 1/17/259/3 | 1/12/313/3 | 1/10/328/3 |
| \((0.2, 2)\)          | 1   | 6/55/12/4 | 2/23/11/4 | 1/15/9/5   | 1/13/8/5  |
|                       | 2   | 5/43/29/3 | 2/21/25/3 | 1/15/23/4 | 1/12/26/4 |
|                       | 5   | 4/31/92/3 | 2/20/78/3 | 1/13/80/3 | 1/11/81/3 |
|                       | 10  | 3/24/217/3 | 2/17/183/3 | 1/13/193/3 | 1/11/200/3 |
| \((0.4, 4)\)          | 1   | 12/57/12/4 | 7/31/10/5 | 4/20/8/6 | 3/16/7/5 |
|                       | 2   | 8/41/26/4 | 5/24/23/4 | 4/18/21/4 | 3/15/18/4 |
|                       | 5   | 7/38/75/4 | 4/20/57/4 | 3/14/55/4 | 2/13/52/4 |
|                       | 10  | 7/37/194/4 | 4/20/128/3 | 3/15/128/4 | 2/12/116/4 |

easier to estimate overall: for strongly non-stationary scenarios, the RMSE is quite small in comparison to the actual value of \(\beta_2\). This is certainly due to the very rigid type of assumed non-stationarity: a small perturbation of \(\beta_2\) entails a dramatic change in the dependence structure. The performance for the degrees of freedom (df) is more tricky to interpret. The RMSE increases for larger df values, but there is no general trend as a function of \(\alpha\) or \((\beta_1, \beta_2)\). However, it seems that in most cases, the RMSE for df usually forms a U-shape as a function of \(\alpha\) and attains its minimum between \(\alpha = 1\) and \(\alpha = 1.5\). Moreover, since two large df values might be difficult to distinguish, the corresponding RMSE is often quite high. Finally, the RMSE for \(\alpha\) is quite low and more or less constant as a function of the non-stationarity level. It decreases with increasing df values.

To illustrate large-sample and infill asymptotics properties of the estimator \([17]\), Figure 3 displays boxplots of parameter estimates, as a function of \(m\) and \(S\) for the extremal \(t\) model with \(df = 5\), \(\alpha = 1\) and \((\beta_1, \beta_2) = (0.2, 2)\). As expected, the estimator appears to be
Figure 3: Boxplots of parameter estimates obtained from data generated from the locally isotropic extremal $t$ process with $df = 5, \alpha = 1$ and $(\beta_1, \beta_2) = (0.2, 2)$. Estimator (17) was used, including the 10% closest pairs. Green boxes (left of vertical dashed line) show the performance for a fixed number of locations, $S = 100$, and an increasing number of independent replicates, $m = 10, 20, 50, 100$. Blue boxes (right of vertical dashed line) show the performance for a fixed number of replicates, $m = 100$, and an increasing number of locations, $S = 10, 20, 50, 100$. Horizontal red lines are true values.

consistent as $m$ increases. In addition, parameters are much better estimated if the data are collected at a dense network of sites, which suggests that the estimator might be consistent as $S \to \infty$, although this is not a formal proof.

We now explore the ability of estimator (17) to detect the spatial heterogeneity. For each simulated dataset, we fit the true non-stationary model and the (restricted) stationary counterpart, computing in each case the corresponding CLIC and CBIC diagnostics defined in Section 4.1. These information criteria were computed using finite differences combined with the direct method of Padoan et al. (2010). The empirical percentages that the CLIC and CBIC are in favor of the true underlying model (either stationary if $\beta_2 = 0$, or non-stationary otherwise) are calculated from the 300 experiments and reported in Table 2 for $S = 100$ and
Table 2: Percentage of correct model identifications (stationary vs. non-stationary) using the CLIC/CBIC for the locally isotropic extremal $t$ with various parameter combinations. The log pairwise likelihood defined in (16) uses the 10% closest pairs. Values are computed from 300 independent simulations of $m = 100$ independent max-stable processes simulated at $S = 100$ fixed locations in $[0, 1]^2$.

| $(\beta_1, \beta_2)$ | df | Smoothness parameter $\alpha$ | 0.5 | 1.0 | 1.5 | 1.9 |
|----------------------|----|-----------------------------|-----|-----|-----|-----|
| $(0.1, 0)$           | 1  | 67/83, 68/84, 67/84, 66/80  |
|                      | 2  | 63/81, 64/85, 65/84, 68/85  |
|                      | 5  | 64/81, 66/85, 69/84, 61/82  |
|                      | 10 | 62/74, 64/80, 61/78, 64/83  |
| $(0.1\sqrt{2}, 1)$  | 1  | 85/66, 100/99, 100/100, 100/100 |
|                      | 2  | 88/73, 100/100, 100/100, 100/100 |
|                      | 5  | 91/78, 100/100, 100/100, 100/100 |
|                      | 10 | 93/85, 100/100, 100/100, 100/100 |
| $(0.2, 2)$           | 1  | 99/98, 100/100, 100/100, 100/100 |
|                      | 2  | 100/98, 100/100, 100/100, 100/100 |
|                      | 5  | 100/100, 100/100, 100/100, 100/100 |
|                      | 10 | 100/100, 100/100, 100/100, 100/100 |
| $(0.4, 4)$           | 1  | 100/100, 100/100, 100/100, 100/100 |
|                      | 2  | 100/100, 100/100, 100/100, 100/100 |
|                      | 5  | 100/100, 100/100, 100/100, 100/100 |
|                      | 10 | 100/100, 100/100, 100/100, 100/100 |

$m = 100$. Overall, non-stationarity in the dependence structure seems easily detectable when $\alpha > 0.5$, or when the non-stationarity level is moderate to strong, with almost 100% of success in each case with the CLIC or CBIC. By contrast, the performance is poor in near-stationary cases and with $\alpha = 0.5$; this is especially striking for the CBIC, which penalizes more model complexity. In case of stationarity, the CLIC selects the true model in about 65% of occasions, whereas the CBIC attains about 80% of success. This suggests that these information criteria, but especially the CLIC, have “more power” to select bigger models, and that they should be interpreted with care in practice. This observation agrees with the findings of [Ng and Joe (2014)]. Furthermore, the ability to distinguish between stationarity and non-stationarity improves when more data are available. For example, for fixed $S = 20$ and parameters df = 5, $\alpha = 1$, $(\beta_1, \beta_2) = (0.2, 2)$, the CLIC percentages are 63%, 79%, 93%, 99%, for $m = 10, 20, 50, 100$, respectively; similarly, for fixed $m = 20$, these values are 42%, 79%, 98%, 100%, for $S = 10, 20, 50, 100$, respectively.
**Effect of model misspecification on return levels.** Neglecting non-stationarity when the data are in fact non-stationary might have serious consequences on the estimation of spatial return levels. To assess this, we consider the locally isotropic extremal $t$ model on the Gumbel scale, with $\text{df} = 5$ and $\alpha = 1.5$. For $(\beta_1, \beta_2) = (0.1, 0)$ (stationary case) and $(\beta_1, \beta_2) = (0.4, 4)$ (strongly non-stationary case), we compute return levels for the integral $\text{INT}_j = \int_{S_j} Z(s)ds$, the spatial minimum $\text{MIN}_j = \min_{s \in S_j} \{Z(s)\}$, and the spatial maximum $\text{MAX}_j = \max_{s \in S_j} \{Z(s)\}$, $j = 1, 2$, of the max-stable process $Z(s)$ over the domains $S_1 = [0, 0.2] \times [0, 1]$ and $S_2 = [0.8, 1] \times [0, 1]$. In practice, these domains are pixelated using a fine grid comprising 105 grid points with equal spacings of 0.05. Assuming that $Z(s)$ describes the annual maximum process for some quantity of interest, we then derive the $N$-year return level for $\text{INT}_j$ and $\text{MIN}_j$ as the empirical $(1 - 1/N)$-quantile calculated from one million independent simulations of $Z(s)$. Return levels $z_{N;\text{MAX}_j}$ for $\text{MAX}_j$ are derived using the exact formula $z_{N;\text{MAX}_j} = \log\{\theta(S_j)\} - \log\{-\log(1 - 1/N)\}$ and an estimate of the areal extremal coefficient $\theta(S_j)$ (Lantuéjoul et al., 2011). The latter determines the effective number of independent extremes in region $S_j$; for the stationary case, one finds $\theta(S_1) = \theta(S_2) \approx 8.6$, and for the non-stationary case, $\theta(S_1) \approx 4.2$, $\theta(S_2) \approx 23.6$, indicating that extremal dependence in $S_1$ is overall much stronger than in $S_2$. Results are shown in Figure 4.
One can see that mis-specification (and therefore also mis-estimation) of spatial dependence strongly affects the return levels of spatial quantities. Underestimation of dependence implies underestimation of return levels for INT$_j$ and MIN$_j$ and overestimation of return levels for MAX$_j$ (and vice versa). Although this depends on the level of non-stationarity, the underlying parameters, and marginal distributions, in practice it is crucial to capture correctly the non-stationarity in the dependence structure.

5 Analysis of temperature maxima

We now discuss an application to a temperature dataset recorded in Colorado during the period 1895-1997, which is freely available on the National Center for Atmospheric Research website. We selected stations in the Front Range area, with at least 40 years of data, and extracted maxima over the months May–September (roughly corresponding to annual maxima), avoiding therefore the modeling of seasonality. Figure 5 illustrates the locations of the monitoring stations kept for the analysis, and summarizes the data availability.

As our focus is on the dependence structure, GEV($\mu, \sigma, \xi$) distributions were fitted independently at each station, assuming that maxima are independent and identically distributed over time, and the annual maxima were then transformed to the unit Fréchet scale using the estimated parameters and the probability integral transform. Quantile-quantile plots (not shown) suggest that marginal fits are relatively good, and further analysis does not show evidence of a consistent temporal trend over the region. Histograms of estimated parameters for the different stations are displayed in Figure 6. The shape parameters are all negative, indicating that distributions of temperature annual maxima have an upper bound, which seems physically plausible. The correlation between location and shape parameters is about 0.4.

We then fit a large variety of stationary and non-stationary extremal $t$ models to the transformed data using the pairwise likelihood estimator (17) including all pairs of locations. These models, summarized in Table 4 in Appendix D, are based on the Paciorek–Schervish
correlation function (9) combined with (11) and are parametrized as in (12). They are either stationary (models 1–2) or non-stationary (models 3–22), locally isotropic (models 1, 3–5, 12, 14–16) or anisotropic (models 2, 6–11, 13, 17–22), based on correlation mixtures of the form (15) (models 1–11) or non-mixtures (models 12–22). In the non-stationary models, altitude, longitude and latitude are used as covariates (on top of the intercept) in the modeling of the dependence ranges $\omega_x(s), \omega_y(s)$, the anisotropy parameter $\delta(s)$ and the mixture coefficient $a(s)$, as suggested in (13) and Section 3.3. Logarithmic links are used for
Figure 6: Histograms of estimated location $\mu$ (left), scale $\sigma$ (middle) and shape $\xi$ (right) parameters obtained from the fit of the GEV($\mu, \sigma, \xi$) distribution at all stations.

$\omega_x(s), \omega_y(s)$ and logit links for $\delta(s), a(s)$.

Figure 7 reports the estimated CLIC and CBIC values of the fitted models, computed using the direct approach outlined in Section 4.1. These two diagnostics agree on at least two main conclusions:

(i) **Mixture models fit generally better**, although they have three more parameters than their non-mixture counterparts. Often, one mixture component is very rough with an estimated smoothness parameter $\hat{\alpha}$ very close to zero (nugget effect), while the other component is much smoother with $\hat{\alpha} \approx 1.4–2$. The smooth component tends to be dominant in flat regions, while the rough component takes over in the mountainous region.

(ii) **Altitude is a major covariate** to be considered in the modeling of extremal dependence, whereas inclusion of further covariates (longitude or latitude) does not improve the fit by much. In non-mixture models, there is a huge drop in CLIC or CBIC values between model 1 (stationary isotropic model with three parameters) and model 3 (locally isotropic model, including altitude as a covariare, with only four parameters). In mixture models, point (i) underscores the importance of altitude as a covariate controlling the importance of the nugget effect over space.
Figure 7: Difference of estimated CLIC (left) and CBIC (right) values for all max-stable models fitted with respect to the best fit. Vertical blue lines mark the separation between mixture (1–11) and non-mixture (12–22) models. The large CLIC/CBIC values for models 10 and 18 are probably due to numerical instabilities in the computation of the sandwich matrices in (17), since altitude and longitude act as confounding factors.

Among non-mixture models, it is worth considering non-stationary non-isotropic models with covariates included in the dependence ranges $\omega_x(s)$, $\omega_y(s)$ and the anisotropy parameter $\delta(s)$. The best non-mixture model is model 8 (respectively 6) according to the CLIC (respectively CBIC). For mixture models with altitude included in the mixture coefficient $a(s)$, use of further covariates in $\omega_x(s)$, $\omega_y(s)$, $\delta(s)$ does not improve the fit by much, although both diagnostics agree to select the most heavily parametrized model as the best model.

Figure 8 displays bivariate kernel density estimators for the pairs of points corresponding to empirical and fitted extremal coefficients, for model 1 (stationary isotropic model, our benchmark), model 6 (best non-mixture model according to the CBIC) and model 22 (best mixture model). Empirical estimates are calculated using the projection method of Marcon et al. (2014) based on the non-parametric Pickands dependence estimator of Capéraà et al. (1997). All three models 1, 6 and 22 show that extremal dependence is slightly underesti-
Figure 8: Bivariate kernel density estimators for pairs of empirical and fitted extremal coefficients, displayed for (left) model 1 (stationary isotropic extremal model), (middle) model 6 (best non-mixture model) and (right) model 22 (best mixture model). A good fit should have points concentrated around the white diagonal line.

Estimated (with a majority of points lying above the diagonal line), but extremal coefficients for non-stationary models tend to be more spread and generally closer to the diagonal. The sum of squared distances between fitted and empirical extremal coefficients is 14.1, 12.4, 10.8 for models 1, 6, 22, respectively. Clearly, the stationary isotropic model provides the worse fit, which confirms our previous conclusions, and even more strongly supports the need for non-stationary dependence structures to incorporate meaningful covariates.

Estimated parameters and standard errors are reported in Table 3 for models 1, 6, and 22. For all models, the estimated degree of freedom is about 4, suggesting that the process is far from being of Brown–Resnick type and that extremal dependence is rather strong at large distances. The estimated process for models 1 and 6 is quite rough (especially for model 1), given their low smoothness parameter. This might suggest that some remaining latent dependence structure has not been captured by these models. Regression coefficients for model 6 are all significantly positive at the 95% confidence level, except for the term $\delta(s)$, which is more difficult to estimate; in particular, the estimate for the altitude coefficients are $-1.9$ (respectively $-1.5$) for $\omega_x(s)$ (respectively $\omega_y(s)$), meaning that “storms” are about 80% smaller 1000 m higher. Regarding model 22, none of the regression coeffi-
Table 3: Estimated parameters (standard errors) for models 1, 6 and 22. Covariates included in the different regression terms are altitude (in km above mean sea level), and longitude/latitude (in km with respect the point of coordinates (−106.6° W, 39° N) lying on the continental divide). Logarithmic links are used for \( \omega_x(s), \omega_y(s) \) and logit links for \( \delta(s), a(s) \).

| #  | Regression terms | Smoothness | df  |
|----|------------------|------------|-----|
| 1  | \( \omega_x(s) \) | Intercept  | Altitude | Longitude | Latitude |
|    |                  | 4.7 (1.4)  | -1.9 (0.3) | -0.01 (0.6) | -0.05 (0.1) |
| 6  | \( \omega_x(s) \) |            | -1.5 (0.3) |          |          |
|    | \( \omega_y(s) \) |            | -0.32 (1.4) | 0.29 (0.4) |          |
|    | \( \delta(s) \)   | 0.3 (0.5)  |          |          |          |
| 22 | \( \omega_x(s) \) | 3.8 (2.7)  | -1.2 (0.8) | -0.01 (0.6) | -0.05 (0.1) |
|    | \( \omega_y(s) \) | 2.8 (10.6) | 0.6 (4.8)  | -0.32 (1.4) | 0.29 (0.4) |
|    | \( \delta(s) \)   | 1.5 (29.6) | -1.1 (13.2) | -0.02 (3.7) | 0.18 (2.0) |
|    | \( a(s) \)        | -1.9 (0.3) | 1.1 (0.2)  |          |          |

Coefficients for \( \omega_x(s), \omega_y(s), \delta(s) \) are significantly positive, suggesting that this model is definitely over-parametrized and that a direct interpretation of these coefficients might be misleading. By contrast, the mixture coefficient \( a(s) \) and the smoothness parameters \( \alpha_1, \alpha_2 \) are quite well identified. As mentioned above, the fitted model has a very rough component that is dominant in the mountains, and a very smooth component that is dominant in the flat regions. Specifically, the proportion of the nugget effect is about 10%, 33%, 64%, 85% at 1000 m, 2000 m, 3000 m, 4000 m, respectively.

6 Discussion

The problem of building and fitting sensible non-stationary dependence models for spatial extremes is not trivial. We have tackled this problem by proposing a very general construction, combining max-stable processes (in particular the extremal \( t \) model), non-stationary correlation functions, and mixtures. The advocated locally elliptic model is based on Paciorek and Schervish (2006) and allows various non-stationary patterns to be flexibly captured in the extremal dependence structure by incorporating meaningful covariates. We have performed inference using pairwise likelihoods, which are computationally convenient, and we have shown by simulation that pairwise likelihoods can efficiently estimate the unknown
parameters, provided that the station network is dense. However, more efficient approaches based on full likelihoods (Stephenson and Tawn 2005; Wadsworth and Tawn 2014; Thibaud and Opitz 2014) might be devised for the extremal $t$ model.

Various non-stationary max-stable models, including altitude, longitude and latitude as covariates, were fitted to a dataset of temperature maxima in Colorado, and these models were shown to provide a better fit with respect to the traditional stationary and isotropic max-stable counterpart, although there is still room for improvement. In particular, we have identified altitude as an important covariate and have shown that extremal dependence is weaker at higher altitudes. In future work, other covariates, such as the slope or solar radiation, might be used to improve the fit, perhaps from satellite data or regional climate computer models. The development of dimension-reduction techniques, when a large number of covariates is available, is also an open problem for research. Alternatively, more flexible non-stationary models might be constructed from a Bayesian perspective, though inference may be tricky and computationally very intensive if standard Markov chain Monte Carlo algorithms are used.

The creation of models for asymptotic independence, a degenerate case in the max-stable paradigm, is also an important issue when data are non-stationary. One possibility would be to “invert” the non-stationary max-stable models proposed above (see Wadsworth and Tawn 2012; Davison et al. 2013).

Finally, we have focused in this work on componentwise maxima, but more efficient approaches and more detailed modeling might be achieved by considering high threshold exceedances (Huser and Davison 2014). This approach, however, entails additional complications such as the modeling of temporal dependence, the selection of a suitable threshold and the non-validity of extremal models at low levels, which might be even more difficult to handle when the data are non-stationary.
Appendix

A A max-stable process that is locally mixing but not mixing everywhere

Consider two independent stationary simple max-stable processes $Z_1(s)$ and $Z_2(s)$ defined on a common space $S \subset \mathbb{R}^d$. Furthermore, let $S_1, S_2$ denote two half-spaces of $S$, infinite in extent, i.e., $S_1 \cup S_2 = S$, $S_1 \cap S_2 = \emptyset$, and there exist sequences $s_n^1 \in S_1$ and $s_n^2 \in S_2$ such that $\|s_1^1 - s_n^1\| \to \infty$ and $\|s_1^2 - s_n^2\| \to \infty$, as $n \to \infty$. Now, define

$$Z(s) = I(s \in S_1)Z_1(s) + I(s \in S_2)Z_2(s),$$

where $I(\cdot)$ is the indicator function. Suppose also that $Z_1(s)$ is mixing (e.g., the Smith model) and that $Z_2(s)$ is not mixing (e.g., the Schlather model). It can easily be checked that $Z(s)$ is a (non-stationary) simple max-stable process and that its extremal coefficient satisfies

$$\begin{cases} 
\theta(s_1^1, s_1^1) \to 2, \\
\theta(s_1^2, s_2^2) \to l < 2, \quad n \to \infty, \\
\theta(s_1^1, s_2^2) = 2,
\end{cases}$$

(18)

where the last equation also holds when the roles of sequences are interchanged. Since (18) holds for any similarly defined sequences $s_n^1 \in S_1$, $s_n^2 \in S_2$, the process $Z(s)$ is locally mixing in $S_1$, but not in $S_2$. Conclusion: $Z(s)$ is not mixing everywhere.
B  Bivariate margins for the Smith–Stephenson model

B.1 General expression

To derive the bivariate margins for model (5), we need to compute the corresponding exponent measure \( V_D(z_1, z_2) \) given in expression (14), which may be written as

\[
V_D(z_1, z_2) = E \left[ \max \left\{ \frac{\phi_d(s_1 - U; \Omega_1)}{z_1}, \frac{\phi_d(s_2 - U; \Omega_2)}{z_2} \right\} \right],
\]

where, by abuse of notation, we have written \( \Omega_1 = \Omega_{s_1} \) and \( \Omega_2 = \Omega_{s_2} \). Let \( A_1 \) and \( A_2 \) denote the expectations in (19) and (20), respectively, so that we have \( V_D(z_1, z_2) = A_1/z_1 + A_2/z_2 \).

We may write

\[
A_j = \int_{B_j \subset \mathbb{R}^d} \phi_d(s_1 - u; \Omega_1) du = \Pr(N \in B_j^*), \quad j = 1, 2,
\]

where \( N = (N_1, \ldots, N_d) \sim \mathcal{N}(\mathbf{0}, I_d) \) is a vector of \( d \) independent standard Gaussian random variables, and \( B_j^* \) \((j = 1, 2)\) is the set defined through the changes of variables \( v = \Omega_j^{-1/2}(s_j - u) \). To see what the domain \( B_1^* \) looks like, let \( v = \Omega_1^{-1/2}(s_1 - u) \) and \( h = s_2 - s_1 \). The following expressions are equivalent:

\[
\frac{\phi_d(s_1 - u; \Omega_1)}{z_1} > \frac{\phi_d(s_2 - u; \Omega_2)}{z_2} \quad \iff \quad (s_1 - u)^T \Omega_1^{-1} (s_1 - u) < (s_2 - u)^T \Omega_2^{-1} (s_2 - u) + 2 \log \left( \frac{z_2 | \Omega_2 |^{1/2}}{z_1 | \Omega_1 |^{1/2}} \right)
\]

\[
\iff \quad v^T v < (\Omega_1^{1/2}v + h)^T \Omega_2^{-1} (\Omega_1^{1/2}v + h) + 2 \log \left( \frac{z_2 | \Omega_2 |^{1/2}}{z_1 | \Omega_1 |^{1/2}} \right)
\]

\[
\iff \quad v^T P v + Q^T v < L,
\]

where the quantities \( P \in \mathbb{R}^{d \times d} \), \( Q \in \mathbb{R}^d \) and \( L \in \mathbb{R} \) in (22) are defined as

\[
P = \left( I_d - \Omega_1^{T/2} \Omega_2^{-1} \Omega_1^{1/2} \right), \quad Q = -2 \Omega_1^{T/2} \Omega_2^{-1} h, \quad L = h^T \Omega_2^{-1} h + 2 \log \left( \frac{z_2 | \Omega_2 |^{1/2}}{z_1 | \Omega_1 |^{1/2}} \right).
\]
Hence, for $j = 1$, (21) boils down to $A_1 = \Pr (N \in B_1^*) = \Pr (N^T P N + Q^T N < L)$. To evaluate this expression, we need to compute the distribution of quadratic forms of Gaussian variates, which is not obvious. To simplify this, and since $P$ is symmetric, we consider its spectral decomposition, i.e., $P = M \Lambda M^T$ where $M \in \mathbb{R}^{d \times d}$ is an orthogonal basis of eigenvectors, and $\Lambda \in \mathbb{R}^{d \times d}$ is the corresponding diagonal matrix of eigenvalues. By symmetry of the Gaussian density, one has

$$A_1 = \Pr (N \in B_1^*) = \Pr (M N \in B_1^*) = \Pr (N^T \Lambda N + Q^T M N < L). \quad (23)$$

Now, we split the matrix of eigenvalues $\Lambda$ into two parts corresponding to zero and non-zero eigenvalues, namely $\Lambda_0 \in \mathbb{R}^{p_0 \times p_0}$, $\Lambda_* \in \mathbb{R}^{(d-p_0) \times (d-p_0)}$ respectively, with $p_0 \in \{0, \ldots, d\}$. Let also $M_0 \in \mathbb{R}^{d \times p_0}$ and $M_* \in \mathbb{R}^{d \times (d-p_0)}$ denote the corresponding matrices of eigenvectors, and $N_0 \sim \mathcal{N}_{p_0} (0, I_{p_0})$ and $N_* \sim \mathcal{N}_{d-p_0} (0, I_{d-p_0})$ be independent Gaussian variates. Then, equation (23) may be written as

$$A_1 = \Pr \left\{ (N_*^T \Lambda_* N_* + Q^T M_* N_*) + (Q^T M_0 N_0) < L \right\} = \Pr \left\{ (N_* - c)^T \Lambda_* (N_* - c) + \|Q^T M_0\| N < R \right\},$$

where $N \sim \mathcal{N}(0, 1)$, $c = -2^{-1} \Lambda_*^{-1} M_*^T Q = \Lambda_*^{-1} M_*^T \Omega_1^{1/2} \Omega_2^{-1/2} h \in \mathbb{R}^{d-p_0}$ and

$$R = L + \frac{1}{4} Q^T M_* \Lambda_*^{-1} M_*^T Q = h^T \Omega_2^{-1} h + 2 \log \left( \frac{\sigma_2 |\Omega_2|^{1/2}}{\sigma_1 |\Omega_1|^{1/2}} \right) + h^T \Omega_2^{-1} \Omega_1^{1/2} M_* \Lambda_*^{-1} M_*^T \Omega_1^{1/2} \Omega_2^{-1} h.$$

Hence, the required probability is equal to

$$A_1 = \Pr \left( \sum_{j=1}^{d-p_0} \lambda_j c_j^2 + 2 \|h^T \Omega_2^{-1} \Omega_1^{1/2} M_0\| N < R \right), \quad (24)$$

where $\lambda_j$ is the $j$th non-zero eigenvalue of the matrix $P$, $c_j$ is the $j$th element of the vector $c$, and $\chi^2_{\nu; nc}$ is a non-central $\chi^2$ random variate with $\nu$ degrees of freedom and non-centrality parameter $nc$. Let $H_\psi (x)$ be the distribution of $\sum_{j=1}^{k} a_j \chi^2_{\nu, j; nc_j} + \tau N$, where $\psi^T = (a^T, \nu^T, nc^T, \tau)$ is the vector of parameters. Then, $A_1 = H_\psi (R)$ with $a^T = (\lambda_1, \ldots, \lambda_{d-p_0})$, 31
\( \mathbf{v}^T = (1, \ldots, 1), \mathbf{n} \mathbf{c}^T = (c_1^2, \ldots, c_{d-p_0}^2) \) and \( \tau = 2\|h^T \mathbf{\Omega}_2^{-1} \mathbf{\Omega}_1^{1/2} \mathbf{M}_0\|. \) The computation of \( H_{\psi}(x) \) is difficult in full generality, but might be approximated using saddlepoint approximations (Kuonen 1999), \( \chi^2 \) approximations (Liu et al. 2009) or numerical inversions of characteristic functions (Davies 1973, 1980). The computation of \( A_2 \) in (21) can be performed similarly (by symmetry of the arguments), and the exponent measure may be written as \( V_D(z_1, z_2) = z_1^{-1} H_{\psi_{1,2}}(R_{1,2}) + z_2^{-1} H_{\psi_{2,1}}(R_{2,1}) \), where \( \psi_{1,2}, R_{1,2} \) are defined above without subscripts, and \( \psi_{2,1}, R_{2,1} \) are deduced by a simple interchange of labels. Let \( h_\psi(x) \) and \( h'_\psi(x) \) denote the density function corresponding to \( H_\psi(x) \) and its derivative, respectively. The partial derivatives of the exponent measure are

\[
- \frac{\partial V_D(z_1, z_2)}{\partial z_1} = \frac{1}{z_1} H_{\psi_{1,2}}(R_{1,2}) + \frac{2}{z_1^2} h_{\psi_{1,2}}(R_{1,2}) - \frac{2}{z_1 z_2} h_{\psi_{2,1}}(R_{2,1}),
\]

\[
- \frac{\partial V_D(z_1, z_2)}{\partial z_2} = \frac{1}{z_2} H_{\psi_{2,1}}(R_{2,1}) + \frac{2}{z_2^2} h_{\psi_{2,1}}(R_{2,1}) - \frac{2}{z_1 z_2} h_{\psi_{1,2}}(R_{1,2}),
\]

\[
- \frac{\partial^2 V_D(z_1, z_2)}{\partial z_1 \partial z_2} = \frac{2}{z_1 z_2} h_{\psi_{1,2}}(R_{1,2}) + \frac{4}{z_1^2 z_2} h'_{\psi_{1,2}}(R_{1,2}) + \frac{2}{z_1 z_2^2} h_{\psi_{2,1}}(R_{2,1}) + \frac{4}{z_1^2 z_2} h'_{\psi_{2,1}}(R_{2,1}),
\]

and the bivariate density corresponding to (3) may be written as

\[
\frac{\partial^2 \exp\{-V_D(z_1, z_2)\}}{\partial z_1 \partial z_2} = \left\{ \frac{\partial V_D(z_1, z_2)}{\partial z_1} \frac{\partial V_D(z_1, z_2)}{\partial z_2} - \frac{\partial^2 V_D(z_1, z_2)}{\partial z_1 \partial z_2} \right\} \exp\{-V_D(z_1, z_2)\}.
\]

**B.2 Special cases**

**Stationary case.** When \( \mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega} \), the matrix \( \mathbf{P} \) in (22) vanishes, so all its eigenvalues are zero and \( \mathbf{M} = \mathbf{I}_d \). Hence, \( p_* = 0, p_0 = d \), and the probability in (24) reduces to

\[
A_1 = \Pr \left\{ 2\|h^T \mathbf{\Omega}^{-1/2}\|N < h^T \mathbf{\Omega}^{-1} h + 2 \log \left( \frac{z_2}{z_1} \right) \right\}
\]

\[
= \Phi \left\{ \frac{(h^T \mathbf{\Omega}^{-1} h)^{1/2}}{2} + \frac{1}{(h^T \mathbf{\Omega}^{-1} h)^{1/2}} \log \left( \frac{z_2}{z_1} \right) \right\},
\]

where \( \Phi(\cdot) \) denotes the standard Gaussian distribution, recovering the expressions given in Smith (1990). The term \( A_2 \) in (20) is found by symmetry. The extremal coefficient is \( \theta(s_1, s_2) = 2 \Phi\{ (h^T \mathbf{\Omega}^{-1} h)^{1/2}/2 \} \). If \( \mathbf{\Omega} = \omega^2 \mathbf{I}_d \), the model is stationary isotropic with \( \theta(s_1, s_2) = 2 \Phi\{ \|h\|/(2\omega) \} \), as illustrated in the top left panel of Figure 1.
Non-stationary, locally isotropic case. When $\Omega_1 = \omega_1^2 I_d$, $\Omega_2 = \omega_2^2 I_d$, with $\omega_1, \omega_2 > 0$ and $\omega_1 \neq \omega_2$, the matrix $P$ defined in (22) is $P = (1 - \omega_1^2 \omega_2^{-2}) I_d$. Therefore, its eigenvalues $\lambda_1, \ldots, \lambda_d$ are all equal to $1 - \omega_1^2 \omega_2^{-2} \neq 0$, so $p_* = d$, $p_0 = 0$. Hence, the probability in (24) reduces to

$$A_1 = \Pr \left\{ (1 - \omega_1^2 \omega_2^{-2}) \chi^2_{d, ||c||^2} < \frac{||h||^2}{\omega_2^2 - \omega_1^2} + 2 \log \left( \frac{z_2 \omega_2^d}{z_1 \omega_1^d} \right) \right\}$$

$$= I(\omega_1 > \omega_2) + (-1)^{I(\omega_1 > \omega_2)} \text{CHI}^2_{d, ||c||^2} \left[ \frac{\omega_2^2}{\omega_2^2 - \omega_1^2} \left\{ \frac{||h||^2}{\omega_2^2 - \omega_1^2} + 2 \log \left( \frac{z_2 \omega_2^d}{z_1 \omega_1^d} \right) \right\} \right],$$

where $\text{CHI}^2_{d, ||c||^2}(\cdot)$ is the distribution function of $\chi^2_{d, ||c||^2}$, $||c||^2 = ||h||^2 \omega_1^2 (\omega_2^2 - \omega_1^2)^{-2}$, and $I(\cdot)$ is the indicator function. The term $A_2$ in (20) can be deduced by interchanging the labels. These expressions match those found in Smith and Stephenson (2009). The extremal coefficient may be expressed as

$$\theta(s_1, s_2) = 1 + \text{CHI}^2_{d, ||h||^2 \min_{1,2}} \left[ \frac{\max_{1,2}}{\max_{1,2}^2 - \min_{1,2}^2} \left\{ \frac{||h||^2}{\max_{1,2}^2 - \min_{1,2}^2} + 2 \log \left( \frac{\max_{1,2}}{\min_{1,2}} \right) \right\} \right]$$

$$- \text{CHI}^2_{d, ||h||^2 \max_{1,2}} \left[ \frac{\min_{1,2}}{\max_{1,2}^2 - \min_{1,2}^2} \left\{ \frac{||h||^2}{\max_{1,2}^2 - \min_{1,2}^2} + 2 \log \left( \frac{\max_{1,2}}{\min_{1,2}} \right) \right\} \right],$$

where $\max_{1,2} = \max(\omega_1, \omega_2)$ and $\min_{1,2} = \min(\omega_1, \omega_2)$. The model is illustrated in the top right panel of Figure 1.

Non-stationary, homogeneously anisotropic case. When $\Omega_1 = \omega_1^2 R$ and $\Omega_2 = \omega_2^2 R$, and $R \in \mathbb{R}^{d \times d}$ is a correlation matrix, the matrix $P$ defined in (22) is $P = (1 - \omega_1^2 \omega_2^{-2}) I_d$. Hence, the probability in (24) is of the same form as (25), except that the Euclidean distance $||h||$ is replaced by the Mahalanobis distance $(h^T R^{-1} h)^{1/2}$, and likewise for the extremal coefficient. See also Smith and Stephenson (2009). This model is illustrated in the bottom left panel of Figure 1.
C Mixing properties of the Smith–Stephenson model

For simplicity, we consider the locally isotropic model with \( \Omega_s = \omega^2(s)I_d \) in (5). For the homogeneously anisotropic case with \( \Omega_s = \omega^2(s)R \), where \( R \) is a correlation matrix, simply notice that \( (h_n^T R^{-1} h_n)^{1/2} = O(\|h_n\|) \), and then follow the same arguments. To show that the locally isotropic model is mixing everywhere if and only if \( \omega(s) = o(\|s\|) \), we fix a point \( s_0 \in S \) and an admissible trajectory \( s_n \in S \) (i.e., \( h_n = \|s_n - s_0\| \to \infty \)), and we check that the extremal coefficient satisfies \( \theta(s_0, s_n) \to 2 \), as \( n \to \infty \), if and only if \( \omega(s) = o(\|s\|) \).

Let \( \omega_0 = \omega(s_0) \) and \( \omega_n = \omega(s_n) \), and without loss of generality, assume that \( \omega_n > \omega_0 \). Since \( \chi^2_{\nu,nc} \) is equal in distribution to the random variable \( nc + 2\sqrt{nc} \mathcal{N}(0, 1) + \chi^2_{\nu} \), one has

\[
\theta(s_0, s_n) = 1 + \Pr \left[ \left( \frac{2\|h_n\|s_0 \mathcal{N}(0, 1)}{\omega_n^2 - \omega_0^2} \right) + \chi^2_d \leq \frac{\omega_n^2}{\omega_n^2 - \omega_0^2} + 2d \log \left( \frac{\omega_n}{\omega_0} \right) \right] - \Pr \left[ \frac{2\|h_n\|s_0 \mathcal{N}(0, 1)}{\omega_n^2 - \omega_0^2} + \chi^2_d \leq \frac{\omega_n^2}{\omega_n^2 - \omega_0^2} + 2d \log \left( \frac{\omega_n}{\omega_0} \right) \right] \]

\[
= 1 + \Pr \left\{ \mathcal{N}(0, 1) \geq \frac{2\|h_n\|s_0 \mathcal{N}(0, 1)}{\omega_n^2 - \omega_0^2} \right\} + \frac{(\omega_n^2 - \omega_0^2)\chi^2_d}{2\|h_n\|s_0} \leq \frac{\|h_n\|^2}{2\omega_0} + \frac{d\omega_n^2}{\|h_n\|s_0} \log \left( \frac{\omega_n}{\omega_0} \right) \}

(26)

(27)

(28)

(29)

Now, since \( \omega_n > \omega_0 > 0 \), we have that \( \omega_n \to 0 \), as \( n \to \infty \). We have the following cases:

1. Assume that \( \omega(s) = o(\|s\|^{1/2}) \), i.e., \( \omega(\|s\|^{-1/2} \to 0 \), as \( \|s\| \to \infty \). This implies that \( \omega_n\|h_n\|^{-1/2} \to 0 \), as \( n \to \infty \). Then, by Slutsky’s Theorem, \( (\omega_n^2 - \omega_0^2)\chi^2_d/(2\|h_n\|s_0) \to 0 \) and \( (\omega_n^2 - \omega_0^2)\chi^2_d/(2\|h_n\|s_0) \to 0 \). The probabilities in (28) and (29) thus converge to 1 and 0, respectively. Hence, \( \theta(s_0, s_n) \to 2 \), as \( n \to \infty \).

2. Assume that \( \omega(s) = o(\|s\|) \), but \( \omega(\|s\|^{-1/2}) \to \infty \), as \( \|s\| \to \infty \). This implies that \( \omega_n\|h_n\|^{-1} \to 0 \) and \( \omega_n\|h_n\|^{-1/2} \to \infty \), as \( n \to \infty \). Then, by Slutsky’s Theorem, \( 2\|h_n\|s_0 \mathcal{N}(0, 1)/(\omega_n^2 - \omega_0^2) \to 0 \) and \( (\omega_n^2 - \omega_0^2)\chi^2_d/(2\|h_n\|s_0) \to 0 \). The probabi-
ties in (26) and (29) thus converge to 1 and 0, respectively. Hence, \( \theta(s_0, s_n) \rightarrow 2 \), as \( n \rightarrow \infty \).

3. Assume that \( \omega(s) = O(\|s\|) \), i.e., \( \omega(s)\|s\|^{-1} \rightarrow C > 0 \), as \( \|s\| \rightarrow \infty \). This implies that \( \omega_n \|h_n\|^{-1} \rightarrow C \), as \( n \rightarrow \infty \). Then, by Slutsky’s Theorem, \( 2\|h_n\|\mathcal{N}(0, 1)/(\omega_n^2 - \omega_0^2) \xrightarrow{P} 0 \), \( 2\|h_n\|\omega_n \mathcal{N}(0, 1)/(\omega_n^2 - \omega_0^2) \xrightarrow{D} \mathcal{N}(0, 4C^{-2}) \). The probabilities in (26) and (27) thus converge to 1 and \( \Pr\{\mathcal{N}(0, 4C^{-2}) + \chi^2_d \leq -C^{-2}\} > 0 \), respectively. Hence, \( \theta(s_0, s_n) \rightarrow l \), with \( 1 < l < 2 \), as \( n \rightarrow \infty \).

4. Assume that \( \omega(s)\|s\|^{-1} \rightarrow \infty \), as \( \|s\| \rightarrow \infty \). This implies that \( \omega_n \|h_n\|^{-1} \rightarrow \infty \), as \( n \rightarrow \infty \). Then, by Slutsky’s Theorem, one has \( 2\|h_n\|\omega_0 \mathcal{N}(0, 1)/(\omega_n^2 - \omega_0^2) \xrightarrow{P} 0 \), \( 2\|h_n\|\omega_n \mathcal{N}(0, 1)/(\omega_n^2 - \omega_0^2) \xrightarrow{P} 0 \). Both probabilities in (26) and (27) converge to 1. Hence, \( \theta(s_0, s_n) \rightarrow 1 \), as \( n \rightarrow \infty \).

Similar arguments may be applied to treat the case \( \omega_0 > \omega_n \rightarrow 0 \), as \( n \rightarrow \infty \), and we find that in this case \( \theta(s_0, s_n) \rightarrow 2 \), as \( n \rightarrow \infty \). Therefore, since \( s_0 \) and \( s_n \) are chosen arbitrarily, the process is mixing everywhere if and only if \( \omega(s) = o(\|s\|) \). In particular, this is the case if \( \omega(s) \) is uniformly bounded on \( S \). By contrast, if \( \omega(s) \) tends to infinity faster than \( \|s\| \), events occurring at \( s_0 \) and \( s_n \) become more and more dependent as their distance increases and they are eventually perfectly dependent (at infinity).
### D Summary of max-stable models fitted to temperature annual maxima

Table 4: Extremal $t$ max-stable models fitted to the temperature maxima. For each of these models, we report whether they are stationary (Stat.), locally isotropic (Iso.), and based on mixtures of correlation functions (Mix.). If a model is non-stationary, altitude (Alt.), longitude (Lon.) and latitude (Lat.) may be used as covariates in the dependence ranges $\omega_x(s), \omega_y(s)$, the anisotropy parameter $\delta(s)$ and the mixture coefficient $a(s)$; recall \([13]\) and \([15]\). If a model is locally isotropic, $\omega_x(s) = \omega_y(s)$ and $\delta(s) = 0$. Mixture models are constructed from two correlation functions of the form \([9]\) combined with \([11]\), with different smoothness parameters $\alpha_1, \alpha_2$, but based on the same matrix $\Omega_s$. The total number of parameters (Nb. par.) is also reported.

| #  | Stat. | Iso. | Mix. | $\omega_x(s)$ | $\omega_y(s)$ | $\delta(s)$ | $a(s)$ | Nb. par. |
|----|-------|------|------|----------------|----------------|----------------|--------|----------|
| 1  | Yes   | Yes  | No   | none           | —              | —              | —      | 3        |
| 2  | Yes   | No   | No   | none           | none           | none           | —      | 5        |
| 3  | No    | Yes  | No   | Alt.           | —              | —              | —      | 4        |
| 4  | No    | Yes  | No   | Alt./Lon.      | —              | —              | —      | 5        |
| 5  | No    | Yes  | No   | Alt./Lon./Lat. | —              | —              | —      | 6        |
| 6  | No    | No   | No   | Alt.           | Alt.           | none           | —      | 7        |
| 7  | No    | No   | No   | Alt./Lon.      | Alt./Lon.      | none           | —      | 9        |
| 8  | No    | No   | No   | Alt./Lon./Lat. | Alt./Lon./Lat. | none           | —      | 11       |
| 9  | No    | No   | No   | Alt.           | Alt.           | Alt.           | —      | 8        |
| 10 | No    | No   | No   | Alt./Lon.      | Alt./Lon.      | Alt./Lon.      | —      | 11       |
| 11 | No    | No   | No   | Alt./Lon./Lat. | Alt./Lon./Lat. | Alt./Lon./Lat. | —      | 14       |
| 12 | No    | Yes  | Yes  | none           | —              | —              | Alt.   | 6        |
| 13 | No    | No   | Yes  | none           | none           | none           | Alt.   | 8        |
| 14 | No    | Yes  | Yes  | Alt.           | —              | —              | Alt.   | 7        |
| 15 | No    | Yes  | Yes  | Alt./Lon.      | —              | —              | Alt.   | 8        |
| 16 | No    | Yes  | Yes  | Alt./Lon./Lat. | —              | —              | Alt.   | 9        |
| 17 | No    | No   | Yes  | Alt.           | Alt.           | none           | Alt.   | 10       |
| 18 | No    | No   | Yes  | Alt./Lon.      | Alt./Lon.      | none           | Alt.   | 12       |
| 19 | No    | No   | Yes  | Alt./Lon./Lat. | Alt./Lon./Lat. | none           | Alt.   | 14       |
| 20 | No    | No   | Yes  | Alt.           | Alt.           | Alt.           | Alt.   | 11       |
| 21 | No    | No   | Yes  | Alt./Lon.      | Alt./Lon.      | Alt./Lon.      | Alt.   | 14       |
| 22 | No    | No   | Yes  | Alt./Lon./Lat. | Alt./Lon./Lat. | Alt./Lon./Lat. | Alt.   | 17       |
References

Anderes, E. B. and Stein, M. L. (2008) Estimating Deformations of Isotropic Gaussian Random Fields on the Plane. *Annals of Statistics* 36(2), 719–741.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Chichester: Wiley. ISBN 9780471976479.

Blanchet, J. and Davison, A. C. (2011) Spatial Modelling of Extreme Snow Depth. *Annals of Applied Statistics* 5(3), 1699–1725.

Brown, B. M. and Resnick, S. I. (1977) Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability* 14(4), 732–739.

Capéraà, P., Fougères, A.-L. and Genest, C. (1997) A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas. *Biometrika* 84(3), 567–577.

Castruccio, S. and Genton, M. G. (2014) Emulation of Global 3D Spatio-Temporal Temperature: A Distributed Computing Approach to Model One Billion Data Points. Under review.

Castruccio, S., Huser, R. and Genton, M. G. (2014) High-Order Composite Likelihood Inference for Multivariate or Spatial Extremes. arXiv:1411.0086v1.

Castruccio, S. and Stein, M. L. (2013) Global Space-Time Models for Climate Ensembles. *Annals of Applied Statistics* 7(3), 1593–1611.

Chavez-Demoulin, V. and Davison, A. C. (2005) Generalized Additive Modelling of Sample Extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(1), 207–222.

Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. London: Springer. ISBN 9781852334598.

Cooley, D. S., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt-Shamseldin, E. C., Omolo, B. O. and Sun, Y. (2012) A Survey of Spatial Extremes: Measuring Spatial Dependence and Modeling Spatial Effects. *REVSTAT* 10(1), 135–165.

Cooley, D. S., Naveau, P. and Nychka, D. (2007) Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of American Statistical Association* 102(479), 824–840.

Cooley, D. S., Naveau, P. and Poncet, P. (2006) Variograms for Spatial Max-Stable Random Fields. In *Dependence in Probability and Statistics*, eds P. Bertail, P. Doukhan and
Davies, R. B. (1973) Numerical Inversion of a Characteristic Function. *Biometrika* 60(2), 415–417.

Davies, R. B. (1980) Algorithm AS 155: The Distribution of a Linear Combination of $\chi^2$ Random Variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29(1980), 323–333.

Davis, R. A., Klüppelberg, C. and Steinkohl, C. (2013) Statistical Inference for Max-Stable Processes in Space and Time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(5).

Davison, A. C. and Gholamrezaee, M. M. (2012) Geostatistics of Extremes. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences* 468(2138), 581–608.

Davison, A. C. and Huser, R. (2014) Statistics of Extremes. *Annual Review of Statistics and its Application*. Accepted.

Davison, A. C., Huser, R. and Thibaud, E. (2013) Geostatistics of Dependent and Asymptotically Independent Extremes. *Mathematical Geosciences* 45(5), 511–529.

Davison, A. C., Padoan, S. and Ribatet, M. (2012) Statistical Modelling of Spatial Extremes (with Discussion). *Statistical Science* 27(2), 161–186.

Demarta, S. and McNeil, A. J. (2005) The $t$ Copula and Related Copulas. *International Statistical Review* 73(1), 111–129.

Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2014) Estimation of Hüsler–Reiss Distributions and Brown–Resnick Processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* To appear.

Fuentes, M. (2001) A High Frequency Kriging Approach for Non-Stationary Environmental Processes. *Environmetrics* 12(5), 469–483.

Genton, M. G., Ma, Y. and Sang, H. (2011) On the Likelihood Function of Gaussian Max-Stable Processes. *Biometrika* 98(2), 481–488.

Gneiting, T. and Guttorp, P. (2010) Continuous-Parameter Stochastic Process Theory. In *Handbook of Spatial Statistics*, eds A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, pp. 17–28. Chapman & Hall.

Guttorp, P. and Gneiting, T. (2006) Studies in the History of Probability and Statistics XLIX On the Matérn Correlation Family. *Biometrika* 93(4), 989–995.
de Haan, L. (1984) A Spectral Representation for Max-Stable Processes. *Annals of Probability* **12**(4), 1194–1204.

de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction*. New York: Springer. ISBN 9780387239460.

Handcock, M. S. and Stein, M. L. (1993) A Bayesian Analysis of Kriging. *Technometrics* **35**(4), 403–410.

Huser, R. and Davison, A. C. (2013) Composite Likelihood Estimation for the Brown–Resnick Process. *Biometrika* **100**(2), 511–518.

Huser, R. and Davison, A. C. (2014) Space-Time Modelling of Extreme Events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 439–461.

Hüsler, J. and Reiss, R.-D. (1989) Maxima of Normal Random Vectors: Between Independence and Complete Dependence. *Statistics & Probability Letters* **7**(4), 283–286.

Jeon, S. and Smith, R. L. (2012) Dependence Structure of Spatial Extremes Using Threshold Approach. arXiv:1209.6344v1.

Jun, M. and Stein, M. L. (2007) An Approach to Producing Space-Time Covariance Functions on Spheres. *Technometrics* **49**(4), 468–479.

Jun, M. and Stein, M. L. (2008) Nonstationary Covariance Models for Global Data. *Annals of Applied Statistics* **2**(4), 1271–1289.

Kabluchko, Z. and Schlather, M. (2010) Ergodic Properties of Max-Infinite Divisible Processes. *Stochastic Processes and their Applications* **120**(3), 281–295.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary Max-Stable Fields Associated to Negative Definite Functions. *Annals of Probability* **37**(5), 2042–2065.

Kuonen, D. (1999) Saddlepoint Approximations for Distributions of Quadratic Forms in Normal Variables. *Biometrika* **86**(4), 929–935.

Lantuéjoul, C., Bacro, J.-N. and Bel, L. (2011) Storm Processes and Stochastic Geometry. *Extremes* **14**(4), 413–428.

Lindgren, F., Rue, H. and Lindström, J. (2011) An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.
Lindsay, B. G. (1988) Composite Likelihood Methods. *Contemporary Mathematics* **80**, 221–239.

Liu, H., Tang, Y. and Zhang, H. H. (2009) A New Chi-Square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables. *Computational Statistics & Data Analysis* **53**(4), 853–856.

Marcon, G., Padoan, S. A., Naveau, P. and Muliere, P. (2014) Multivariate Nonparametric Estimation of the Pickands Dependence Function Using Bernstein Polynomials. arXiv:1405.5228v1.

Ng, C. T. and Joe, H. (2014) Model Comparison with Composite Likelihood Information Criteria. *Bernoulli* **20**(4), 1738–1764.

Nikoloulopoulos, A. K., Joe, H. and Li, H. (2009) Extreme Value Properties of Multivariate $t$ Copulas. *Extremes* **12**(2), 129–148.

Northrop, P. J. and Jonathan, P. (2011) Threshold Modelling of Spatially-Dependent Non-Stationary Extremes with Application to Hurricane-Induced Wave Heights (with Discussion). *Environmetrics* **22**, 799–809.

Nychka, D., Wikle, C. K. and Royle, J. A. (2002) Multiresolution Models for Nonstationary Spatial Covariance Functions. *Statistical Modelling* **2**(4), 315–331.

Opitz, T. (2013) Extremal $t$ Processes: Elliptical Domain of Attraction and a Spectral Representation. *Journal of Multivariate Analysis* **122**(1), 409–413.

Paciorek, C. J. and Schervish, M. (2006) Spatial Modelling Using a New Class of Nonstationary Covariance Functions. *Environmetrics* **17**(5), 483–506.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association* **105**(489), 263–277.

Perrin, O. and Monestiez, P. (1999) Modelling of Non-Stationary Spatial Structure Using Parametric Radial Basis Deformations. In *geoENV II – Geostatistics for Environmental Applications*, eds J. Gómez-Hernández, A. Soares and R. Froidevaux, volume 10 of *Quantitative Geology and Geostatistics*, pp. 175–186. Springer.

Reich, B. J., Eidsvik, J., Guindani, M., Nail, A. J. and Schmidt, A. M. (2011) A Class of Covariate-Dependent Spatiotemporal Covariance Functions For The Analysis of Daily Ozone Concentration. *Annals of Applied Statistics* **5**(4), 2465–2487.

Reich, B. J. and Shaby, B. A. (2012) A Hierarchical Max-Stable Spatial Model for Extreme Precipitation. *Annals of Applied Statistics* **6**(4), 1430–1451.
Ribatet, M. (2013) Spatial Extremes: Max-Stable Processes at Work. *Journal de la Société Française de Statistique* **154**(2), 156–177.

Ribatet, M., Cooley, D. S. and Davison, A. C. (2012) Bayesian Inference from Composite Likelihoods, with an Application to Spatial Extremes. *Statistica Sinica* **22**(2), 813–845.

Sampson, P. D. and Guttorp, P. (1992) Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association* **87**(417), 108–119.

Sang, H. and Gelfand, A. (2009) Hierarchical Modeling for Extreme Values Observed over Space and Time. *Environmental and Ecological Statistics* **16**(3), 407–426.

Schlather, M. (2002) Models for Stationary Max-Stable Random Fields. *Extremes* **5**(1), 33–44.

Schmidt, A. M. and O’Hagan, A. (2003) Bayesian Inference for Non-Stationary Spatial Covariance Structure via Spatial Deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(3), 743–758.

Smith, E. L. and Stephenson, A. G. (2009) An Extended Gaussian Max-Stable Process Model for Spatial Extremes. *Journal of Statistical Planning and Inference* **139**(4), 1266–1275.

Smith, R. L. (1990) Max-Stable Processes and Spatial Extremes. Unpublished.

Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. First edition. New York: Springer. ISBN 9780387986296.

Stephenson, A. and Tawn, J. A. (2005) Exploiting Occurrence Times in Likelihood Inference for Componentwise Maxima. *Biometrika* **92**(1), 213–227.

Thibaud, E., Mutzner, R. and Davison, A. C. (2013) Threshold Modeling of Extreme Spatial Rainfall. *Water Resources Research* **49**(8), 4633–4644.

Thibaud, E. and Opitz, T. (2014) Efficient Inference and Simulation for Elliptical Pareto Processes. arXiv:1401.0168v1.

Varin, C., Reid, N. and Firth, D. (2011) An Overview of Composite Likelihood Methods. *Statistica Sinica* **21**(2011), 5–42.

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence Modelling for Spatial Extremes. *Biometrika* **99**(2), 253–272.

Wadsworth, J. L. and Tawn, J. A. (2014) Efficient Inference for Spatial Extreme Value Processes Associated to Log-Gaussian Random Functions. *Biometrika* **101**(1), 1–15.