

Higher-dimensional spatial extremes via single-site conditioning

J. L. Wadsworth and J. A. Tawn

Department of Mathematics and Statistics, Fylde College, Lancaster University, LA1 4YF

Abstract

Currently available models for spatial extremes suffer either from inflexibility in the dependence structures that they can capture, lack of scalability to high dimensions, or in most cases, both of these. We present an approach to spatial extreme value theory based on the conditional multivariate extreme value model, whereby the limit theory is formed through conditioning upon the value at a particular site being extreme. The ensuing methodology allows for a flexible class of dependence structures, as well as models that can be fitted in high dimensions. To overcome issues of conditioning on a single site, we suggest a joint inference scheme based on all observation locations, and implement an importance sampling algorithm to provide spatial realizations and estimates of quantities conditioning upon the process being extreme at any of one of an arbitrary set of locations. The modelling approach is applied to Australian summer temperature extremes, permitting assessment of the spatial extent of high temperature events over the continent.

Keywords: asymptotic independence; conditional extreme value model; extremal dependence; importance sampling; Pareto process; spatial modelling

1 Introduction

1.1 Background

In this work, our main motivation is to facilitate modelling of a continuous spatial process $\{Y(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ when it reaches extreme levels. We assume that n replicates of the process $Y_i, i = 1, \dots, n$ are available in time, and that observations have been collected at a finite set of locations s_1, \dots, s_d . To understand the risk posed by extreme events of Y , one firstly seeks general characterizations of spatial stochastic processes, given that some facet of them is extreme. These characterizations in turn yield statistical models that can be applied to appropriately selected extreme data, and used to investigate the probabilities of events that are more extreme than those observed to date.

Spatial extreme value theory has attracted much attention in recent years. A large tranche of literature deals with *max-stable processes*, which arise through considering the limits of affinely normalized maxima $\{M_n(s) = \max_{1 \leq i \leq n} [Y_i(s) - b_n(s)]/a_n(s) : s \in \mathcal{S}\}$ as the number of independent or weakly dependent copies n tends to infinity. The functions $a_n(s) > 0, b_n(s)$ are related only to the marginal distributions of Y ; for details of this approach see e.g. Davison et al. (2012) or de Haan and Ferreira (2007, Chapter 9). However, applications where the object of interest really is the spatial pointwise maximum, taken over a large number of repetitions, are relatively scarce. Moreover, inference for max-stable processes is notoriously difficult due to the fact that the pointwise maximum function M_n , and its limiting counterpart, is composed of a number of different underlying processes Y_i . This can be alleviated using the approach of Stephenson and Tawn (2005) that includes information on which Y_i contribute to M_n , but this can lead to bias if d is relatively large compared to n (Wadsworth, 2015). Full likelihood inference ignoring this information, either via data augmentation (Thibaud et al., 2016) or Monte Carlo expectation maximization (Huser et al., 2019) is still difficult to implement and remains limited to moderate (up to ≈ 20) numbers of locations.

The analog of max-stable processes for suitable definitions of threshold exceedances is Pareto processes (Ferreira and de Haan, 2014). This characterization focuses on the behaviour of $\{[Y(s) - b_n(s)]/a_n(s) : \sup_{s \in \mathcal{S}} [Y(s) - b_n(s)]/a_n(s) > 0\}$, with generalizations of this approach given by Dombry and Ribatet (2015) and de Fondeville and Davison (2020). Likelihoods for the ensuing models are typically much easier to handle than for the corresponding max-stable process models, but the need for *censored* likelihoods as a bias reduction technique (e.g. Huser et al., 2016) still inhibits scalability. All realistic models for spatial extremes are based on suitably modified Gaussian processes, meaning that their censored likelihoods contain many evaluations of high-dimensional Gaussian cdfs, and standard implementations are limited to ≈ 30 locations. For a certain popular class of Pareto processes, de Fondeville and Davison (2018) used gradient score methods to avoid computing costly components of the likelihood, and combined this with coding efficiencies to permit inference on several hundred sites. Their application to a 3600-location problem is a notable exception to the small d phenomenon observed elsewhere.

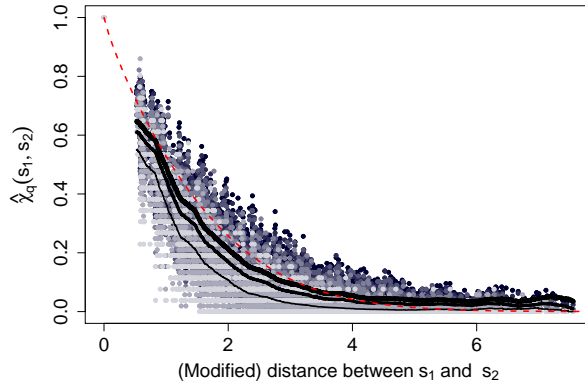


Figure 1: Estimates $\hat{\chi}_q(s_1, s_2)$ of $\chi_q(s_1, s_2)$ plotted against a modified version of distance $\|s_1 - s_2\|$ (see Section 6 for details of the spatial deformation). Darker points represent those estimated at lower quantiles, and lighter points at higher quantiles, moving from $q = 0.95$ (darkest) to $q = 0.995$ (lightest). The solid lines represent smoothed versions of $\hat{\chi}_{0.95}(s_1, s_2)$ (thickest), $\hat{\chi}_{0.97}(s_1, s_2)$ (medium) and $\hat{\chi}_{0.99}(s_1, s_2)$ (thinnest). The dashed red line is the estimate of $\chi_q(s_1, s_2)$ from the Brown–Resnick Pareto process, which does not vary with q .

In spite of such progress, when the goal of the analysis is extrapolation from observed levels further into the tail of the distribution, both max-stable and Pareto processes suffer from serious drawbacks in their applicability. Non-trivial limit processes arise as n tends to infinity only when the underlying process Y exhibits *asymptotic dependence*, meaning that for marginal distributions $Y(s_j) \sim F_{s_j}$, the extremal dependence measure $\chi(s_1, s_2) > 0$ for all $s_1 \neq s_2$, where $\chi(s_1, s_2) = \lim_{q \rightarrow 1} \chi_q(s_1, s_2)$ and

$$\chi_q(s_1, s_2) = \mathbb{P}\{F_{s_1}(Y(s_1)) > q, F_{s_2}(Y(s_2)) > q\} / (1 - q) = \mathbb{P}\{Y(s_1) > F_{s_1}^{-1}(q) \mid Y(s_2) > F_{s_2}^{-1}(q)\}. \quad (1)$$

When $\chi(s_1, s_2) = 0$ instead, as is the case for all Gaussian processes that are not perfectly dependent, the limiting max-stable process is a collection of independent random variables, and the Pareto process cannot be sensibly defined, since it degenerates everywhere except where the supremum is realized. Testing the assumption of asymptotic dependence is critical for application of max-stable or Pareto models, since if it does not hold then their incorrect use will lead to bias in estimation of extreme event probabilities.

A common feature of environmental data is that the dependence weakens as the level of the process increases: that is, for fixed s_1, s_2 , $\chi_q(s_1, s_2) \searrow$ as $q \rightarrow 1$. Figure 1 displays estimates of $\{\chi_q(s_1, s_2) : s_1 \neq s_2, q \in [0.95, 0.995]\}$, plotted against distance on the x -axis and in different shades for different quantiles, for the Australian summer temperature data that will be analyzed in Section 6. Dependence clearly weakens with distance as expected. The smoothed lines give averages of the estimates over distance at the $q = 0.95, 0.97$ and 0.99 quantiles. The decrease in these lines indicates that dependence also weakens at higher quantiles, but this would not be captured by a fitted Pareto process (also shown), for which $\chi_q(s_1, s_2) > 0$ does not depend on q .

In contrast to the asymptotic dependence case, little work has been done on developing asymptotically justifiable models for asymptotically independent extremes, for which $\chi_q(s_1, s_2) \searrow 0$. Partly, this is because one has to reconsider the meaning of “asymptotically justifiable” when the limits from classical extreme value theory are trivial. Based on a subasymptotic argument, [Wadsworth and Tawn \(2012\)](#) suggested a class of models that might be broadly applicable to asymptotically independent extremes, whilst the Gaussian process forms another possibility ([Bortot et al., 2000](#)).

Often it is unclear whether data display asymptotic dependence or asymptotic independence, but models often only cover one dependence type. [Huser et al. \(2017\)](#) and [Huser and Wadsworth \(2019\)](#) present spatial models that can capture both possibilities, although the same dependence class must hold over the entire spatial domain of interest. Furthermore, except for their Gaussian sub-models, independence between sites at long range is not possible. As such these models are suited only to smaller domains on which dependence persists. Very recent work ([Hazra et al., 2021](#)) extends the ideas of [Huser et al. \(2017\)](#) to address this issue.

1.2 Conditional extreme value theory: overview

We seek models for extremes of Y that are asymptotically justified, appropriate for the features of environmental data, and that can be fitted to reasonably large numbers of observation locations. Our approach to this task is to exploit the so-called conditional extreme value model and adapt to the spatial setting.

Conditional extreme value theory ([Heffernan and Tawn, 2004](#); [Heffernan and Resnick, 2007](#)) focuses on the behaviour of a random vector \mathbf{Y} , given that a fixed component of that random vector, say Y_j , is large. In contrast

to classical multivariate extreme value theory, which leads to multivariate max-stable and Pareto distributions, resulting limit distributions can offer non-trivial descriptions of vectors exhibiting asymptotic dependence or asymptotic independence. We clarify that by “conditional extreme value theory” or the “conditional model”, we mean theory or models derived explicitly through conditioning on a single component of a random vector, or value at a single site of a spatial process, being extreme. Other extreme value models (e.g. Pareto processes) are derived by conditioning on different extreme events, but the specific use of this terminology has become somewhat standard in the literature on extreme values (Das and Resnick, 2011b; Resnick and Zeber, 2014). In the time-series setting, there is related work that focuses on so-called tail chains, which describe the distribution of the process $\{Y_t : t \in \mathcal{T} \subset \mathbb{R}_+\}$ following an extreme event at $t = 0$ (Basrak and Segers, 2009; Resnick and Zeber, 2014; Papastathopoulos et al., 2017). In the spatial case, we lose the directionality of time, and this situation has been much less explored.

To take advantage of conditional extreme value theory in a spatial setting, we condition on the process Y being extreme at an arbitrary reference location s_0 . Based on an early formulation of the ideas in this paper, Tawn et al. (2018) illustrated that the parameters of the Heffernan and Tawn (2004) model that describe pairwise dependence vary smoothly relative to distance between sites, giving an indication of the potential for extending that multivariate model to spatial processes when conditioning on a single location s_0 . In practice, whilst this specific conditioning event might suit certain applications, there is often no natural reference location. More generally it is desirable to condition on the process being extreme anywhere over the domain of interest, $\mathcal{S} \subset \mathbb{R}^2$, as in the case of Pareto processes. We aim to combine the best of both worlds. Through the conditional extreme value approach we develop models that have flexible extremal dependence structures, accommodating asymptotic dependence and asymptotic independence. We then show how distributions conditioning on a single site can be combined to obtain inference on quantities of interest conditioning upon the maximum of the process at any collection of locations being large. This is achieved via an importance sampling algorithm, which also leads to an approximate method to simulate directly from the distribution of the field conditionally upon the maximum being large.

Aside from giving rise to flexible models for asymptotically (in)dependent data, the approach we propose can be scaled to accommodate several hundred observation locations. Whether asymptotically independent or dependent, most models for spatial extremes are driven by properties of the joint tail of Y , i.e., when all sites may be large together. Because of this, models are usually fitted via censored likelihood, so that locations with extreme observations contribute; see e.g. Thibaud and Opitz (2015) or Huser and Wadsworth (2019) for examples. This censoring is a computational burden due to the integrals involved, although Zhang et al. (2021) recently proposed the addition of a nugget effect to simplify likelihoods for simulation-based inference. The different asymptotic arguments for conditional extreme value theory do not necessitate censored likelihoods, making them inherently more scalable, although care can be required in ensuring that the implementation is sensible; further comments on this are given in Sections 3 and 6.

In Section 2.1 we detail the main assumption that will underlie the models built in Section 3, and which is illustrated through examples in Section 2.2.

1.3 Definitions and notation

Recall the definition of $\chi(s_1, s_2)$ given through equation (1). We specialize to the case of stationary processes Y , for which $\chi(s_1, s_2) \equiv \chi(h)$, $h = s_1 - s_2$. A process is termed asymptotically dependent if $\chi(h) > 0$ for all lags $h = (h_1, h_2)^\top$, and asymptotically independent if $\chi(h) = 0$ for all $\|h\| > 0$, with $\|\cdot\|$ the Euclidean distance. If $\chi(h) > 0$ for $\|h\| < \Delta_\theta$, and $\chi(h) = 0$ for $\|h\| \geq \Delta_\theta$, with $\theta = \arccos h_1/\|h\|$ the direction of h , we call the process *directionally lag-asymptotically dependent*, or simply *lag-asymptotically dependent* if the process is isotropic, so that $\Delta_\theta \equiv \Delta$.

Notationally, all vectors of length greater than one are expressed in boldface, with the exception of those denoting spatial location, e.g. $s_j = (s_{j,1}, s_{j,2})^\top$, including spatial lags $h = s_1 - s_2$. By convention, arithmetic operations on vectors are applied componentwise, with scalar values recycled as necessary. For example, if $f : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^d$, $c \in \mathbb{R}$, then $f(\mathbf{x} + c) = (f(x_1 + c), \dots, f(x_d + c))^\top$.

2 Main assumption and examples

2.1 Main assumption

The assumptions underlying conditional extreme value theory are most simply expressed following standardization to exponential-tailed margins, achieved in practice via the probability integral transformation. Let $\{X(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ be a stationary stochastic process with continuous sample paths and margins satisfying $P(X(s_j) > x) \sim ce^{-x}$, $x \rightarrow \infty$, for some $c > 0$. Where we wish to account for negative dependence, it is also assumed that $P(X(s_j) < x) \sim c^*e^{-|x|}$, as $x \rightarrow -\infty$, $c^* > 0$. For such cases, Keef et al. (2013) proposed

modelling with Laplace margins. We assume that there exist functions $\{a_{s-s_0} : \mathbb{R} \rightarrow \mathbb{R}, s \in \mathcal{S}\}$, with $a_0(x) = x$ and $\{b_{s-s_0} : \mathbb{R} \rightarrow (0, \infty), s \in \mathcal{S}\}$, such that for any $s_0 \in \mathcal{S}$, any $d \in \mathbb{N}$ and any collection of sites $s_1, \dots, s_d \in \mathcal{S}$,

$$\left(\left\{ \frac{X(s_i) - a_{s_i-s_0}(X(s_0))}{b_{s_i-s_0}(X(s_0))} \right\}_{i=1, \dots, d}, X(s_0) - t \right) | X(s_0) > t \xrightarrow{d} (\{Z^0(s_i)\}_{i=1, \dots, d}, E), \quad t \rightarrow \infty; \quad (2)$$

that is, convergence in distribution of the normalized process to $\{Z^0(s) : s \in \mathcal{S}\}$ in the sense of finite-dimensional distributions. In the limit, the variable $E \sim \text{Exp}(1)$ is independent of the residual process Z^0 , which satisfies $Z^0(s_0) = 0$ almost surely, but is non-degenerate for all $s \neq s_0$ and places no mass at $+\infty$. Note that in assumption (2) it is irrelevant whether the conditioning site s_0 represents one of the sites $\{s_1, \dots, s_d\}$. When assumption (2) is employed for modelling (see Sections 3 and 4) one indeed needs to condition on observed values. However, when simulating new events (see Section 5), the conditioning site could be any site in the domain.

An equivalent limiting formulation under the existence of a joint density for the process X is obtained by conditioning upon the precise value of $X(s_0)$. This can be seen via l'Hôpital's rule:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\mathbb{P} \left(\left\{ \frac{X(s_i) - a_{s_i-s_0}(X(s_0))}{b_{s_i-s_0}(X(s_0))} \right\}_{i=1, \dots, d} \leq \mathbf{z}, X(s_0) > t \right)}{\mathbb{P}(X(s_0) > t)} &= \lim_{t \rightarrow \infty} \frac{\frac{\partial}{\partial t} \mathbb{P} \left(\left\{ \frac{X(s_i) - a_{s_i-s_0}(X(s_0))}{b_{s_i-s_0}(X(s_0))} \right\}_{i=1, \dots, d} \leq \mathbf{z}, X(s_0) > t \right)}{\frac{\partial}{\partial t} \mathbb{P}(X(s_0) > t)} \\ &= \lim_{t \rightarrow \infty} \mathbb{P} \left(\left\{ \frac{X(s_i) - a_{s_i-s_0}(t)}{b_{s_i-s_0}(t)} \right\}_{i=1, \dots, d} \leq \mathbf{z} \mid X(s_0) = t \right). \end{aligned}$$

The independence of the conditioning variable is also assured under this alternative formulation of the assumption, see e.g. [Wadsworth et al. \(2017, Proposition 5\)](#). Since all the processes that we will consider have densities, this alternative version may sometimes be useful (see Section 2.2). Although the assumption of a density is very common for spatial statistical modelling, it may be of theoretical interest to consider the situation where this is not the case. We do not attempt to treat this here.

The functions a_{s-s_0}, b_{s-s_0} appearing in limit (2) can be characterized to some degree. [Heffernan and Resnick \(2007\)](#) detail various requirements in a bivariate setting under the assumption that the conditioning variable has a regularly varying tail, i.e., power law type decay, but with no marginal assumptions on the other variable. If the conditioning variable has an exponential tail, these requirements translate to the existence of functions $\psi_{s-s_0}^1, \psi_{s-s_0}^2$ such that for all $c \in \mathbb{R}$

$$\psi_{s-s_0}^1(c) = \lim_{t \rightarrow \infty} b_{s-s_0}(t+c)/b_{s-s_0}(t) \quad \psi_{s-s_0}^2(c) = \lim_{t \rightarrow \infty} \{a_{s-s_0}(t+c) - a_{s-s_0}(t)\}/b_{s-s_0}(t), \quad (3)$$

with local uniform convergence on compact subsets of \mathbb{R} ; see also [Papastathopoulos and Tawn \(2016\)](#). Conditions (3) do not provide detailed information since the non-conditioning variable can have any marginal distribution, but any function a_{s-s_0} or b_{s-s_0} in (2), and later in Section 3, should satisfy these. Some modest additional structure is given in Proposition 3 of Appendix A under assumptions on the support of the limit distribution. Generally the literature on conditional extremes is split according to whether the non-conditioning variable(s) are assumed to have a standard marginal form, such as the exponential tails in assumption (2). Standardization occurs in most applied literature (e.g. [Keef et al., 2013](#)) with the prescription for the assumed normalization following from a variety of theoretical examples, in conjunction with checks on the modelling assumptions. We follow this general approach with the identical exponential-tailed margins in assumption (2). In more probabilistic literature (e.g. [Das and Resnick, 2011a](#); [Drees and Janßen, 2017](#)) standardization is typically not assumed, or involves heavy tails rather than exponential tails, such as in [Heffernan and Resnick \(2007\)](#).

A version of assumption (2) suited only to processes exhibiting asymptotic dependence has been studied and applied in various places in the literature. In the spatial context, [Engelke et al. \(2014, 2015\)](#) exploit (2) with $a_{s-s_0}(x) \equiv x, b_{s-s_0}(x) \equiv 1$, i.e., the normalization does not depend on space. This is similar to the time series setting of [Basrak and Segers \(2009\)](#). The conditional extreme value approach has also been applied under asymptotic dependence in emerging applications such as graphical models and Markov trees ([Engelke and Hitz, 2020](#); [Segers, 2020](#)). Allowing a greater range of normalizations in (2) is the key feature that permits treatment of a substantially broader set of underlying processes.

The more established approaches to spatial extreme value theory based on max-stable and Pareto processes can be justified for extremes of asymptotically dependent data not only through convergence of finite-dimensional distributions, but also functional regular variation (e.g. [Hult and Lindskog, 2005](#); [de Fondeville and Davison, 2020](#); [de Haan and Ferreira, 2007](#), Chapter 9). We note that our assumption in convergence (2) is restricted to the case of finite-dimensional distributions. Functional convergence is not treated here, and although this would not alter the methodology that we introduce, we highlight that this is an area of potential further mathematical interest.

2.2 Examples

We present examples of widely used spatial processes that satisfy limit (2), and use these to identify useful structures for model building. Table 1 summarizes the examples of Section 2.2.1–2.2.5.

Table 1: Normalization functions and limit processes for theoretical examples given in Sections 2.2.1–2.2.5. For the process of [Huser and Wadsworth \(2019\)](#), V in (12) is taken as a marginally transformed Gaussian process. In the fourth row, $\ell_{s-s_0}(x)$ is a slowly-varying function of x , with $\lim_{x \rightarrow \infty} \ell_{s-s_0}(x) = 0$, whose form is given by (9).

Process	$a_{s-s_0}(x)$	$b_{s-s_0}(x)$	$Z^0(s)$
Gaussian	$\rho(s-s_0)^2 x$	$1 + a_{s-s_0}(x)^{1/2}$	Gaussian
Brown–Resnick	x	1	Gaussian
t_ν	x	1	$t_{\nu+1}$ (transformed margins)
Inverted Brown–Resnick	$\ell_{s-s_0}(x)x$	$a_{s-s_0}(x)/(\log x)^{1/2}$	Independence (reverse Gumbel margins)
Huser and Wadsworth (2019) ($\lambda < 1$)	$\rho(s-s_0)^2 x$	$1 + a_{s-s_0}(x)^{1/2}$	Gaussian

2.2.1 Gaussian process

Let $\{Y(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ be a standard stationary Gaussian processes with correlation function $\rho(h) \geq 0$ and let $X(s) = -\log(1 - \Phi(Y(s)))$ be the same process transformed to standard exponential margins. Then, taking $a_{s-s_0}(x) = \rho(s-s_0)^2 x$ and $b_{s-s_0}(x) = x^{1/2}$, the limit process Z^0 has finite-dimensional distributions that are Gaussian with zero mean and covariance structure defined by the matrix

$$\Sigma_0 = (2\rho_{k,0}\rho_{l,0}(\rho_{k,l} - \rho_{k,0}\rho_{l,0}))_{1 \leq k,l \leq d},$$

with $\rho_{k,0} = \rho(s_k - s_0)$ etc. See [Heffernan and Tawn \(2004\)](#) for more details of this derivation in a multivariate setting. This limit representation looks problematic as the process Z^0 becomes degenerate as $\rho(s_k - s_0) \rightarrow 0$, owing to the fact that the scale normalization, $X(s_0)^{1/2}$, is still present when $\rho(s_k - s_0) = 0$ even though $X(s_k)$ and $X(s_0)$ are then independent. To avoid this, following [Wadsworth and Tawn \(2013\)](#), we can instead consider

$$\left\{ \frac{X(s_i) - \rho(s_i - s_0)^2 X(s_0)}{\rho(s_i - s_0) X(s_0)^{1/2} + 1} \right\}_{i=1, \dots, d},$$

i.e. taking $b_{s-s_0}(x) = 1 + \rho(s-s_0)x^{1/2}$, for which the limit process Z^0 is Gaussian with zero mean and finite-dimensional covariance structure

$$\Sigma_0 = 2(\rho_{k,l} - \rho_{k,0}\rho_{l,0})_{1 \leq k,l \leq d}.$$

Consequently, Z^0 has the conditional distribution of a Gaussian process with correlation function ρ , conditional upon the event $Z^0(s_0) = 0$. However, when $\rho(s_k - s_0) = 0$, the limit process Z^0 does not have Gaussian margins, but identical margins to X . As such, there is still a discontinuity in the limit behaviour once independence is reached.

2.2.2 Brown–Resnick process

Let $\{X(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ be a Brown–Resnick process ([Kablichko et al., 2009](#)) with Gumbel margins. That is, X can be expressed as

$$X(s) = \bigvee_{i=1}^{\infty} E_i + W_i(s) - \sigma^2(s)/2 \quad (4)$$

where E_i are points of a Poisson process on \mathbb{R} with intensity $e^{-x} dx$, W_i are independent and identically distributed copies of a centred Gaussian process with stationary increments, and $\sigma^2(s) = \mathbb{E}\{W(s)^2\}$. The variogram of the process W is given by

$$\gamma(s_1, s_2) = \mathbb{E}\{[W(s_1) - W(s_2)]^2\}, \quad (5)$$

and if $W(0) = 0$ almost surely (a.s.), then $\sigma^2(s) = \gamma(s, 0)$. Note that the representation (4) is not unique, and e.g., [Dieker and Mikosch \(2015\)](#) provide alternative representations with the same distribution.

[Engelke et al. \(2015\)](#) show that, for such a process, the finite-dimensional distributions of extremal increments $\{X(s) - X(s_0) | X(s_0) > t\}$ converges as $t \rightarrow \infty$ to a multivariate Gaussian distribution with mean vector and covariance matrix

$$\mathbf{m}_0 = (-\gamma(s_i, s_0)/2)_{i=1, \dots, d}, \quad \Omega_0 = (\gamma(s_i, s_0)/2 + \gamma(s_k, s_0)/2 - \gamma(s_i, s_k)/2)_{1 \leq i, k \leq d}, \quad (6)$$

respectively. As such, the diagonal elements of the covariance matrix are $(\gamma(s_i, s_0))_{i=1, \dots, d}$, i.e., $\mathbf{m}_0 = -\text{diag}(\Omega_0)/2$. This is the same limiting formulation as (2), with $a_{s-s_0}(x) = x$, $b_{s-s_0}(x) = 1$, and Z^0 a Gaussian process whose moment structure is determined by expression (6).

2.2.3 t process

The t process, with $\nu > 0$ degrees of freedom, arises as a particular Gaussian scale mixture. Specifically, taking

$$Y(s) = RW(s), \quad (7)$$

with W a standard Gaussian process with correlation function $\rho(h)$, and $R^{-2} \sim \text{Gamma}(\nu/2, \nu/2)$, the finite-dimensional distributions $\mathbf{Y} = (Y(s_1), \dots, Y(s_d))$ have density

$$f_\nu^d(\mathbf{y}; \boldsymbol{\mu}, \Sigma) = C_\nu^d [1 + (\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) / \nu]^{-(\nu+d)/2},$$

with normalization constant $C_\nu^d = \Gamma((\nu+d)/2) / \Gamma(\nu/2) |\Sigma|^{1/2} (\nu\pi)^{d/2}$; we write $\mathbf{Y} \sim \text{St}_\nu^d(\boldsymbol{\mu}, \Sigma)$. Calculations for the bivariate t_ν distribution were given in Keef (2006); here we extend these to arbitrary dimension. Suppose $\mathbf{Y} \sim \text{St}_\nu^d(\mathbf{0}, \Sigma)$, with dispersion matrix $\Sigma = (\rho_{k,l})$ a correlation matrix, and let \mathbf{Y}_{-j} represent \mathbf{Y} without component j . Then

$$\frac{\mathbf{Y}_{-j} - \boldsymbol{\rho}_j Y_j}{(\nu + Y_j^2)^{1/2}} (\nu + 1)^{1/2} \sim \text{St}_{\nu+1}^{d-1}(\mathbf{0}, (\rho_{k,l} - \rho_{j,k}\rho_{j,l})_{k,l \neq j}), \quad (8)$$

with $\boldsymbol{\rho}_j \in (-1, 1)^{d-1}$ the j th column of Σ , without the j th component. Let $T: \mathbb{R} \rightarrow \mathbb{R}$ be a monotonic increasing transformation and consider $X = T(Y)$, such that $P(X(s_j) > x) \sim e^{-x}$, with $\mathbf{X} = T(\mathbf{Y})$ the transformed finite-dimensional random vector. A suitable choice of transformation T satisfies $T^{-1}(x) \sim K^{1/\nu} e^{x/\nu}$, $x \rightarrow \infty$, where $K = C_\nu^1 \nu^{(\nu-1)/2}$. We then have

$$\begin{aligned} P(\mathbf{X}_{-j} - t \leq \mathbf{z} | X_j = t) &= P(T(\mathbf{Y}_{-j}) \leq \mathbf{z} + t | T(Y_j) = t) \\ &= P\left(\frac{\mathbf{Y}_{-j} - \boldsymbol{\rho}_j T^{-1}(t)}{(\nu + T^{-1}(t)^2)^{1/2}} \leq \frac{T^{-1}(\mathbf{z} + t) - \boldsymbol{\rho}_j T^{-1}(t)}{(\nu + T^{-1}(t)^2)^{1/2}} \middle| Y_j = T^{-1}(t)\right), \end{aligned}$$

which, by (8) is the cdf of the $\text{St}_{\nu+1}^{d-1}(\mathbf{0}, (\rho_{k,l} - \rho_{j,k}\rho_{j,l})_{k,l \neq j} / (\nu + 1))$ distribution. Using the asymptotic form of $T^{-1}(x)$, the argument of this distribution function,

$$\frac{T^{-1}(\mathbf{z} + t) - \boldsymbol{\rho}_j T^{-1}(t)}{(\nu + T^{-1}(t)^2)^{1/2}} \rightarrow e^{\mathbf{z}/\nu} - \boldsymbol{\rho}_j, \quad t \rightarrow \infty.$$

Consequently, $\mathbf{X} - X_j | X_j > t \xrightarrow{d} \mathbf{Z}^j$, where $Z_j^j = 0$ and

$$P(\mathbf{Z}_{-j}^j \leq \mathbf{z}) = F_{\nu+1}^{d-1}(e^{\mathbf{z}/\nu}; \boldsymbol{\rho}_j, (\rho_{k,l} - \rho_{j,k}\rho_{j,l})_{k,l \neq j} / (\nu + 1)),$$

where $F_{\nu+1}^{d-1}(\cdot, \boldsymbol{\mu}, \Sigma)$ is the cdf of $\text{St}_{\nu+1}^{d-1}(\boldsymbol{\mu}, \Sigma)$. As such, we conclude for the spatial process that $a_{s-s_0}(x) = x$, $b_{s-s_0}(x) = 1$. Note that this distribution has mass on lines through $-\infty$, which arise due to the fact that in representation (7), large values of R cause both large and small values of Y , depending upon the sign of W .

The process $Z^0(s)$, defined through its finite-dimensional distributions $\mathbf{Z}^0 = \{Z^0(s_1), \dots, Z^0(s_d)\}$, is thus a transformed version of the $t_{\nu+1}$ process with $Z^0(s_0) = 0$ a.s., with positive probability of being equal to $-\infty$ at any other site. That probability is smaller the stronger the dependence with the conditioning site; i.e., locally around the conditioning site, there is a high probability of $Z^0 > -\infty$.

2.2.4 Inverted Brown–Resnick process

Wadsworth and Tawn (2012) introduced the class of inverted max-stable processes as those processes whose upper joint tail has the same dependence structure as the lower joint tail of a max-stable process. That is, if Y is a max-stable process and T represents a monotonically-decreasing marginal transformation, then $T(Y)$ is an inverted max-stable process. Applying the transformation $T(x) = e^{-x}$ to (4) yields the inverted Brown–Resnick process, with standard exponential margins.

Papastathopoulos and Tawn (2016) consider conditional limits of the bivariate margins of the inverted Brown–Resnick process. They show that the normalizations required to obtain a non-degenerate limit are

$$a_{s_i-s_0}(x) = x \exp \left\{ \gamma(s_i, s_0)/4 - (\gamma(s_i, s_0)/2)^{1/2} (2 \log x)^{1/2} + (\gamma(s_i, s_0)/2)^{1/2} \frac{\log \log x}{(2 \log x)^{1/2}} \right\} \quad (9)$$

$$b_{s_i-s_0}(x) = a_{s_i-s_0}(x) / (\log x)^{1/2}, \quad (10)$$

with limiting marginal distribution for $Z^0(s_i)$ given by

$$P(Z^0(s_i) \leq z) = 1 - \exp\{-\gamma(s_i, s_0)/2\}^{1/2} \exp\{2z/\gamma(s_i, s_0)^{1/2}\}/(8\pi)^{1/2}. \quad (11)$$

The multiplier of x in equation (9) is a slowly varying function $\ell_{s-s_0}(x)$, meaning that for $c > 0$, $\lim_{x \rightarrow \infty} \ell_{s-s_0}(cx)/\ell_{s-s_0}(x) = 1$. In the case of (9), $\lim_{x \rightarrow \infty} \ell_{s-s_0}(x) = 0$, i.e., $a_{s-s_0}(x) = o(x)$, $x \rightarrow \infty$. For the inverted Brown–Resnick process, the normalization and limiting distribution appear rather unintuitive when considering how γ affects the dependence. Indeed, as $\gamma(s_i, s_0) \rightarrow \infty$, the max-stable process, and thus the inverted max-stable process, approaches independence. Yet, for finite x , the normalization $a_{s_i-s_0}(x)$ becomes large, and the mass of the limiting distribution of $Z^0(s_i)$ is placed at smaller values. Note that replacing z by $\{\gamma(s_i - s_0)/2\}^{1/2}z - \gamma(s_i - s_0)^{1/2} \log\{\gamma(s_i - s_0)/2\}/4$ removes dependence of the limit distribution on γ ; this is equivalent to modifying the normalization functions to

$$\begin{aligned} \tilde{a}_{s_i-s_0}(x) &= a_{s_i-s_0}(x) - b_{s_i-s_0}(x)\gamma(s_i, s_0)^{1/2} \log\{\gamma(s_i, s_0)/2\}/4 \\ \tilde{b}_{s_i-s_0}(x) &= \{\gamma(s_i - s_0)/2\}^{1/2}b_{s_i-s_0}(x). \end{aligned}$$

However, whilst stabilizing the limit in γ , this still does not lead to an easily interpretable normalization in the sense of a and/or b decreasing monotonically with γ . To understand this, it is helpful to consider how such a process is formed. From equation (4), with $\sigma^2(s) = \gamma(s)$, the Brown–Resnick process is the pointwise maximum of location-adjusted Gaussian processes with negative drift. The more negative the drift (i.e., the larger γ) the more likely it is that the pointwise maxima from two locations will stem from different underlying Gaussian processes, which is why independence is achieved in the limit as $\gamma \rightarrow \infty$. Now, large values of the *inverted* Brown–Resnick process correspond to small values of the original process, which are likely to be at the intersection points whereby different Gaussian processes contribute to the suprema. This rather complex construction thus leads to the seemingly unintuitive behaviour. Concerning the limiting process Z^0 , Papastathopoulos and Tawn (2016) state that this corresponds to pointwise independence, and as such all structure lies in the functions a_{s-s_0} , b_{s-s_0} , and reverse Gumbel type margins (11).

2.2.5 Process of Huser and Wadsworth (2019)

Huser and Wadsworth (2019) present a model for spatial extremes obtained by scale mixtures on Pareto margins, or location mixtures on exponential margins. Suppose that V is an asymptotically independent spatial process with unit exponential margins, and Q is an independent unit exponential variate. Then

$$X^*(s) = \delta Q + (1 - \delta)V(s) \quad (12)$$

exhibits asymptotic independence for $\delta \leq 1/2$ and asymptotic dependence for $\delta > 1/2$. Here, for simplicity of presentation, we reparameterize to $X(s) = X^*(s)/(1 - \delta)$, and set $\lambda = \delta/(1 - \delta) \in (0, \infty)$, with $\lambda \in (0, 1]$ corresponding to asymptotic independence. The marginal distribution of X is

$$P(X(s_j) > x) = \frac{1}{1 - \lambda}e^{-x} - \frac{\lambda}{1 - \lambda}e^{-x/\lambda},$$

so that the leading order term is $e^{-x}/(1 - \lambda)$ for $\lambda < 1$ and $\lambda e^{-x/\lambda}/(\lambda - 1)$ for $\lambda > 1$; the case $\lambda = 1$ is obtained as $(1 + x)e^{-x}$ upon taking the limit.

The following proposition establishes the behaviour of interest for the case $\lambda \in (0, 1)$; in particular we find that the conditional limit distribution of the modified process X requires the same normalization and has the same limit distribution as the process V , if the scale normalization required for V is increasing in $V(s_0)$. For notational simplicity in the following we set $V_0 = V(s_0)$, $X_0 = X(s_0)$, $\mathbf{V} = (V(s_1), \dots, V(s_d))$, $\mathbf{X} = (X(s_1), \dots, X(s_d))$, $\mathbf{a}(x) = (a_{s_1-s_0}(x), \dots, a_{s_d-s_0}(x))$, and $\mathbf{b}(x) = (b_{s_1-s_0}(x), \dots, b_{s_d-s_0}(x))$.

Proposition 1. *Suppose that \mathbf{V} has unit exponential margins, $P(V_l > v) = e^{-v}$, $v > 0$, and for $\mathbf{a}(v)$ and $\mathbf{b}(v)$ with twice-differentiable components a_l, b_l satisfying $a'_l(v) \sim \alpha_l$, $a''_l(v) = o(1)$, $b'_l(v)/b_l(v) = o(1)$, as $v \rightarrow \infty$, $l = 1, \dots, d$,*

$$P\left(\frac{\mathbf{V} - \mathbf{a}(V_0)}{\mathbf{b}(V_0)} \leq \mathbf{z} \mid V_0 = v\right) \rightarrow G(\mathbf{z}), \quad v \rightarrow \infty.$$

Suppose further that all first and second order partial derivatives with respect to components of \mathbf{z} converge, as specified in Lemma 1 of Appendix A. Then for $\mathbf{X} = \mathbf{V} + \lambda Q$, with $Q \sim \text{Exp}(1)$ independent of \mathbf{V} , and $\lambda \in (0, 1)$,

$$P\left(\frac{\mathbf{X} - \mathbf{a}(X_0)}{\mathbf{b}(X_0)} \leq \mathbf{z} \mid X_0 = x\right) \rightarrow \begin{cases} G(\mathbf{z}), & \min_{1 \leq l \leq d} b_l(x) \rightarrow \infty \\ \int_0^{q^*} G\left(\mathbf{z} + \frac{(\alpha-1)\lambda q}{\lim_{x \rightarrow \infty} \mathbf{b}(x)}\right) (1 - \lambda)e^{-(1-\lambda)q} dq, & \text{otherwise} \end{cases}$$

as $x \rightarrow \infty$, with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^\top$, and $q^ = \lim_{x \rightarrow \infty} ([\min\{\min_l(a_l(x) + b_l(x)z_l), x\}]/\lambda)_+$.*

For $\lambda > 1$, corresponding to $\delta > 1/2$ in (12), asymptotic dependence arises, and with a rescaling, one can instead express $X(s) = Q + V(s)/\lambda$ such that $P(X(s_j) > x) \sim (1 - 1/\lambda)^{-1} e^{-x}$. In this case, $a_{s-s_0}(x) = x$, $b_{s-s_0}(x) = 1$ and the limiting dependence structure is determined by the distribution of V .

Corollary 1. *Assume the conditions of Proposition 1. The normalization and limit distribution for X are the same as those for V when either:*

(i) All $b_{s-s_0}(x) \rightarrow \infty$, $x \rightarrow \infty$, for all s, s_0 .

(ii) $a_{s-s_0}(x) \sim x$ and $b_{s-s_0}(x) \sim 1$, $x \rightarrow \infty$, for all s, s_0 , as arises under asymptotic dependence for V .

The proof of Proposition 1 is in Appendix A. For the application of Proposition 1, convergence of the partial derivatives needs to be established. Supposing concretely that V is a Gaussian process with margins transformed to be exponential, Lemma 2 and Remark 1 of Appendix A provides this result.

The fact that the normalization and limit distribution are often the same for X and V may seem surprising given that Huser and Wadsworth (2019) noted differences in the so-called coefficient of tail dependence (Ledford and Tawn, 1996) for X according to the relative values of δ in equation (12) and the coefficient of tail dependence for V . This result highlights that the conditional extreme value approach provides a different asymptotic representation of the tail of a random vector or process to those where all variables are considered to be extreme simultaneously. Further elaboration of this can be found in Nolde and Wadsworth (2021).

3 Statistical modelling

Motivated by the limit assumption (2) and the examples in Section 2.2, we suppose that for a high threshold u ,

$$\{X(s)|X(s_0) > u : s \in \mathcal{S}\} \stackrel{d}{\approx} \{a_{s-s_0}(X(s_0)) + b_{s-s_0}(X(s_0))Z^0(s) : s \in \mathcal{S}\}, \quad (13)$$

for some choice of functions a_{s-s_0}, b_{s-s_0} , residual process Z^0 , and that $X(s_0) - u|X(s_0) > u \sim \text{Exp}(1)$ is independent of Z^0 . This specifies a process model conditioning on the value at a particular site s_0 being extreme. The process is observed at d locations, so for inference we are interested in d such specifications, taking each observation site as the conditioning site s_0 . However, assuming stationarity in space, we obtain a common form for a_{s-s_0}, b_{s-s_0} and Z^0 whatever the conditioning site. Inference is described in Section 4, whilst simulation conditioning both on $\{X(s_0) > u\}$ and $\{\max_{i_1 \leq i \leq i_m} X(s_i) > u\}$, for any collection of sites $\{s_{i_1}, \dots, s_{i_m}\}$ with each $s_{i_j} \in \mathcal{S}$, is dealt with in Section 5. For now, we address the specification of a_{s-s_0}, b_{s-s_0} and Z^0 . Each of the proposed sets of functions a_{s-s_0}, b_{s-s_0} satisfies conditions (3).

3.1 Functions a_{s-s_0} and b_{s-s_0}

As noted in Section 1, the function $a_{s-s_0} : \mathbb{R} \rightarrow \mathbb{R}$ must satisfy $a_0(x) = x$. Further, if the process X is asymptotically dependent, then $a_{s-s_0}(x) = x$ for all s , whilst if it is (directionally) lag-asymptotically dependent up to lag Δ_θ , then $a_{s-s_0}(x) = x$ for $\|s - s_0\| \leq \Delta_\theta$. Under asymptotic independence, and if Z^0 has any positive support, we have $a_{s-s_0}(x) < x$, where the bound may only hold asymptotically, i.e., as $x \rightarrow \infty$ (Proposition 3, Appendix A). Focusing on the isotropic case, we propose the general parametric form for a_{s-s_0} as

$$a_{s-s_0}(x) = x\alpha(s - s_0) = \begin{cases} x & \|s - s_0\| < \Delta \\ x \exp\{-[(\|s - s_0\| - \Delta)/\lambda]^\kappa\}, & \|s - s_0\| \geq \Delta. \end{cases} \quad (14)$$

Taking $\Delta = 0$ provides a flexible model for asymptotic independence, with parameters (λ, κ) to estimate. Taking $\Delta > 0$ but less than the maximum distance between any two sites in the domain would allow for lag-asymptotic dependence; the value of Δ itself could be estimated directly, although profiling over Δ on a grid may be preferable. Taking Δ larger than the maximum distance between any two sites would correspond to asymptotic dependence, in which case (λ, κ) would not be estimated. Equation (14) covers all forms for a_{s-s_0} from lines 1,2,3 and 5 of Table 1, if ρ is a powered exponential correlation function, and line 4 up to a slowly varying function. Furthermore, condition (3) is satisfied as long as $b_{s-s_0}(x) \not\rightarrow 0$ as $x \rightarrow \infty$. We note that other monotonically decreasing functions are also candidates for the second line of (14); certain correlation functions and survival functions are natural choices.

For the function b_{s-s_0} , we propose three forms to achieve different modelling aims.

Model 1 $b_{s-s_0}(x) = [1 + \zeta x^\beta]^{-1}$, $\beta < 0$, $\zeta \geq 0$.

The rationale behind this suggestion is that, if $\zeta > 0$, then $b_{s-s_0}(x) \nearrow 1$, with ζ and β controlling the convergence to the constant value. When $\|s - s_0\| < \Delta$ and data are (lag-)asymptotically dependent, this

permits the model to display some *subasymptotic* dependence, in the sense that for $\|s_k - s_0\| < \Delta$, and x_q the q -quantile of X ,

$$\begin{aligned}\chi_q(s_k, s_0) &= \mathbb{P}(X(s_k) > x_q | X(s_0) > x_q) = \mathbb{P}(X(s_0) + [1 + \zeta X(s_0)^\beta]^{-1} Z^0(s_k) > x_q | X(s_0) > x_q) \\ &= e^{x_q} \int_{x_q}^{\infty} \mathbb{P}(Z^0(s_k) > (x_q - v) / [1 + \zeta v^\beta]^{-1}) e^{-v} dv \\ &= \int_0^{\infty} \mathbb{P}(Z^0(s_k) > -r / [1 + \zeta(r + x_q)^\beta]^{-1}) e^{-r} dr \\ &\searrow \int_0^{\infty} \mathbb{P}(Z^0(s_k) > -r) e^{-r} dr, \quad q \rightarrow 1, x_q \rightarrow \infty.\end{aligned}$$

Here, the dependence measure $\chi_q(s_k, s_0)$ is as defined in equation (1). In practice, spatial data nearly always exhibit values of $\chi_q(s_k, s_0)$ that decrease with the level q , although Pareto process models for asymptotic dependence cannot capture this feature, leading to overestimation of dependence at extreme levels. Taking b_{s-s_0} as a constant function here with $a_{s-s_0}(x) = x$ would also yield asymptotic dependence with χ_q not varying with q . Such a model could be implemented by choice if the estimate of $\beta \ll 0$, or $\zeta \approx 0$, for example.

Model 2 Let $\Delta = 0$ in (14), and $b_{s-s_0}(x) = x^\beta$.

This represents a modelling strategy close to that proposed by [Heffernan and Tawn \(2004\)](#). If the marginal support of Z^0 includes $(0, \infty)$, we require $\beta < 1$ (Proposition 3, Appendix A), whilst conditions (3) imply $\beta \geq 0$. We have

$$\chi_q(s_k, s_0) = \int_0^{\infty} \mathbb{P}[Z^0(s_k) > \{x_q - \alpha(s_k - s_0)(r + x_q)\} / (r + x_q)^\beta] e^{-r} dr \searrow 0, \quad q \rightarrow 1, x_q \rightarrow \infty,$$

so that the rate of convergence to zero is controlled by β , in conjunction with the value of $\alpha(s_k - s_0)$ and the distribution of $Z^0(s_k)$. Such a model may perform well on data that exhibit asymptotic independence, but positive dependence over the whole region of study. In particular, since $b_{s-s_0}(x)$ is still growing even as $a_{s-s_0}(x) \rightarrow 0$ ($\|s - s_0\| \rightarrow \infty$), independence cannot be achieved as $\|s - s_0\| \rightarrow \infty$. This observation motivates our final model.

Model 3 Let $\Delta = 0$ in (14), and $b_{s-s_0}(x) = 1 + a_{s-s_0}(x)^\beta$, $\beta > 0$.

With Model 3, if $a_{s-s_0}(x) \rightarrow 0$ for fixed x as $\|s - s_0\| \rightarrow \infty$, then $b_{s-s_0}(x) \rightarrow 1$ as $\|s - s_0\| \rightarrow \infty$. Thus, for s sufficiently far from s_0 , we would have

$$X(s) | X(s_0) > u \stackrel{d}{\approx} Z^0(s).$$

This final observation indicates that, under Model 3, if s is far enough from s_0 that $X(s)$ and $X(s_0)$ are independent, the margins of the process Z^0 should be the same as the process X .

Models 1–3 provide a wide variety of structures, but we note that other choices could be considered.

3.2 Process $Z^0(s)$

The random field Z^0 must satisfy $Z^0(s_0) = 0$. Supposing initially that we begin with an arbitrary Gaussian process Z_G , there are two natural ways to derive a process from Z_G with this property:

- (i) Set Z^0 to have the distribution of $Z_G | Z_G(s_0) = 0$
- (ii) Set Z^0 to have the distribution of $Z_G(s) - Z_G(s_0)$,

with the resulting processes both again Gaussian. The initial process Z_G might be stationary, and specified by a correlation function ρ and variance σ^2 , or have stationary increments, specified by a variogram γ , as in (5); addition of the drift term $-\gamma(s, s_0)/2$ to the latter yields the process specified by (6).

Taking such a Gaussian process may be a natural and parsimonious choice in many situations. Indeed, allowing non-Gaussian dependence in Z^0 would lead to intractable models for high-dimensions, although we emphasize that the dependence of the modelled process X is not Gaussian. However, the choice of marginal distribution for Z^0 can impact upon the model; in particular, as noted at the end of Section 3.1, we may wish Z^0 to have the same margins as X in certain places, e.g., when $\|s - s_0\|$ is large.

We thus propose the following distributional family, that includes both Gaussian and Laplace marginal distributions. We say that a random variable Z has a *delta-Laplace* distribution, with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, if its density is

$$f(z) = \frac{\delta}{2\sigma\Gamma(1/\delta)} \exp \left\{ - \left| \frac{z - \mu}{\sigma} \right|^\delta \right\}, \quad \delta > 0, \quad (15)$$

with $\Gamma(\cdot)$ the standard gamma function. When $\delta = 1$ this is the Laplace distribution. If X has Laplace margins, as suggested in Keef et al. (2013), Model 3 can be completed by allowing the parameter δ to depend on $s - s_0$, such that $\delta(s - s_0) \rightarrow 1$ as $\|s - s_0\| \rightarrow \infty$. When $\delta = 2$, density (15) is that of the Gaussian distribution with mean μ and variance $\sigma^2/2$. In general, if Z has distribution (15), then $E(Z) = \mu$ and $\text{Var}(Z) = \{\Gamma(3/\delta)/\Gamma(1/\delta)\}\sigma^2$.

In practice, the location and scale structure implied by the processes (i) and (ii) defined on Gaussian margins are passed through to the delta-Laplace margins via the probability integral transform. In some situations it may be desirable to incorporate alternative mean structures by specifying a particular parametric form for $\mu(s - s_0)$ that is different from those implied by (i) and (ii); further discussion on this is made in Section 6.

4 Inference

4.1 Likelihood

The models described in Section 3 are suitable given an extreme in a single location. However, we would like to combine these models to allow for extremes in any observed location. Assuming stationarity, this motivates the use of a composite likelihood to estimate parameters of a_{s-s_0} , b_{s-s_0} and Z^0 , all of which do not depend on the particular conditioning site s_0 , which is now taken to be one of the observation locations. Specifically, denote these parameters by $\boldsymbol{\theta} = (\boldsymbol{\theta}_a, \boldsymbol{\theta}_b, \boldsymbol{\theta}_Z)$, where some subsets of parameters may be scalar. The likelihood based on the n_j points $(X_i(s_1), \dots, X_i(s_d)) = (x_1^i, \dots, x_d^i)$, $i = 1, \dots, n_j$, for which $X_i(s_j) > u$ is denoted

$$L_j(\boldsymbol{\theta}) = \prod_{i=1}^{n_j} f_{Z^j}(\{[x_k^i - a_{s_k-s_j}(x_j^i; \boldsymbol{\theta}_a)]/b_{s_k-s_j}(x_j^i; \boldsymbol{\theta}_b)\}_{k \in \{1, \dots, d\} \setminus j}; \boldsymbol{\theta}_Z) \prod_{k \in \{1, \dots, d\} \setminus j} b_{s_k-s_j}(x_j^i; \boldsymbol{\theta}_b)^{-1}, \quad (16)$$

with f_{Z^j} the density of Z^0 at the observation locations excluding s_j when $s_0 = s_j$. We estimate $\boldsymbol{\theta}$ from the composite likelihood over all d observed sites $L(\boldsymbol{\theta}) = \prod_{j=1}^d L_j(\boldsymbol{\theta})$. Composite likelihoods have been widely used in inference for max-stable processes (Padoan et al., 2010), where lower-dimensional — typically pairwise — densities are used due to being substantially simpler to compute. Our usage here is different, as the components of the composite likelihood are full joint densities. However, any realizations of X that are larger than u at multiple sites will be counted more than once in the likelihood, with different conditioning sites. The motivation for the composite likelihood is thus to get a single set of parameters that, on average, represent well each of the distributions conditioning on a single site being large. Parameter uncertainty can be assessed by bootstrap techniques; see Section 6. The first product in equation (16) arises from an assumption of independent replications of $X|X(s_j) > u$. This will often not be satisfied in practice, but for weak-moderate temporal dependence parameter consistency is unaffected by such misspecification (e.g. Chandler and Bate, 2007), while we can take advantage of a block bootstrap for assessing parameter uncertainty.

An advantage of the conditional approach over other models is that the likelihood does not require integrals as in the censored likelihoods commonly used for asymptotically dependent or asymptotically independent models (e.g., Wadsworth and Tawn, 2014; Thibaud and Opitz, 2015; Huser et al., 2017; Huser and Wadsworth, 2019). Such integrals are typically the limiting factor in the number of sites to which one can fit the model, often meaning a reasonable upper limit is 20–30 sites. Figure 9 in Appendix C displays example computation times for the evaluation of a single likelihood L_j , and the full composite likelihood for different numbers of observation locations d and different numbers of average repetitions n_j , both for Model 3. Based on these findings, it is perfectly feasible to optimize a likelihood for a moderate number of repetitions at 100–200 observation locations with no special computing power. In a recent review paper, Huser and Wadsworth (2022) suggest cutting the computational burden of the composite likelihood $L(\boldsymbol{\theta})$ by taking a product over a subset of the full set of possible conditioning sites.

4.2 Model selection and diagnostics

If all options for a_{s-s_0} , b_{s-s_0} and Z^0 from Sections 3.1–3.2 are considered, the number of candidate models becomes large. One way to inform the structure of the model is to perform pairwise fits, for which Z^0 is a scalar random variable, and plot parameter estimates against distance. This will help to uncover if and how $\alpha(s - s_0)$ should decrease with $\|s - s_0\|$, as well as evolution of parameters of Z^0 with $\|s - s_0\|$.

Although the computation time for the full composite likelihood is reasonable, it is desirable to have a faster way to compare several candidate models. To achieve this one could select a single site j and compare goodness of fit based on AIC using likelihood (16). Should a clear preference for a set of modelling choices emerge over investigation of a few such conditioning sites, these can be carried forward to a full composite likelihood fit.

To assess overall goodness of fit, we suggest (i) assessing whether the distribution of fitted residual processes $\{\widehat{Z}^0(s) = [X(s) - \widehat{a}_{s-s_0}(X(s_0))]/\widehat{b}_{s-s_0}(X(s_0))\}$ is close to the assumed distribution, (ii) comparing parameter estimates from pairwise fits to global fits, and (iii) verifying that simulations from the model are consistent with

the observed distribution when one or more sites is extreme. The assumption of independence of the conditioning variable and the process Z^0 can be tested by summarizing the fitted \widehat{Z}^0 processes with low-dimensional variables such as the empirical mean and variance, and testing whether they are independent of the conditioning variable. If dependence is implicated then a higher threshold could be considered.

As we will show in Section 5, simulation from this model is simple. As a result, diagnostics based on the intended model purpose can be explored, by comparing the behaviour of the data to the fitted model by simulation.

5 Simulation

Inference on quantities of interest, such as the probabilities of different extreme events, can be made via simulation. Given the nature of the model and assumptions, simulation conditional on being extreme at a particular location s_0 is straightforward, and the algorithm is outlined in Section 5.1. In many circumstances, it is desirable to condition instead on the process being extreme at some part of the domain, but not at a specific location. We address methods for this in Section 5.2.

5.1 Simulation given an extreme at a specified site

Under assumption (2) and model (13), simulation of X , given an extreme value above a threshold $v \geq u$ at a specific location s_0 , proceeds as follows:

Algorithm 1.

1. Generate $X(s_0)|X(s_0) > v \sim \text{Exp}(1)$
2. Independently of $X(s_0)$, generate $\{Z^0(s) : s \in \mathcal{S}\}$ from the model as specified in Section 3.2.
3. Set $\{X(s)|X(s_0) > v : s \in \mathcal{S}\} = \{a_{s-s_0}(X(s_0)) + b_{s-s_0}(X(s_0))Z^0(s) : s \in \mathcal{S}\}$.

In practice, Algorithm 1 is used to simulate at a finite m -dimensional collection of sites $\{s_{i_1}, \dots, s_{i_m}\}$, where the observation sites $\{s_1, \dots, s_d\}$ may or may not be included. In particular, the conditioning site s_0 need not correspond to an observation location. We denote the m -dimensional distribution of $X|X(s_0) > v$, obtained via the approximation (13), by Q_0^v . Further, denote by $M = \{i_1, \dots, i_m\}$ the set of indices corresponding to the simulation sites, and by $D = \{1, \dots, d\}$ the indices of the observation locations, where we may have $M = D$. Letting $\mathbf{X}_M = \{X(s_{i_1}), \dots, X(s_{i_m})\}$, Algorithm 1 can be used to estimate quantities of the form

$$\mathbb{E}\{g(\mathbf{X}_M)|X(s_0) > v\} = \mathbb{E}_{Q_0^v}\{g(\mathbf{X}_M)\}, \quad (17)$$

for some $g : \mathbb{R}^m \rightarrow \mathbb{R}^l$, $l \geq 1$. Equation (17) is convenient if one is either interested in conditioning on a specific location, or if $g(\cdot)$ is an indicator function for an event that involves $X(s_0) > v' \geq v$, since then the unconditional probability of this event is easily estimated using the exponential distribution of $X(s_0)$. As an illustration, suppose that the simulation sites are the observation sites, let $\mathbb{1}(A)$ be the indicator function for the occurrence of event A , and $g(\mathbf{X}_M) = \mathbb{1}(X(s_1) > v, \dots, X(s_d) > v)$. Then, taking s_0 as any s_j , $j = 1, \dots, d$ we have

$$\mathbb{P}(X(s_1) > v, \dots, X(s_d) > v) = \mathbb{E}_{Q_0^v}\{\mathbb{1}(X(s_1) > v, \dots, X(s_d) > v)\} \times \mathbb{P}(X(s_0) > v).$$

5.2 Simulation given an extreme in the domain

In place of conditioning on a single site being extreme, interest often lies in estimating quantities of the form

$$\mathbb{E}\left\{g(\mathbf{X}_M)|\max_{i \in M} X(s_i) > v\right\}, \quad (18)$$

where if $m = |M|$ is large and the locations suitably arranged, $\{\max_{i \in M} X(s_i) > v\} \approx \{\sup_{\mathcal{S}} X(s) > v\}$.

We denote the distribution of $\mathbf{X}_M|\max_{i \in M} X(s_i) > v$ by \mathbb{P}^v , abbreviate $X(s_i) = X_i$, $i \in M$ or $i \in D$, and let $Q^v := \sum_{i \in M} \pi_i Q_i^v$, where Q_i^v represents Q_0^v at $s_0 = s_i$ and

$$\pi_i = \frac{\mathbb{P}(X_i > v)}{\sum_{i \in M} \mathbb{P}(X_i > v)},$$

which equals $1/m$ if the margins are identical. In other words Q^v is the mixture distribution formed by selecting each Q_i^v with probability π_i . The support of Q^v is $\{\mathbf{x} \in \mathbb{R}^m : \max_{i \in M} x_i > v\}$. We note that because

we standardize the margins of X , all margins are equal and some of the following algorithms can be simplified. However, we give the more general formulation below.

We firstly describe an existing approach to the problem of estimating (18) based on rejection sampling, together with a number of limitations of this method. To overcome these, we introduce a new approach in Section 5.2.2 based on importance sampling.

5.2.1 Rejection sampling

Keef et al. (2013) propose a method to simulate from the distribution of $\mathbf{X}_D | \max_{i \in D} X(s_i) > u$, i.e., where $v = u$ and $M = D = \{1, \dots, d\}$, the set of observation locations. Note the partition

$$\left\{ \mathbf{x} \in \mathbb{R}^d : \max_{i \in D} x_i > u \right\} = \cup_{j=1}^d \left\{ \mathbf{x} \in \mathbb{R}^d : x_j > u, x_j = \max_{i \in D} x_i \right\},$$

and that $P^u = \sum_{j \in D} \tilde{\pi}_j^u Q_{j, \max}^u$, where $Q_{j, \max}^u$ is the distribution of $\mathbf{X} | \{X_j > u, X_j = \max_{i \in D} X_i\}$, and $\tilde{\pi}_j^u = P(X_j = \max_{i \in D} X_i | \max_{i \in D} X_i > u)$. Their algorithm is as follows:

Algorithm 2.

1. Estimate $\tilde{\pi}_j^u$ using empirical proportions
2. To simulate from $Q_{j, \max}^u$, draw from Q_j^u and reject unless $X_j = \max_{i \in D} X_i$
3. Simulate from P^u by drawing from $Q_{j, \max}^u$ with probability $\tilde{\pi}_j^u$

This rejection method leads directly to draws from (an estimate of) P^u , but has substantial drawbacks:

- (i) One can only condition on being large at the observation locations, rather than any set of locations;
- (ii) For high dimensions, the number of rejections to get a single draw from $Q_{j, \max}^u$ may be very high;
- (iii) To simulate from P^v , $v \gg u$, one needs to firstly simulate from P^u and use the resulting draws to estimate the probabilities $\tilde{\pi}_j^u$, before proceeding with the rejection sampling approach to simulate $Q_{j, \max}^v$;
- (iv) If the maximum never occurs at a particular site in the sample, it will not occur there in simulations either.

The final point can be addressed by simulating from the model to estimate $\tilde{\pi}_j^u$, at the cost of adding complexity to the algorithm. We propose an approach to estimate quantities of the form (18) via importance sampling, that avoids all of these drawbacks.

5.2.2 Importance sampling

Compared to P^v , the distribution Q^v samples more frequently in regions where multiple variables are extreme. However, this frequency is tractable: in the region where exactly k variables are larger than v , Q^v samples k times more observations than P^v , because there are k distributions Q_j^v covering this area.

To formalize this connection, let P be a probability measure on $\mathcal{B}(\mathbb{R}^m)$, the Borel sigma algebra of \mathbb{R}^m , for which $P(X_j > v) > 0$ for all $j \in M$. Letting $R_{\max}^v = \{\mathbf{x} \in \mathbb{R}^m : \max_{i \in M} x_i > v\}$, and $E, E_{P^v}, E_{Q_j^v}$ denote expectation with respect to P, P^v and Q_j^v respectively, we can express

$$E_{P^v}\{g(\mathbf{X}_M)\} = \frac{E\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\}}{E\{\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\}}, \quad E_{Q_j^v}\{g(\mathbf{X}_M)\} = \frac{E\{g(\mathbf{X}_M)\mathbb{1}(X_j > v)\}}{E\{\mathbb{1}(X_j > v)\}},$$

and $E_{Q^v}\{g(\mathbf{X})\} = \sum_{i \in M} \pi_i E_{Q_i^v}\{g(\mathbf{X})\}$. Further note that $R_{\max}^v = \cup_K R_K^v$, where

$$R_K^v = \{\mathbf{x} \in \mathbb{R}^m : x_k > v, \text{ for all } k \in K, \text{ and } x_l \leq v \text{ for all } l \in M \setminus K\}, \quad K \subseteq M,$$

and the union is over all K in the power set of M , minus the empty set, with all R_K^v disjoint.

Proposition 2. For any random vector $\mathbf{X}_M \in \mathbb{R}^m$, for which $P(X_i > v) > 0$ for all $i \in M$, and any function $g : \mathbb{R}^m \rightarrow \mathbb{R}^l$

$$E_{P^v}\{g(\mathbf{X}_M)\} = \frac{\sum_{i \in M} P(X_i > v)}{P(\max_{i \in M} X_i > v)} \sum_K E_{Q^v}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)/|K|\} = \frac{\sum_K E_{Q^v}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)/|K|\}}{\sum_K E_{Q^v}\{\mathbb{1}(\mathbf{X}_M \in R_K^v)/|K|\}}. \quad (19)$$

Proof. We have

$$\mathbb{E}_{P^v}\{g(\mathbf{X}_M)\} = \frac{\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\}}{\mathbb{E}\{\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\}} = \frac{\sum_K \mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\}}{\mathbb{P}(\max_{i \in M} X_i > v)}, \quad (20)$$

whilst for any $j \in K$ we also have

$$\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\} = \frac{\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\mathbb{1}(X_j > v)\}}{\mathbb{E}\{\mathbb{1}(X_j > v)\}} \mathbb{E}\{\mathbb{1}(X_j > v)\}.$$

Averaging over each $j \in K$ we get

$$\begin{aligned} \mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\} &= \frac{1}{|K|} \sum_{j \in K} \frac{\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\mathbb{1}(X_j > v)\}}{\mathbb{E}\{\mathbb{1}(X_j > v)\}} \mathbb{E}\{\mathbb{1}(X_j > v)\} \\ &= \frac{1}{|K|} \sum_{j \in M} \frac{\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\mathbb{1}(X_j > v)\}}{\mathbb{E}\{\mathbb{1}(X_j > v)\}} \pi_j \sum_{i \in M} \mathbb{P}(X_i > v), \end{aligned} \quad (21)$$

the second line following since for $j \in M \setminus K$, $\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\mathbb{1}(X_j > v)\} = 0$, and using the definition of π_j . Putting (20) and (21) together

$$\begin{aligned} \mathbb{E}_{P^v}\{g(\mathbf{X}_M)\} &= \frac{\sum_{i \in M} \mathbb{P}(X_i > v)}{\mathbb{P}(\max_{i \in M} X_i > v)} \sum_K \frac{1}{|K|} \sum_{j \in M} \pi_j \frac{\mathbb{E}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)\mathbb{1}(X_j > v)\}}{\mathbb{E}\{\mathbb{1}(X_j > v)\}} \\ &= \frac{\sum_{i \in M} \mathbb{P}(X_i > v)}{\mathbb{P}(\max_{i \in M} X_i > v)} \sum_K \mathbb{E}_{Q^v}\{g(\mathbf{X}_M)\mathbb{1}(\mathbf{X}_M \in R_K^v)/|K|\}, \end{aligned}$$

which gives the first equality in (19). For the second equality, note that

$$1 = \mathbb{E}_{P^v}\{\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\} = \frac{\sum_{i \in M} \mathbb{P}(X_i > v)}{\mathbb{P}(\max_{i \in M} X_i > v)} \sum_K \mathbb{E}_{Q^v}\{\mathbb{1}(\mathbf{X}_M \in R_{\max}^v)\mathbb{1}(\mathbf{X}_M \in R_K^v)/|K|\},$$

and dividing through by this gives the result. \square

We note that there is nothing specific to our general assumptions or extreme-value theory in this proposition: it holds for any random vector and any v whether v is extreme or not. For our purposes, v is a high marginal threshold and the distribution Q^v is obtained through asymptotic theory via Q_j^v , which can be simulated from as outlined in Section 5.1.

Thanks to Proposition 2, we can estimate $\mathbb{E}\{g(\mathbf{X}_M) | \max_{i \in M} X_i > v\} = \mathbb{E}_{P^v}\{g(\mathbf{X}_M)\}$ as follows:

Algorithm 3.

1. With probability $\pi_j = 1/m$, draw \mathbf{X}_M from Q_j^v , $j \in M$
2. Repeat step 1 n times to get n draws $\mathbf{X}_M^1, \dots, \mathbf{X}_M^n$ from Q^v
3. Estimate the expectation $\mathbb{E}_{P^v}\{g(\mathbf{X}_M)\}$ by

$$\frac{\sum_{i=1}^n g(\mathbf{X}_M^i) / |\{j \in M : X_j^i > v\}|}{\sum_{i=1}^n 1 / |\{j \in M : X_j^i > v\}|}, \quad \mathbf{X}_M^i \sim Q^v. \quad (22)$$

We note that the natural estimate of $\mathbb{E}_{P^v}\{g(\mathbf{X}_M)\}$ from equation (19) would look like

$$\frac{\sum_K \sum_{i=1}^n g(\mathbf{X}_M^i) \mathbb{1}(\mathbf{X}_M^i \in R_K) / |K|}{\sum_K \sum_{i=1}^n \mathbb{1}(\mathbf{X}_M^i \in R_K) / |K|}, \quad \mathbf{X}_M^i \sim Q^v,$$

but since it is only the size of K that matters we do not need to sum over all possible sets K , but simply divide by the number of variables exceeding the threshold. Furthermore, if it is desired to have realizations from the distribution conditioning upon the maximum being large, i.e., of $\mathbf{X}_M | \max_{i \in M} X_i > v$, this can be achieved approximately by using the importance weights in expression (22). That is, we sub-sample $n' < n$ realizations from the collection $\{\mathbf{X}_M^i\}_{i=1}^n$ with probabilities proportional to $\{1/|\{j : X_j^i > v\}|\}_{i=1}^n$.

5.2.3 Unconditional probabilities

Conditioning upon the process X being extreme anywhere in the domain creates a desirable definition of an extreme event that does not require reference to an arbitrary reference location s_0 . However, should we wish to undo the conditioning, we require estimation of the probability $\mathbb{P}(\max_{i \in M} X(s_i) > v)$, in place of the simpler marginal probability $\mathbb{P}(X(s_0) > v)$, which has a known form after marginal transformation. Nonetheless, Algorithm 3 can be exploited to estimate $\mathbb{P}(\max_{i \in M} X(s_i) > v)$, by noting that for any $j \in M$

$$\mathbb{P}\left(\max_{i \in M} X(s_i) > v\right) = \frac{\mathbb{P}(X(s_j) > v)}{\mathbb{P}(X(s_j) > v \mid \max_{i \in M} X(s_i) > v)},$$

for which the denominator can be estimated by Algorithm 3 by taking $g(\mathbf{X}_M) = \mathbb{1}(X_j > v)$.

5.3 Conditional infill simulation of an existing event

Assume that we have observations of the process X at a collection of sites indexed by $D' \subseteq D$ with $|D'| = d'$. We may wish to simulate X , conditionally upon the values of the observed process when extreme for at least one site, at an alternative collection of sites $\{s_{k_1}, \dots, s_{k_l}\}$, where we let $L = \{k_1, \dots, k_l\}$ with $L \cap D' = \emptyset$. If there are missing data, or for checking model fit, then a natural candidate is to take $L = D \setminus D'$, which leads to simulation at all observation locations, conditional upon a subset of them. In contrast to purely Gaussian models, conditional simulation is more challenging from models tailored to spatial extremes, although it is possible for models based on elliptical processes such as those described in Huser et al. (2017). An algorithm was proposed for max-stable processes by Dombry et al. (2013).

Consider a realization X_i , with $X_i(s_j) > u$ for all $j \in J_i \subseteq D$, and $X_i(s_j) < u$ for all $j \notin J_i$. For each $j \in J_i$, we have

$$Z_i^j(s) = \frac{X_i(s) - a_{s-s_j}(X_i(s_j))}{b_{s-s_j}(X_i(s_j))},$$

which, following the discussion in Section 3.2, is modelled as a (marginally-transformed) Gaussian process. Simulation of $\{Z^j(s_k), k \in L\} \mid \{Z^j(s_i), i \in D'\}$ can thus be achieved exploiting conditional simulation from a Gaussian process. Simplifying notation slightly, let $(\mathbf{Z}_L, \mathbf{Z}_{D'})$ represent an $(l + d')$ -dimensional random vector from the multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix Σ , and transformed to have delta-Laplace margins with mean vector $\boldsymbol{\mu}^\delta$ and scale parameter vector $\boldsymbol{\sigma}^\delta$. Partition $\boldsymbol{\mu} = (\boldsymbol{\mu}_L, \boldsymbol{\mu}_{D'})$, with $\boldsymbol{\mu}_L \in \mathbb{R}^l$, $\boldsymbol{\mu}_{D'} \in \mathbb{R}^{d'}$, and similarly for other vectors, whilst $\Sigma_{LD'} \in \mathbb{R}^{l \times d'}$ etc., represent the partitioned Σ . Let F_δ, Φ, f_δ and ϕ represent the univariate distribution functions and densities, respectively, of the delta-Laplace and Gaussian distributions, where $F_\delta(\mathbf{z}_L; \boldsymbol{\mu}_L^\delta, \boldsymbol{\sigma}_L^\delta) = (F_\delta(z_{L,k_1}; \mu_{L,k_1}^\delta, \sigma_{L,k_1}^\delta), \dots, F_\delta(z_{L,k_l}; \mu_{L,k_l}^\delta, \sigma_{L,k_l}^\delta))^\top$ etc., and denote by $\phi_l(\cdot; \mathbf{m}, \Omega)$ the l -dimensional multivariate Gaussian density with mean \mathbf{m} and covariance Ω .

The conditional density of $(\mathbf{Z}_L \mid \mathbf{Z}_{D'} = \mathbf{z}_{D'})$, denoted $f_{\mathbf{Z}_L \mid \mathbf{Z}_{D'}}$, can be expressed

$$f_{\mathbf{Z}_L \mid \mathbf{Z}_{D'}}(\mathbf{z}_L \mid \mathbf{z}_{D'}) = \phi_l \left[\Phi^{-1} \left\{ F_\delta(\mathbf{z}_L; \boldsymbol{\mu}_L^\delta, \boldsymbol{\sigma}_L^\delta); \boldsymbol{\mu}_L, \boldsymbol{\sigma}_L \right\}; \boldsymbol{\mu}_{L \mid D'}, \Sigma_{L \mid D'} \right] \prod_{k \in L} \frac{f_\delta(z_{L,k}; \mu_k^\delta, \sigma_k^\delta)}{\phi(\Phi^{-1}(F_\delta(z_{L,k}; \mu_{L,k}^\delta, \sigma_{L,k}^\delta); \mu_{L,k}, \sigma_{L,k}); \mu_{L,k}, \sigma_{L,k})}, \quad (23)$$

with

$$\begin{aligned} \boldsymbol{\mu}_{L \mid D'} &= \boldsymbol{\mu}_L - \Sigma_{LD'} \Sigma_{D'D'}^{-1} (\boldsymbol{\mu}_{D'} - \Phi^{-1} \{ F_\delta(\mathbf{z}_{D'}; \boldsymbol{\mu}_{D'}^\delta, \boldsymbol{\sigma}_{D'}^\delta); \boldsymbol{\mu}_{D'}, \boldsymbol{\sigma}_{D'} \}) \\ \Sigma_{L \mid D'} &= \Sigma_{LL} - \Sigma_{LD'} \Sigma_{D'D'}^{-1} \Sigma_{D'L}. \end{aligned}$$

To simulate from density (23):

Algorithm 4.

1. Given $\mathbf{z}_{D'}$, $\boldsymbol{\mu}, \Sigma, \boldsymbol{\mu}^\delta, \boldsymbol{\sigma}^\delta$, calculate $\boldsymbol{\mu}_{L \mid D'}$ and $\Sigma_{L \mid D'}$
2. Simulate \mathbf{Y}_L from the l -dimensional Gaussian with mean $\boldsymbol{\mu}_{L \mid D'}$ and covariance $\Sigma_{L \mid D'}$
3. Apply the transformation $\mathbf{Z}_L = F_\delta^{-1} \{ \Phi(\mathbf{Y}_L; \boldsymbol{\mu}_L, \boldsymbol{\sigma}_L); \boldsymbol{\mu}_L^\delta, \boldsymbol{\sigma}_L^\delta \}$.

Note the final step of Algorithm 4 is not a typical application of the probability integral transform, since the margins of \mathbf{Y}_L do not have location and scale parameters $\boldsymbol{\mu}_L, \boldsymbol{\sigma}_L$, but rather $\boldsymbol{\mu}_{L \mid D'}, \boldsymbol{\sigma}_{L \mid D'}$ where the latter is the square root of the diagonal of $\Sigma_{L \mid D'}$. When $\delta = 2$ and $\boldsymbol{\sigma}^\delta = \boldsymbol{\sigma} \{ \Gamma(1/2) / \Gamma(3/2) \}^{1/2} = 2^{1/2} \boldsymbol{\sigma}$, then this is just standard conditional simulation from the Gaussian.

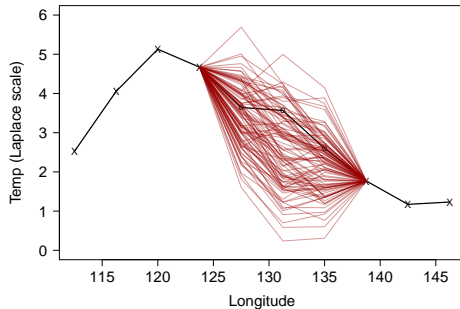


Figure 2: Example of conditional simulation. A realization of X (black solid line) is displayed at 10 sites but the information from three sites (depicted by circles) is treated as missing. The red lines depict conditional simulations based on the other seven values, and being extreme at one of the sites above the modelling threshold. The model is the fitted model from Section 6, fitted to all observations.

Once draws from $\{Z^j(s_k), k \in L\} | \{Z^j(s_i), i \in D'\}$ have been made, then the process $X_i(s) | X_i(s_j) > u$ is recovered by setting

$$X_i(s_k) | \{X_i(s_j) > u\} = a_{s_k-s_j}(X_i(s_j)) + b_{s_k-s_j}(X_i(s_j))Z^j(s_k), \quad k \in L,$$

and this can be done for each $j \in J_i$. That is, we are conditioning both on the values at observed sites, and on the process being extreme at a specific site $s_j \in J_i$.

The conditional simulation using Algorithm 4 is illustrated in Figure 2, using a subset of the Australian temperature data analyzed in Section 6. For these data, which are gridded and complete, conditional simulation may serve to help understand the fit of the model, but for other datasets it could be useful to deal with missing observations or to simulate at unobserved sites. The intended use of the conditional simulation may dictate whether one of these sites is more interesting or whether inference should be combined across all sites in J_i .

6 Australian temperature extremes

6.1 Data

We analyze temperature data for Australia from the HadGHCND global gridded dataset (Caesar et al., 2006). These are observed data translated onto a relatively coarse $2.5^\circ \times 3.75^\circ$ grid, with 72 points covering Australia; see the top-left panel of Figure 3. A complete record of daily maximum temperatures is available from 1/1/1957 - 31/12/2014, and to account for the seasonality of temperature data we focus on the summer temperatures recorded in December, January and February. This yields a total of 5234 days of observations at the 72 locations. The same data, recorded until the end of 2011, were analyzed in Winter et al. (2016) using non-spatial models, where the focus lay in investigating the effects of the El Niño Southern Oscillation (ENSO).

We transform the marginal distributions to unit Laplace, using the empirical distribution function at each site. In common with most works on extreme value dependence modelling, we do not attempt to explicitly account for the small amount of uncertainty introduced through use of estimated distribution functions. The choice of the double exponential-tailed Laplace distribution is motivated by the fact that we may anticipate independence at long range, and by allowing the margins of Z^0 to follow distribution (15), this feature can be captured. An alternative to empirical transformation is to use a semi-parametric marginal transformation, modelling the upper tail with a univariate generalized Pareto distribution. In high-dimensional datasets this becomes more laborious as one needs to select a threshold and check the model at each site; failing to do so can lead to poor marginal fits that may impact upon dependence structure estimation. On the other hand, the semi-parametric transformation is more useful for back-transforming simulations to the original scale in the extremes.

6.2 Exploratory analysis

Exploratory analysis indicates anisotropy and some spatial non-stationarity. This latter issue has received relatively little attention in an extreme value context, although Huser and Genton (2016) consider non-stationary max-stable processes when covariate information is available. In more classical spatial modelling, one approach to dealing with non-stationarity is to deform the coordinate system, as suggested by Sampson and Guttorp (1992). The essential idea of this technique is to transform from one coordinate system (the “G-plane”) in

which non-stationarity is detected, to another configuration (the ‘‘D-plane’’), where stationarity might be more reasonably assumed. Such deformations have the advantage of being based on the data only, without need for covariates. We choose here to adopt such an approach, using the methodology of [Smith \(1996\)](#). Denote by $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ the mapping from the G-plane to D-plane, with $s_j = (s_{j,1}, s_{j,2})^\top \in \mathbb{R}^2$ and $\tau(s_j) = (\tau(s_j)_1, \tau(s_j)_2)^\top \in \mathbb{R}^2$. The technique of [Smith \(1996\)](#) is designed for estimation on all data and is driven by the covariance matrix: the task is to find a deformation function τ such that $\text{Cov}\{Y(\tau(s_i)), Y(\tau(s_j))\}$ only depends on $\tau(s_i), \tau(s_j)$ through $\|\tau(s_i) - \tau(s_j)\|$. However, if patterns of non-stationarity are similar in the extremes to the body then this procedure should improve estimation in the extremes nonetheless. A similar argument was made by [Eastoe and Tawn \(2009\)](#) in relation to marginal non-stationarity due to covariate effects. Assuming d locations in the G-plane, [Smith \(1996\)](#) parameterizes

$$\tau(s_j) = \left(\kappa^2 s_{j,1} + \psi \kappa \lambda s_{j,2} + \sum_{i=1}^d \omega_{1,i} \xi_i(s_j), \quad \psi \kappa \lambda s_{j,1} + \lambda^2 s_{j,2} + \sum_{i=1}^d \omega_{2,i} \xi_i(s_j) \right)^\top, \quad \kappa, \lambda > 0, \psi \in \mathbb{R}, \quad (24)$$

with $\xi_i(s_j) = \frac{1}{2} \{(s_{j,1} - s_{i,1})^2 + (s_{j,2} - s_{i,2})^2\} \log\{(s_{j,1} - s_{i,1})^2 + (s_{j,2} - s_{i,2})^2\}$, and $\omega_{r,i}$, $r = 1, 2$, weights that satisfy $\sum_{i=1}^d \omega_{r,i} = \sum_{i=1}^d \omega_{r,i} s_{1,i} = \sum_{i=1}^d \omega_{r,i} s_{2,i} = 0$. In practice, following [Smith \(1996\)](#), we set most $\omega_{r,i}$ to zero, and focus on using a smaller number of geographically dispersed *anchor sites* to estimate the deformation. Figure 3 displays the G-plane and D-plane, highlighting the sites used in the estimation. Also displayed are estimates of $\chi_q(s_i, s_j)$ and $\chi_q(\tau(s_i), \tau(s_j))$ for $q = 0.95$ against distance in each coordinate system. The more stationary the data, the less variability one expects to observe in χ_q for a given spatial lag $\|s_i - s_j\|$ or $\|\tau(s_i) - \tau(s_j)\|$: a modest improvement seems to arise using the D-plane over the G-plane. A number of factors account for this: reduced anisotropy, accounting for differences in distance between 1° of latitude and longitude, as well as non-stationarity itself. Slightly different D-planes would be estimated using different anchor sites, and there is no clear way to optimize this aspect given the combinatorial possibilities. We proceed using the D-plane coordinates displayed in Figure 3.

6.3 Fitted model

To select a model, we applied the principles of Section 4.2. Based on likelihood values, Model 3 from Section 3 with Z^0 defined by a Gaussian process $Z_G | Z_G(s_0) = 0$, for which Z_G is stationary with powered exponential correlation, marginally transformed to delta-Laplace scale, was deemed to fit the data best. The threshold for observations being taken as extreme at a given site was set at the 97.5% quantile of the Laplace distribution, which yields an average of 130 exceedances per site. The fit of the model indicates asymptotic independence of the data and independence at long range, which is supported by estimates from a pairwise fit of the model, displayed in Figure 7 of Appendix B. Figure 7 supports parameterizing δ to decrease from 2 to 1 as distance increases, and so we set $\delta(s - s_0) = 1 + \exp\{-\|s - s_0\|/\delta_1\}^{\delta_2}$. Parameter estimates are displayed in Table 2; Table 3 in Appendix B recalls the definition of each parameter for convenience. Estimates of β were constrained to be less than one, but are very close. The uncertainty in the parameter estimates is visualized in Figure 8 of Appendix B. To produce this figure, we used a stationary bootstrap procedure ([Politis and Romano, 1994](#)) to account for the effect of temporal dependence, resampling entire fields with block lengths following a geometric distribution with mean length ten. The maximum composite likelihood estimates from the full dataset have then been subtracted to create a common scale.

Table 2: Parameter estimates to two decimal places.

	κ	λ	β	ϕ	μ	δ_1	δ_2	ν	σ
All	1.82	1.33	1.00	2.01	-0.08	1.08	1.74	1.89	0.88
Z only	-	-	-	2.06	-	1.36	2.31	1.87	0.97
El Niño	1.76	1.30	1.00	2.07	-0.06	0.84	1.36	1.87	0.86
Z only	-	-	-	2.16	-	1.28	2.10	1.85	0.97
La Niña	1.80	1.32	1.00	2.02	-0.03	0.99	1.65	1.87	0.87
Z only	-	-	-	2.12	-	1.37	2.43	1.84	0.99
La Nada	1.82	1.33	1.00	1.99	-0.10	1.08	1.74	1.88	0.87
Z only	-	-	-	2.07	-	1.37	2.32	1.86	0.97

The assumed model implies that the collection of residual processes $\{Z^j(s)\}_{j=1}^{72}$ have a distribution defined by

1. Dependence: $Z_G | Z_G(s_j) = 0$, where Z_G is a stationary Gaussian process with mean μ , and covariance function $\text{Cov}(s, s+h) = \sigma^2 \exp\{-\|h\|/\phi\}^\nu$;

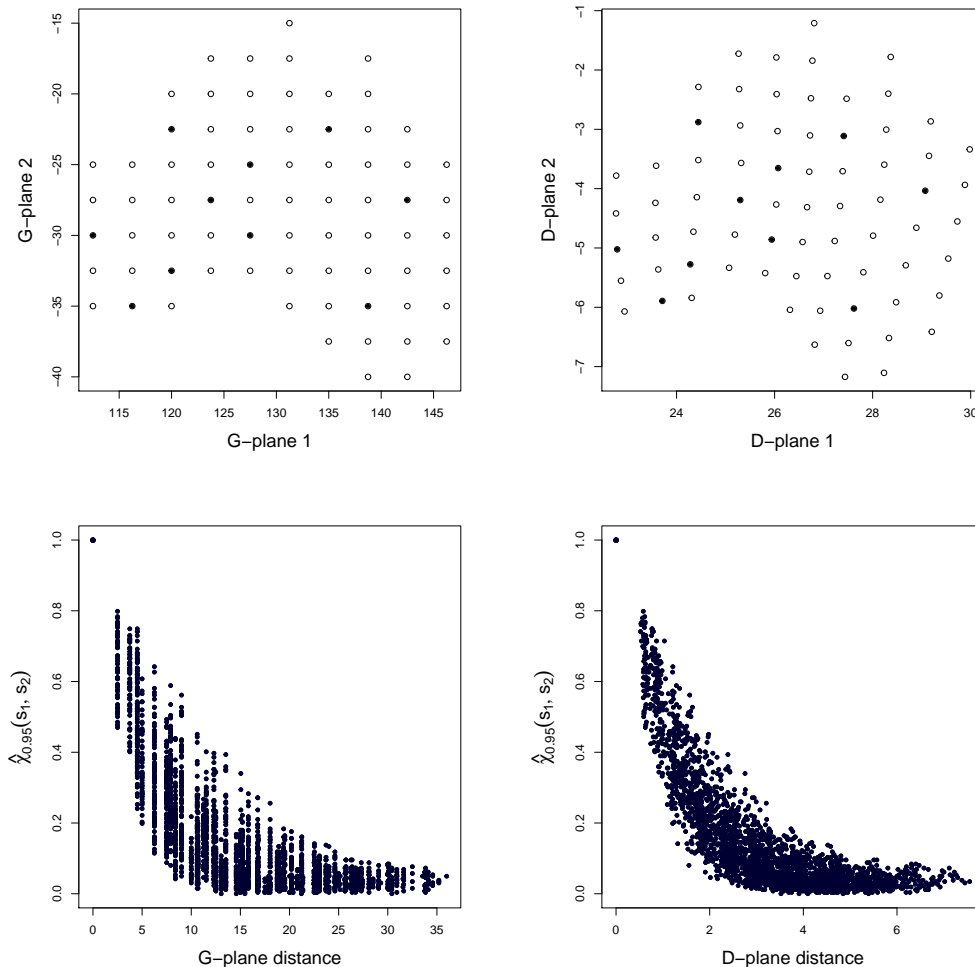


Figure 3: Top: G-plane and D-plane, with anchor sites for which $\omega_{r,i} \neq 0$ in (24) highlighted. Bottom: estimates of $\chi_q(s_i, s_j)$ (left) and $\chi_q(\tau(s_i), \tau(s_j))$ (right) for $q = 0.95$ plotted against distance in the coordinate system.

2. Margins: delta-Laplace distribution with shape $\delta(s - s_0)$, and location and scale parameter to match that of the conditional field $Z_G|Z_G(s_j) = 0$.

As mentioned in Section 4.2, one test of model fit therefore is to examine how closely the residual processes obtained by

$$Z^j(s) = \{X(s) - \hat{\alpha}(s - s_j)X(s_j)\} / [1 + \{\hat{\alpha}(s - s_j)X(s_j)\}^{\hat{\beta}}]$$

follow this structure. Upon examination of the empirical Z^j , we find a lack of fit, which translates into inadequate reproduction of extreme events. This emanates from an incorrect mean structure. For the possibilities outlined in Section 3.2, the mean of $Z_G|Z_G(s_j) = 0$ is either increasing or decreasing from zero, or identically zero, as $\|s - s_j\|$ increases. However, in practice here the mean increases then decreases again towards zero; see Figure 7. This is indicative of the long-range independence in this dataset, since if $X(s_k)$ is independent of $X(s_j)$, with $\alpha(s_k - s_j) \approx 0$, then $Z^j(s_k) = X(s_k)$, which has delta-Laplace margins with $\delta = 1, \sigma = 1, \mu = 0$. One remedy for this would be to attempt to parameterize the non-monotonic form. A further possibility in the current context of gridded data, is to extract the fitted Z^j and re-fit the model for these residual processes using the empirical means $(1/n_j) \sum_{i=1}^{n_j} Z_i^j(s_k)$, $k \in \{1, \dots, 72\} \setminus j$. A disadvantage of this approach is that it would not allow simulation at a new location without placing further spatial structure on the means. Where this is not an issue, as here, an advantage is that it can help alleviate symptoms of non-stationarity not accounted for by the deformation, and we thus adopt this approach. Table 2 also displays the parameter estimates for Z^j where the Gaussian process model with delta-Laplace margins has been re-fitted, using the empirical means. The estimates of δ_1, δ_2 and σ have the largest change, with the effect illustrated in Figure 7. Notably $\hat{\sigma} \approx 1$, ensuring the correct margins for $Z^0(s)$ for large $\|s - s_0\|$.

As a check on the modelling assumptions, we investigate the independence of $X(s_0)|X(s_0) > u$ and Z^0 . Figure 5 in Appendix B displays Kendall's τ coefficients between $X(s_j)|X(s_j) > u$ and the mean and variance of the corresponding $Z^j(s)$ for all conditioning sites s_j . Based on these diagnostics, independence of $X(s_0)|X(s_0) >$

u and Z^0 might be an adequate, but certainly not perfect, assumption. Nonetheless, parameter estimates do not change much when the model is fitted at a higher threshold. Furthermore, other diagnostics such as plots of data simulated from the model versus the observed data, such as the examples shown in Figure 6, suggest a reasonable model fit.

Using the parameter estimates from the re-fitted Z^j , and the original estimates for parameters not relating to Z^j , we employ the importance sampling techniques of Section 5.2.2 to estimate the expected number of grid locations exceeding a certain quantile, given that at least one location exceeds that quantile. Figure 4 displays results for quantiles ranging from 97.5% - 99.9999%, with empirical values where available; the in-sample agreement appears reasonable, although there is some question about whether the expected number decreases rapidly enough. Such a summary could not easily be calculated using the pairwise methods in Winter et al. (2016). To do so, one would have to fit $72 \times 71 = 5112$ pairwise models, calculate the Z^j by concatenating empirical residuals, and then using these with the 10,224 estimated parameters to implement the rejection scheme described in Section 5.2.1. Furthermore, the use of only empirical residuals restricts the shape of new events: where sites are effectively independent from the conditioning site, i.e., $a_{s-s_0}(x) \approx 0$, simulated new events will look just like past events in these areas. The figure also displays the estimate from the Brown–Resnick Pareto process, fitted to all processes for which the value $\max_{1 \leq j \leq 72} X(s_j)$ exceeds its 97.5% quantile. Whilst this provides a reasonable estimate close to the fitting threshold, it is clearly inappropriate for extrapolation further into the tail.

6.4 Investigation of El Niño effect

In Australia, the value of ENSO is known to impact the marginal distribution of extreme temperatures at specific locations (Winter et al., 2016). To investigate the possible effect of ENSO on the spatial extent of high temperature events, separate models were fitted to data from El Niño, La Niña and “La Nada” seasons. El Niño (respectively La Niña) events are defined here as those for which the sea surface temperature (SST) anomaly from 1980–2010 mean levels is higher than $+1^\circ$ (respectively lower than -1°); La Nada events are the remaining ones. The anomalies were taken from <https://www.esrl.noaa.gov/psd/gcos wgsp/Timeseries/Data/nino34.long.anom.data>. The covariate was defined by averaging monthly SST anomalies over the summer, so that each summer season is either El Niño, La Niña or La Nada. Marginal transformations were made separately for the three categories, whilst dependence parameter estimates are given in Table 2. There is some modest deviation from the combined parameter estimates particularly for the El Niño years where δ decays more rapidly, indicating slightly increased variability in these years. Figure 4 displays estimates of the expected number of exceedances from the models fit to the different ENSO regimes: La Niña years are slightly higher and El Niño slightly lower, but differences do not appear very large. However, differences in marginal quantiles make the practical interpretation less straightforward. In the South-East of Australia particularly, the high quantiles represent hotter temperatures under El Niño than La Niña, and a temperature that represents an exceedance of a marginal 99.5% quantile in La Niña conditions — which from Figure 4 occur in around 3.5 grid squares on average — might be closer to a 97.5% marginal quantile, or even lower, in El Niño conditions, which Figure 4 suggests might affect 5 or more grid squares on average. The ENSO regime may also affect the temporal persistence of extreme heat events, but this is not captured by our purely spatial models.

7 Discussion

The conditional approach to spatial extreme value analysis offers a number of advantages: flexible, asymptotically-motivated dependence structures that can capture asymptotic (in)dependence; the models can be fit in reasonably high dimensions; and conditional simulation at unobserved locations is simple. The principal drawback of the approach lies in the more complicated interpretation of what constitutes “the model” for a given dataset. When conditioning only on a single site being extreme, interpretation is straightforward. When wishing to condition on any of a set of sites being extreme, we need to combine these individual models in a way that entails no clearly defined overall model for the process given that it is extreme somewhere over space. However, for the purpose of *inference* on many quantities of interest, the importance sampling techniques ensure that this is not an issue.

Owing to computational limitations, there have been relatively few attempts at high-dimensional inference for extremes. In general, higher-dimensional problems do not only present computational issues, but they are often accompanied by additional modelling challenges. The spatial domain of interest is likely to be larger, leading to potentially more diverse behaviour in the dependence. In the Australian temperature example, we observed non-stationarity, and independence at longer range. These issues are less likely to arise when focusing on smaller areas, and these complexities should be kept in mind as spatial extremes moves into a higher-dimensional phase.

A natural next step is to extend this approach in to a multivariate or space-time setting. For example,

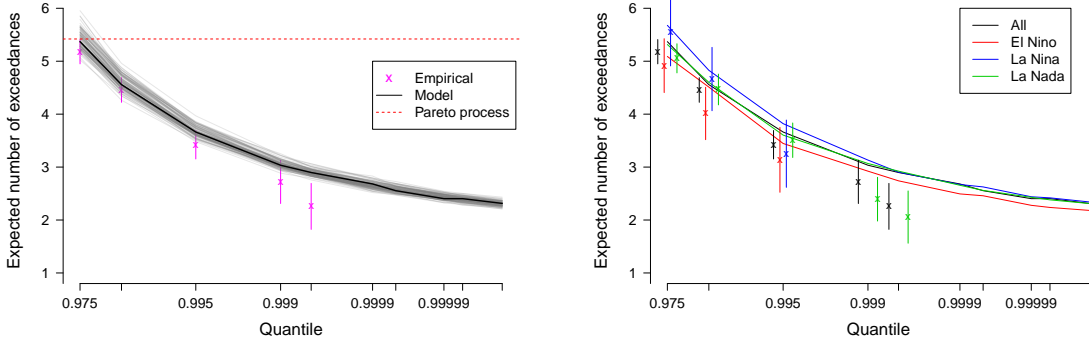


Figure 4: Expected number of exceedances of a particular marginal quantile given that there is at least one exceedance of that quantile somewhere over Australia. Points marked \times are this quantity estimated directly from the data, with bars giving approximate 95% confidence intervals; there are no data to estimate directly above the 99.95% quantile. Left: thick solid line represents model-based estimation, with estimates from bootstrap samples in grey. Right: different colours denote estimates from the models fitted to all data, as well as that stratified by ENSO covariate.

convergence (2) can be generalized by replacing all locations $s \in \mathcal{S}$ by $(s, t) \in \mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{R}$, with the conditioning location (s_0, t_0) . For modelling purposes, there are additional considerations of the dependence regime in both space and time, and how these interact. However, in principle one can have asymptotic (in)dependence in both space and time, with potentially different behaviour in the different dimensions. This spatio-temporal setting is the focus of recent work in [Simpson and Wadsworth \(2021\)](#).

Further extensions to this approach allow for different forms of the functions a_{s-s_0}, b_{s-s_0} or residual process Z^0 . Building on an early draft of this paper [Shooter et al. \(2019, 2021\)](#) have proposed alternative parametric models for the normalization functions, while [Simpson et al. \(2020\)](#) use a spline for the a_{s-s_0} term, avoiding an imposed parametric form. The latter also explore the use of Gaussian Markov random field structures for Z^0 in order to allow for a substantially larger number of observation locations, through the machinery of the integrated nested Laplace approximation (INLA).

Acknowledgements

We thank Simon Brown at the UK Met Office for the data analyzed in Section 6. J. Wadsworth gratefully acknowledges funding from EPSRC grant EP/P002838/1.

Data and code

The data and code for the analysis of Section 6 are available as Supplementary Material. The most recent data can be obtained from <https://www.metoffice.gov.uk/hadobs/hadghcnd/>, subject to the conditions detailed at the URL.

A Additional results and proofs

Before the proof of Proposition 1, Lemma 1 specifies conditions under which there is some flexibility in the normalization leading to convergence.

Lemma 1. *Suppose that for $\mathbf{a}(v)$ and $\mathbf{b}(v)$ with twice-differentiable components a_l, b_l satisfying $a'_l(v) \sim \alpha_l$, $a''_l(v) = o(1)$, $b'_l(v)/b_l(v) = o(1)$, as $v \rightarrow \infty$, for $l = 1, \dots, d$,*

$$\mathrm{P} \left(\frac{\mathbf{V} - \mathbf{a}(V_0)}{\mathbf{b}(V_0)} \leq \mathbf{z} \mid V_0 = v \right) = \mathrm{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) \rightarrow G(\mathbf{z}), \quad v \rightarrow \infty,$$

and that all first and second order partial derivatives converge, i.e.

$$\begin{aligned} \frac{\partial}{\partial z_k} \mathrm{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) &\rightarrow \frac{\partial}{\partial z_k} G(\mathbf{z}), \\ \frac{\partial^2}{\partial z_k \partial z_l} \mathrm{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) &\rightarrow \frac{\partial^2}{\partial z_k \partial z_l} G(\mathbf{z}), \quad v \rightarrow \infty. \end{aligned}$$

Then

(i) for $\mathbf{h}^1(v) = (h_1^1(v), \dots, h_d^1(v))^\top$, $\mathbf{h}^2(v) = (h_1^2(v), \dots, h_d^2(v))^\top$, with $h_l^1(v) = o(1)$, $h_l^2(v) = o(1)$, for all $l = 1, \dots, d$,

$$\mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)[1 + \mathbf{h}^1(v)]} + \mathbf{h}^2(v) \leq \mathbf{z} \mid V_0 = v \right) \rightarrow G(\mathbf{z}), \quad v \rightarrow \infty.$$

(ii) for \mathbf{h}^1 as above but $h_l^2(v) = O(1)$, i.e., some components may be asymptotically non-zero constants, whilst some may converge to zero,

$$\mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)[1 + \mathbf{h}^1(v)]} + \mathbf{h}^2(v) \leq \mathbf{z} \mid V_0 = v \right) \rightarrow G(\mathbf{z} - \lim_{v \rightarrow \infty} \mathbf{h}^2(v)), \quad v \rightarrow \infty.$$

Proof. (i) Write

$$\begin{aligned} \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) &= F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v), \\ \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)[1 + \mathbf{h}^1(v)]} + \mathbf{h}^2(v) \leq \mathbf{z} \mid V_0 = v \right) &= F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v) + \mathbf{b}(v)\mathbf{h}^1(v)\mathbf{z} - \mathbf{b}(v)\mathbf{h}^2(v) + O(\mathbf{b}(v)\mathbf{h}^1(v)\mathbf{h}^2(v))|v). \end{aligned}$$

Now write $\mathbf{g}(v) = \mathbf{b}(v)\mathbf{h}^1(v)\mathbf{z} - \mathbf{b}(v)\mathbf{h}^2(v) + O(\mathbf{b}(v)\mathbf{h}^1(v)\mathbf{h}^2(v))$, with $g_l(v) = o(b_l(v))$, and consider the Taylor expansion

$$F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v) + \mathbf{g}(v)|v) = F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v) + \nabla F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v)^\top \mathbf{g}(v) + O(\max(\vee_l h_l^1(v), \vee_l h_l^2(v))^2). \quad (25)$$

The components of $\nabla F_{V|V_0}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v)^\top$ are $F_{V|V_0}^{(l)}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v)$, where $F_{V|V_0}^{(l)}(\mathbf{x}|v) = \frac{\partial}{\partial y_l} F_{V|V_0}(\mathbf{y}|v)|_{\mathbf{y}=\mathbf{x}}$. The convergence

$$\frac{\partial}{\partial z_l} \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(V_0)}{\mathbf{b}(V_0)} \leq \mathbf{z} \mid V_0 = v \right) \rightarrow \frac{\partial}{\partial z_l} G(\mathbf{z})$$

is equivalent to

$$F_{V|V_0}^{(l)}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v)b_l(v) \rightarrow \frac{\partial}{\partial z_l} G(\mathbf{z}),$$

hence

$$F_{V|V_0}^{(l)}(\mathbf{b}(v)\mathbf{z} + \mathbf{a}(v)|v) = \frac{\partial}{\partial z_l} G(\mathbf{z}) \{b_l(v)\}^{-1} [1 + o(1)], \quad (26)$$

with a similar approach for the mixed partial derivatives. Substituting (26) into (25) yields that the second term in the expansion is $O(\max(\vee_l h_l^1(v), \vee_l h_l^2(v)))$, whilst the next order term is $O(\max(\vee_l h_l^1(v), \vee_l h_l^2(v))^2)$.

(ii) The proof is very similar to part (i), except the Taylor expansion is about $\mathbf{b}(v)(\mathbf{z} - \lim_{v \rightarrow \infty} \mathbf{h}^2(v)) + \mathbf{a}(v)$. \square

Proof of Proposition 1. Let $q_x^* = ([\min\{\min_l(a_l(x) + b_l(x)z_l), x\}]/\lambda)_+$ with $y_+ = \max(y, 0)$. We have

$$\begin{aligned} \lim_{x \rightarrow \infty} \mathbb{P} \left(\frac{\mathbf{X} - \mathbf{a}(X_0)}{\mathbf{b}(X_0)} \leq \mathbf{z} \mid X_0 = x \right) &= \lim_{x \rightarrow \infty} \int_0^{q_x^*} \mathbb{P} \left(\frac{\mathbf{X} - \mathbf{a}(X_0)}{\mathbf{b}(X_0)} \leq \mathbf{z}, Q = q \mid X_0 = x \right) dq \\ &= \lim_{x \rightarrow \infty} \int_0^{q_x^*} \mathbb{P} \left(\frac{\mathbf{X} - \mathbf{a}(x)}{\mathbf{b}(x)} \leq \mathbf{z} \mid Q = q, X_0 = x \right) f_{Q|X_0}(q|x) dq \\ &= \lim_{x \rightarrow \infty} \int_0^{q_x^*} \mathbb{P} \left(\frac{\mathbf{V} + \lambda q - \mathbf{a}(v(x) + \lambda q)}{\mathbf{b}(v(x) + \lambda q)} \leq \mathbf{z} \mid V_0 = v(x), Q = q \right) \\ &\quad \times f_{Q|X_0}(q|x) dq \\ &= \lim_{x \rightarrow \infty} \int_0^{q_x^*} \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v(x)) + (1 - \alpha)\lambda q + o(1)}{\mathbf{b}(v(x))[1 + O(\nabla \mathbf{b}(v(x))/\mathbf{b}(v(x)))]} \leq \mathbf{z} \mid V_0 = v(x), Q = q \right) \\ &\quad \times f_{Q|X_0}(q|x) dq, \end{aligned}$$

with $v(x) = x - \lambda q$ and $f_{Q|X_0}(q|x)$ the conditional density of $Q|X_0 = x$. The integrand is dominated by $f_{Q|X_0}(q|x)$, and

$$f_{Q|X_0}(q|x) = \frac{f_{Q,V_0}(q, x - \lambda q)}{f_{X_0}(x)} = \frac{e^{-q} e^{-(x - \lambda q)} \mathbb{1}(q > 0) \mathbb{1}(x - \lambda q > 0)}{e^{-x} [(1 - \lambda)^{-1} - (1 - \lambda)^{-1} e^{-(1/\lambda - 1)x}]} \rightarrow (1 - \lambda) e^{-(1 - \lambda)q} \mathbb{1}(q > 0), \quad x \rightarrow \infty;$$

note that $f_{Q|X_0}(q|x)$ is uniformly bounded by the integrable function $g(q) = (1 - \lambda) e^{-(1 - \lambda)q} [1 - e^{-(1/\lambda - 1)R}]^{-1}$ for all $x > R$. By Lemma 1, as $v \rightarrow \infty$,

$$\mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v) + (1 - \alpha)\lambda q + o(1)}{\mathbf{b}(v)[1 + O(\nabla \mathbf{b}(v)/\mathbf{b}(v))]} \leq \mathbf{z} \mid V = v, Q = q \right) \rightarrow G \left(\mathbf{z} + \frac{(\alpha - 1)\lambda q}{\lim_{v \rightarrow \infty} \mathbf{b}(v)} \right),$$

which simplifies when $\lim_{v \rightarrow \infty} b_l(v) = \infty$ for all l . Dominated convergence then yields the result stated. \square

Lemma 2 (Convergence of Gaussian partial derivatives). *Suppose $(\mathbf{Y}, Y_0)^\top$ follows a $(d + 1)$ -dimensional Gaussian distribution, and let $(\mathbf{V}, V_0)^\top = T((\mathbf{Y}, Y_0)^\top) \in \mathbb{R}_+^{d+1}$ be a random vector with unit exponential margins and Gaussian copula. Denote by $\boldsymbol{\rho}_0 > \mathbf{0}$ the d -vector of correlation parameters between \mathbf{Y} and Y_0 . Then for $\mathbf{a}(v) = \boldsymbol{\rho}_0^2 v$, $\mathbf{b}(v) = 1 + (\boldsymbol{\rho}_0 v)^{1/2}$ and any $r \leq d$,*

$$\frac{\partial^r}{\partial z_1 \cdots \partial z_r} \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) \rightarrow \frac{\partial^r}{\partial z_1 \cdots \partial z_r} G(\mathbf{z}).$$

Proof. We can express

$$\begin{aligned} \mathbb{P} \left(\frac{\mathbf{V} - \mathbf{a}(v)}{\mathbf{b}(v)} \leq \mathbf{z} \mid V_0 = v \right) &= \mathbb{P} \left[\mathbf{Y} \leq T^{-1} \{ \mathbf{b}(T(y)) \mathbf{z} + \mathbf{a}(T(y)) \} \mid T(Y_0) = T(y) \right] \\ &= \Phi_d [T^{-1} \{ \mathbf{b}(T(y)) \mathbf{z} + \mathbf{a}(T(y)) \} - \boldsymbol{\rho}_0 \mathbf{y}; \Sigma_0] \end{aligned}$$

where $\Phi_d(\cdot; \Sigma_0)$ is the cdf of the centred d -variate Gaussian with covariance matrix $\Sigma_0 = (\rho_{k,l} - \rho_{k,0}\rho_{l,0})_{1 \leq k, l \leq d}$. Taking the derivative yields

$$\begin{aligned} \frac{\partial^r}{\partial z_1 \cdots \partial z_r} \mathbb{P} \left[\mathbf{Y} \leq T^{-1} \{ \mathbf{b}(T(y)) \mathbf{z} + \mathbf{a}(T(y)) \} \mid T(Y_0) = T(y) \right] \\ = \Phi_d^{(1:r)} [T^{-1} \{ \mathbf{b}(T(y)) \mathbf{z} + \mathbf{a}(T(y)) \}; \Sigma_0] \prod_{k=1}^r \frac{\partial}{\partial z_k} T^{-1} \{ b_k(T(y)) z_k + a_k(T(y)) \}, \end{aligned} \quad (27)$$

where $\Phi_d^{(1:r)}(\cdot | \Sigma_0)$ is the r th-order mixed partial derivative of Φ_d . The components

$$\frac{\partial}{\partial z_k} T^{-1} \{ b_k(T(y)) z_k + a_k(T(y)) \} = \frac{b_k(T(y))}{T' [T^{-1} \{ b_k(T(y)) z_k + a_k(T(y)) \}]}. \quad (28)$$

Now $T(y) = -\log\{1 - \Phi(y)\} = y^2/2 + O(\log y)$, $y \rightarrow \infty$, whilst $T^{-1}(x) = (2x)^{1/2} + O(\log x/x^{1/2})$, $x \rightarrow \infty$, and $T'(y) = \phi(y)/\{1 - \Phi(y)\} \sim y$, $y \rightarrow \infty$. Further, $b_k(T(y)) = 1 + \rho_{0,k}y/\sqrt{2} + O(\log y)$, whilst $a_k(T(y)) = \rho_{0,k}^2 y^2/2 + O(\log y)$. Combining these,

$$T^{-1} \{ b_k(T(y)) z_k + a_k(T(y)) \} = \rho_{0,k} y + z_k/\sqrt{2} + o(1),$$

and equation (28) converges to $1/\sqrt{2}$, whilst (27) converges to

$$\Phi_d^{(1:r)} \left(\mathbf{z}/\sqrt{2}; \Sigma|_0 \right) 2^{-r/2},$$

which is the r th-order mixed partial derivative of the Gaussian limit distribution. \square

Remark 1. For application of Proposition 1, we also require convergence of the second derivative $\partial^2/(\partial z_k)^2$. Iteration of the above manipulations leads to a conclusion that the derivative converges to $\Phi_d^{(kk)}(\mathbf{z}/\sqrt{2}; \Sigma_0)/2$, with $\Phi_d^{(kk)}(\mathbf{z}) = \partial^2 \Phi_d(\mathbf{y})/(\partial y_k)^2|_{\mathbf{y}=\mathbf{z}}$.

Proposition 3. *Suppose that (X_1, X_2) have identical margins with infinite upper endpoint and*

1. $\lim_{x \rightarrow \infty} \mathbb{P}(X_1 > x | X_2 > x) = 0$
2. $\lim_{t \rightarrow \infty} \mathbb{P} \{ \{X_1 - a(X_2)\}/b(X_2) > z | X_2 > t \} = G(z)$,

with G a non-degenerate distribution function for which $\lim_{z \rightarrow \infty} G(z) = 1$. Then if there exists t_0 such that $a(t) > 0$ is monotonically non-decreasing in t for all $t > t_0$

(i) If $G(z) \in (0, 1)$ for some $z \in (0, \infty)$, $a(t) < t$ for all sufficiently large t .

(ii) If $G(z) \in (0, 1)$ for all $z \in (0, \infty)$, then $b(t) = o(t)$, $t \rightarrow \infty$.

Proof. (i). When $z > 0$,

$$0 \leq \mathbb{P}(X_1 > a(X_2) + b(X_2)z | X_2 > t) \leq \mathbb{P}(X_1 > a(X_2) | X_2 > t) \leq \mathbb{P}(X_1 > a(t) | X_2 > t),$$

for sufficiently large t . But since $\mathbb{P}(X_1 > t | X_2 > t) \rightarrow 0$, we must have $a(t) < t$ for all large t to get convergence to $G(z) \in (0, 1)$.

(ii). If $b(t)$ is constant or decreasing then clearly the statement holds, so suppose it is increasing as $t \rightarrow \infty$. Similarly to above

$$0 \leq \mathbb{P}(X_1 > a(X_2) + b(X_2)z | X_2 > t) \leq \mathbb{P}(X_1 > b(X_2)z | X_2 > t) \leq \mathbb{P}(X_1 > b(t)z | X_2 > t)$$

for $z > 0$ and large t . Now if $b(t) \sim ct$ for some $c > 0$ then for any $z > 1/c$, and sufficiently large t , $0 < \mathbb{P}(X_1 > b(t)z | X_2 > t) \leq \mathbb{P}(X_1 > t | X_2 > t) \rightarrow 0$. But since $G(z) \in (0, 1)$ for any $z \in (0, \infty)$, we must have $b(t) = o(t)$. \square

B Supporting information for Section 6

B.1 Model information

Table 3: Description of parameters in fitted model

Parameter	Description
κ	Shape parameter in $\alpha(s - s_0)$ (eq. (14))
λ	Scale parameter in $\alpha(s - s_0)$ (eq. (14))
β	Power in $b_{s-s_0} = 1 + a_{s-s_0}(x)^\beta$
ϕ	Scale parameter in Gaussian correlation
ν	Shape parameter in Gaussian correlation
σ	Scale parameter of Z_G used to determine structure of Z^j
μ	Mean parameter of Z_G used to determine structure of Z^j
δ_1	Scale parameter in $\delta(s - s_0) = 1 + \exp\{-\ s - s_0\ /\delta_1\}^{\delta_2}$
δ_2	Shape parameter in $\delta(s - s_0) = 1 + \exp\{-\ s - s_0\ /\delta_1\}^{\delta_2}$

B.2 Diagnostic plots and parameter uncertainty

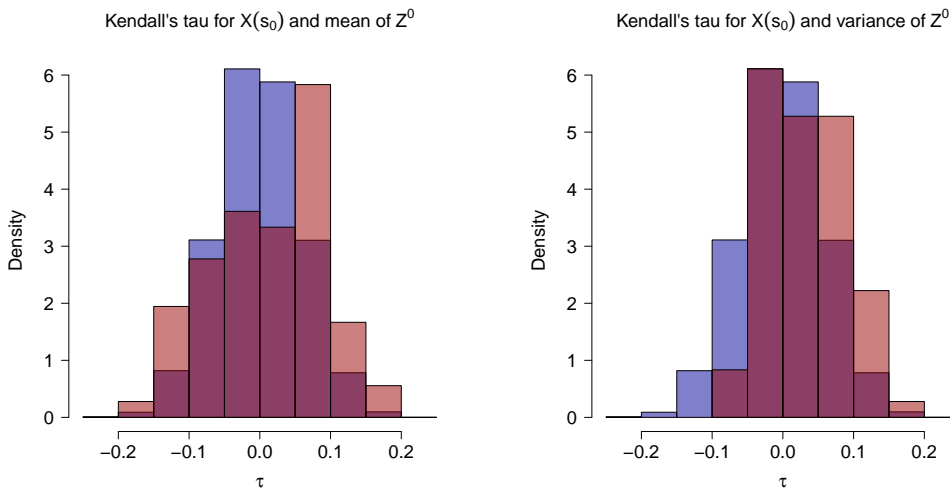


Figure 5: Histograms (red) of Kendall's τ coefficients for: $X^i(s_j) | X^i(s_j) > u$ and the mean (left) and variance (right) of $Z_i^j(s)$, $i = 1, \dots, n_j$ for the 72 sites, with average sample size $n_j \approx 130$. For comparison, in blue is the histogram of the approximate null Kendall's τ distribution obtained from 10000 samples of 130 independent bivariate datapoints to give an impression of the null distribution.

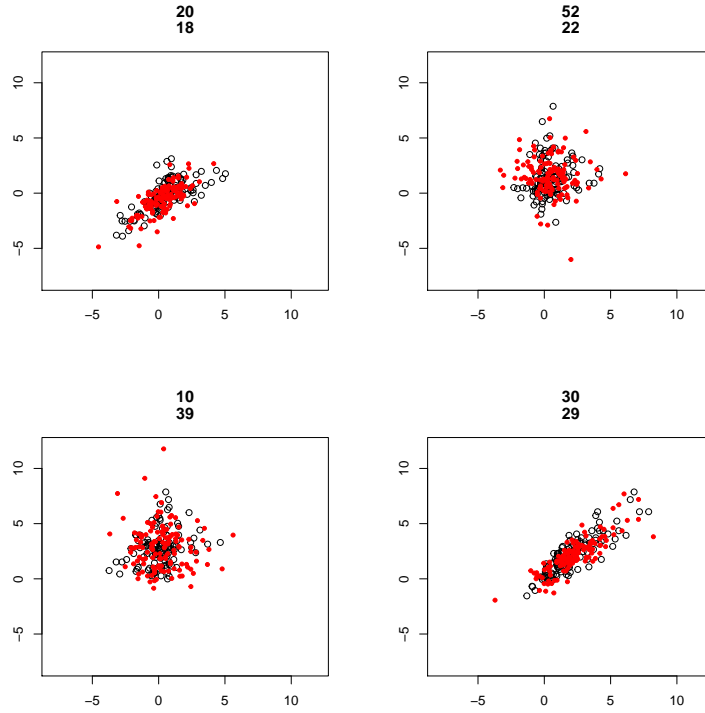


Figure 6: Example pairwise plots of observed data (black unfilled circles) and corresponding data simulated from the model (red filled circles). The site pairings are given in the figure heading, and the conditioning site is 48, which is not one of the sites appearing in the pairs. The observed data correspond to observations at the two sites when the observation at the conditioning site exceeds its threshold. The simulated data are generated from the model conditioning on the same site exceeding the same threshold.

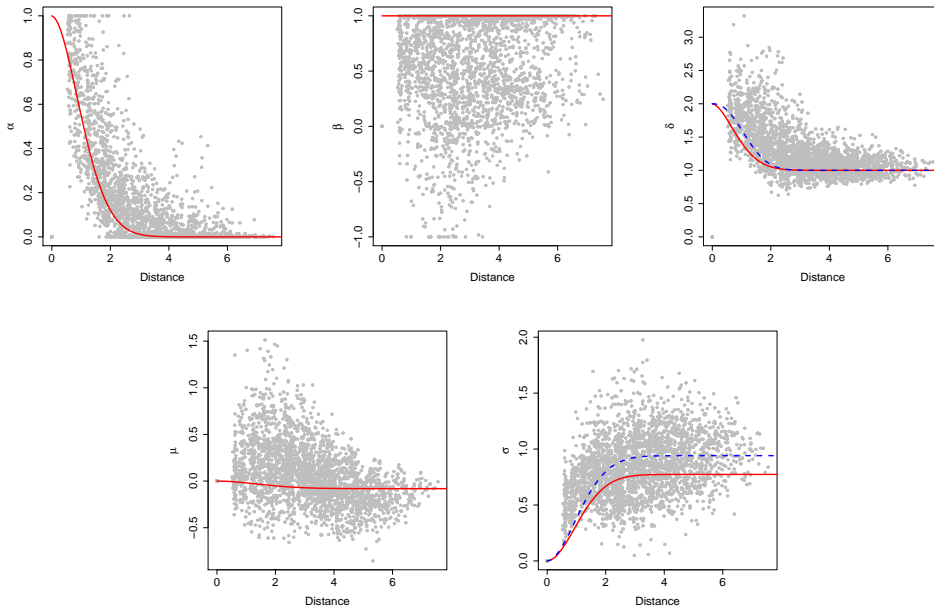


Figure 7: Parameter estimates from pairwise model fits using the same structure as Model 3, i.e., $a(x) = \alpha x$, $b(x) = 1 + (\alpha x)^\beta$ and Z following a delta Laplace distribution. Here β was constrained between $(-1, 1)$. Fits were made using composite likelihoods where a single set of parameters was assumed to apply for each pair, whichever the conditioning variable. Distance is in the deformed coordinate space. Solid red lines display implied values from the full fitted model; dashed blue lines display implied estimates from the model refitted to the extracted residual processes Z^j .

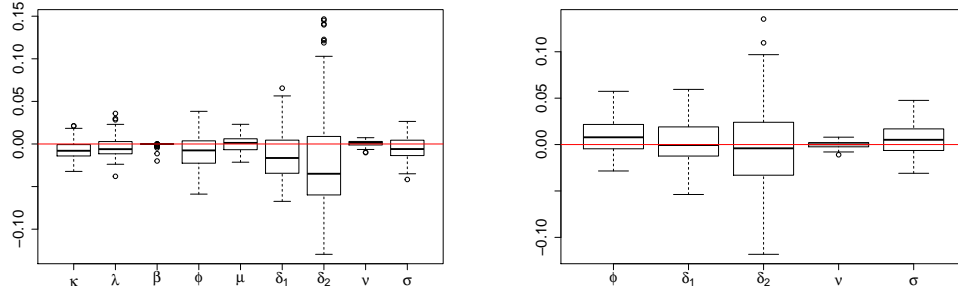


Figure 8: Distribution of estimates from 100 bootstrap repetitions, with population MLE subtracted. Left: full model fit; right: fit to extracted Z^j using empirical means.

C Computation time

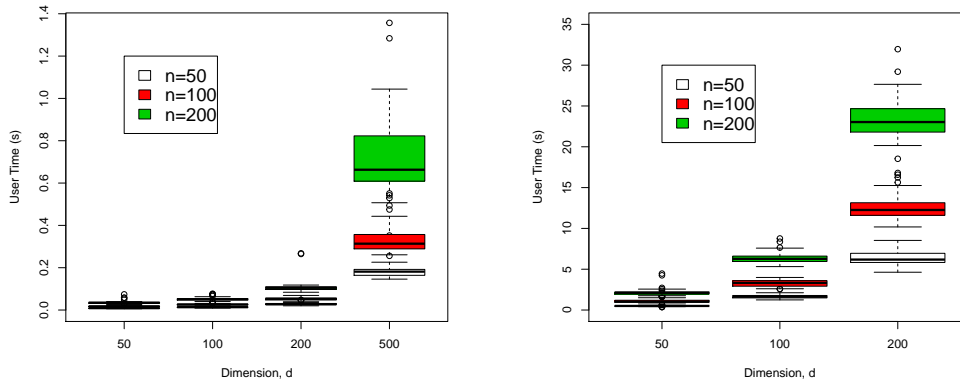


Figure 9: Left: computation time for a likelihood conditioning on a single site exceeding the threshold, for different numbers of sites, d , and average numbers of exceedances at the conditioning site, n . Right: computation time for a composite likelihood over all sites. The boxplots are created from 50 repetitions on different data.

Figure 9 is intended as a rough guide to how many dimensions one might reasonably attempt to optimize a likelihood in, based on our R code implementation, which makes use of the `mvnfast` R library (Fasiolo, 2016). Computation times for the composite likelihood generally slightly exceed d times the computation for the likelihood conditioning at a single site.

References

- Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080.
- Bortot, P., Coles, S. G., and Tawn, J. A. (2000). The multivariate Gaussian tail model: an application to oceanographic data. *Journal of the Royal Statistical Society: Series C*, 49(1):31–49.
- Caesar, J., Alexander, L., and Vose, R. (2006). Large-scale changes in observe daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres*, 111:1–10.
- Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.
- Das, B. and Resnick, S. I. (2011a). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252.
- Das, B. and Resnick, S. I. (2011b). Detecting a conditional extreme value model. *Extremes*, 14(1):29–61.

- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.
- de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.
- de Fondeville, R. and Davison, A. C. (2020). Functional peaks-over-threshold analysis. *arXiv:2002.02711*.
- de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: an Introduction*. Springer Science & Business Media.
- Dieker, A. and Mikosch, T. (2015). Exact simulation of Brown–Resnick random fields at a finite number of locations. *Extremes*, 18(2):301–314.
- Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124.
- Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and its Interface*, 8:9–17.
- Drees, H. and Janßen, A. (2017). Conditional extreme value models: fallacies and pitfalls. *Extremes*, 20(4):777–805.
- Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58(1):25–45.
- Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B*, 82(4):871–932.
- Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society: Series B*, 77:239–265.
- Engelke, S., Malinowski, A., Oesting, M., and Schlather, M. (2014). Statistical inference for max-stable processes by conditioning on extreme events. *Advances in Applied Probability*, 46:478–495.
- Fasiolo, M. (2016). *An introduction to mvnfast. R package version 0.1.6*.
- Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.
- Hazra, A., Huser, R., and Bolin, D. (2021). Realistic and fast modeling of spatial extremes over large geographical domains. *arXiv preprint arXiv:2112.10248*.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B*, 66(3):497–546.
- Hult, H. and Lindskog, F. (2005). Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their applications*, 115(2):249–274.
- Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103.
- Huser, R., Dombry, C., Ribatet, M., and Genton, M. G. (2019). Full likelihood inference for max-stable data. *Stat*, 8(1):e218.
- Huser, R. and Genton, M. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21:471–491.
- Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, 21:166–186.
- Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(1):e1537.
- Huser, R. G. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.

- 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.
- Keef, C. (2006). Spatial dependence of river flooding and extreme rainfall. PhD thesis, Lancaster University.
- Keef, C., Tawn, J. A., and Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.
- Nolde, N. and Wadsworth, J. L. (2021). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, to appear.
- 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.
- Papastathopoulos, I., Strokorb, K., Tawn, J. A., and Butler, A. (2017). Extreme events of Markov chains. *Advances in Applied Probability*, 49(1):134–161.
- Papastathopoulos, I. and Tawn, J. A. (2016). Conditioned limit laws for inverted max-stable processes. *Journal of Multivariate Analysis*, 150:214–228.
- Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.
- Resnick, S. I. and Zeber, D. (2014). Transition kernels and the conditional extreme value model. *Extremes*, 17(2):263–287.
- Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.
- Segers, J. (2020). One-versus multi-component regular variation and extremes of markov trees. *Advances in Applied Probability*, 52(3):855–878.
- Shooter, R., Ross, E., Tawn, J. A., and Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30(6):e2562.
- Shooter, R., Tawn, J. A., Ross, E., and Jonathan, P. (2021). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24(2):241–265.
- Simpson, E. S., Opitz, T., and Wadsworth, J. L. (2020). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *arXiv:2011.04486*.
- Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:100482.
- Smith, R. L. (1996). Estimating nonstationary spatial correlations. Unpublished manuscript.
- Stephenson, A. and Tawn, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227.
- Tawn, J. A., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28:39–58.
- Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *Annals of Applied Statistics*, 10(4):2303–2324.
- Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.
- Wadsworth, J. and Tawn, J. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.
- Wadsworth, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102(3):705–711.

- 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.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B*, 79(1):149–175.
- Winter, H. C., Tawn, J. A., and Brown, S. J. (2016). Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia. *Annals of Applied Statistics*, 10(4):2075–2101.
- Zhang, L., Shaby, B. A., and Wadsworth, J. L. (2021). Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations. *Journal of the American Statistical Association*, pages 1–13.