Spatial Statistics 28 (2018) 39–58

Contents lists available at ScienceDirect

Spatial Statistics

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

Modelling spatial extreme events with environmental applications

Jonathan Tawn [a,*], Rob Shooter [b], Ross Towe [c], Rob Lamb [d,e]

[a] Department of Mathematics and Statistics, Lancaster University, United Kingdom
[b] STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, United Kingdom
[c] School of Computing and Communications, Lancaster University, United Kingdom
[d] JBA Trust, Skipton, United Kingdom
[e] Lancaster Environment Centre, Lancaster University, United Kingdom

ARTICLE INFO

Article history:
Received 31 October 2017
Accepted 24 April 2018
Available online 4 May 2018

Keywords:
Conditional multivariate extreme values
Copula
Gaussian processes
Max-stable processes
Pareto processes
Spatial extremes

ABSTRACT

Spatial extreme value analysis has been an area of rapid growth in the last decade. The focus has been on modelling the spatial componentwise maxima by max-stable processes. Here, we will explain the limitations of these modelling approaches and show how spatial models can be developed that overcome these deficiencies by exploiting the flexible conditional multivariate extremes models of Heffernan and Tawn (2004). We illustrate the benefits of these new spatial models through applications to North Sea wave analysis and to widespread UK river flood risk analysis.

1. Introduction

In many environmental applications data are collected from a number of spatial locations, for example numerous locations across an ocean basin or locations across a river network. Historically interest has been in the extremal behaviour at individual sites. However, our interest lies in developing a framework in which it is possible to estimate probabilities of joint events over space. For example, for wave heights we may want to know the probability of no offshore structure being damaged in a storm, and for river levels the probability that the total damages from a flood exceed £1 billion. Probabilities of the occurrence of extreme spatial events are of particular interest to the reinsurance industry for deriving aggregate financial loss distributions, and also to governments in terms of risk assessment and emergency planning.

* Corresponding author.
E-mail address: j.tawn@lancaster.ac.uk (J. Tawn).

https://doi.org/10.1016/j.spasta.2018.04.007

To answer such questions we take an asymptotically justified model for the joint occurrence of extreme values of an event over space. Our reason for this is that we aim to extrapolate to spatial events that are larger than any previously observed, so we cannot rely on empirical evidence alone. Asymptotic theory therefore provides a principled approach to develop our models and understanding. Such a spatial model requires both marginal distributions and the dependence structure of the spatial process to be explicitly characterised. It is the challenge of modelling the extremal dependence structure that will be the primary focus of this paper. As closed form probabilities cannot be derived for the spatial events of interest to us, we aim to develop methods that enable straightforward simulation of extreme spatial events from which probabilities can be derived using Monte Carlo methods.

Let $\{Y(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ denote a stationary spatial process indexed by s over a set $\mathcal{S}$ with marginal distribution function F which has upper endpoint y_F. In practice we observe replicates of $\{Y(s) : s \in \mathcal{S}\}$ at a finite set of points $\{Y(s_j) : j = 1, \ldots, n\}$, and at times $t = 1, \ldots, n$. Hence $Y_t(s)$ denotes the process observed at time t at location s. We are interested in the extreme values of Y over the entire set of $\mathcal{S}$. For this paper, we assume that the entire spatial process is independent and identically distributed in time, i.e., $\{Y_i(s); s \in \mathcal{S}\}$ is independent of $\{Y_j(s); s \in \mathcal{S}\}$ for all $i, j = 1, \ldots, n$ with $i \neq j$. Thus our focus is on the spatial dependence behaviour of the process only. However, unlike in many applications of spatial statistics, we have a large number of independent and identically distributed replicates of the spatial process from which to make our inference.

In many spatial extreme value problems the aim is to characterise the extremal behaviour of the spatial process $Y(s)$. A complication is that without a natural ordering scheme in more than the one dimension the definition of an extreme event is not well-defined. A range of approaches can be taken, as follows.

Max-stable processes Consider componentwise maxima over n independent and identically distributed copies of $\{Y(s), s \in \mathcal{S}\}$, i.e.,

$$\{M_n(s); s \in \mathcal{S}\} = \{\max_{1 \leq t \leq n} Y_t(s); s \in \mathcal{S}\}. \tag{1.1}$$

Here, and throughout this paper, operations are carried out componentwise, i.e., site specifically.

Pareto processes Consider the process obtained by characterising the limiting behaviour of

$$\{Y(s); s \in \mathcal{S} \mid \max_{s \in \mathcal{S}} Y(s) > u\} \tag{1.2}$$

as $u \to y_F$.

Conditional extremes processes We propose to characterise the behaviour of

$$\{Y(s); s \in \mathcal{S} \mid Y(s_0) > u\} \tag{1.3}$$

for any $s_0 \in \mathcal{S}$ as $u \to y_F$.

When suitably linearly normalised, $\{M_n(s); s \in \mathcal{S}\}$ converges (as $n \to \infty$) to a max-stable process; see Smith (1990), Schlather (2002), Padoan et al. (2010) and Davison et al. (2012). This is the most widely used approach to spatial extremes due to its historical link to the families of univariate and multivariate extreme value distributions (all finite dimensional distributions of a max-stable process are multivariate extreme distributions) and also for its elegant mathematical properties. However, this approach cannot be used to answer questions about original events for $Y(s)$ since $M_n(s)$ is a composition of a number of different events, and hence this formulation cannot be used to answer our motivating questions. Furthermore, the spatial dependence structure for $M_n(s)$ is restrictive and so fails to accommodate a wide class of events including Gaussian processes; see the discussion of $\chi(\tau)$ below.

Using the underlying mathematical formulation of max-stable processes, Ferreira and de Haan (2014) obtain a limiting form of the process (1.2), which we outline in Section 2.3. Note that Dombry and Ribatet (2015) alternatively condition on other functionals of the process being extreme, and obtain a class of limiting processes known as ℓ-Pareto processes.

Our proposal differs in two ways from that used for Pareto or ℓ-Pareto processes. We condition on the extreme event in conditional representation (1.3) being large at a specific site. We also

exploit the normalisation structure from Heffernan and Tawn (2004) in the conditional approach (1.3) that uses a different normalisation of $Y(s)$ to achieve a more general (and more flexible) limiting representation. We will take the conditional extremes process approach (1.3) which we outline in Section 3.2. However, we also give further details of max-stable and Pareto processes to help explain their weaknesses for our needs and to show how our approach differs from them.

To help to first identify the differences between the approaches, let us introduce two pairwise spatial extremal dependence measures, $\{\chi(\tau), \bar{\chi}(\tau)\}$, which are natural extensions of multivariate measures defined by Coles et al. (1999) to stationary spatial processes. Consider a pair of sites $(s, s+\tau)$, each in $\mathcal{S}$. Then $\chi(\tau)$ is defined by the following limit probability

$$\chi(\tau) = \lim_{y \to y_F} \mathbb{P}(Y(s+\tau) > y \mid Y(s) > y), \tag{1.4}$$

if it exists. Additionally, $\bar{\chi}(\tau)$ is determined by the following asymptotic equivalence, as $y \to y_F$

$$\mathbb{P}(Y(s+\tau) > y \mid Y(s) > y) \sim \mathcal{L}\left(1/\bar{F}(y)\right) \{\bar{F}(y)\}^{[1-\bar{\chi}(\tau)]/[1+\bar{\chi}(\tau)]},$$

where $\mathcal{L}$ is a slowly varying function at infinity and $\bar{F}(y) = 1 - F(y)$. Here $0 \leq \chi(\tau) \leq 1$ and $-1 < \bar{\chi}(\tau) \leq 1$. For each of $\chi(\tau)$ and $\bar{\chi}(\tau)$, larger values correspond to stronger levels of extremal dependence.

If $\chi(\tau) > 0$, then $\bar{\chi}(\tau) = 1$ and the largest values of the process can occur simultaneously at two sites τ apart, a property known as asymptotic dependence at lag τ. However, if $\chi(\tau) = 0$ then in the limit the largest values at sites τ distance apart must occur in different spatial events, and the process is said to have asymptotic independence at τ. For processes with $\chi(\tau) = 0$, the quantity $\bar{\chi}(\tau)$ is a helpful measure for determining the level of asymptotic independence since it controls the rate at which $\mathbb{P}(Y(s+\tau) > y \mid Y(s) > y)$ converges to zero. In particular, $0 < \bar{\chi}(\tau) \leq 1$ corresponds to positive extremal dependence, $\bar{\chi}(\tau) = 0$ to near extremal independence, and $-1 < \bar{\chi}(\tau) < 0$ to negative extremal dependence.

Determining the pair $\{\chi(\tau), \bar{\chi}(\tau)\}$, for all τ, provides a good summary of the extremal properties of the process. Some spatial extreme value modelling approaches preclude certain types of extremal dependence. For example, for all non-degenerate max-stable processes or Pareto processes that are dependent at lag τ then $\{\chi(\tau), \bar{\chi}(\tau)\} = (c_\tau, 1)$, for some $0 < c_\tau < 1$. However, for all non-degenerate Gaussian processes $\{\chi(\tau), \bar{\chi}(\tau)\} = (0, \rho(\tau))$, where $\rho(\tau)$ is the correlation of the Gaussian process at lag τ. Thus max-stable and Pareto processes are asymptotically dependent, whereas Gaussian processes are asymptotically independent. These measures show that max-stable and Pareto processes fail to capture the spatial extremal dependence features of Gaussian processes. Consequently, if the data were from a Gaussian process but a max-stable process model was fitted then there will be an over-estimation of the risk of jointly large events. Therefore a broader class of spatial extreme value models is required if we are to capture the dependence structures of both these important classes of spatial process. The models we will introduce here have this capability, as well as having sufficient structure in order to model our applications well.

The conditional multivariate extreme value model of Heffernan and Tawn (2004) estimates the form of extremal dependence structure (asymptotic dependence or asymptotic independence) as part of the fitting procedure. The model can handle high dimensional problems (Winter et al., 2016), extremal temporal dependence (Winter and Tawn, 2017), missing values (Keef et al., 2009) and negative dependence (Keef et al., 2013). Examples of the environmental applications include heatwaves, hydrology and oceanography (Jonathan et al., 2013; Keef et al., 2009; Towe et al., 2017; Winter and Tawn, 2016). Here we outline how these multivariate methods can be extended to a spatial framework and clarify what they offer over existing spatial extreme value models.

The layout of the paper is as follows. Section 2 details existing statistical models for spatial extreme values. Section 3 presents the conditional multivariate extreme value model of Heffernan and Tawn (2004) and outlines how this model can be extended to handle spatial extreme problems. Finally, Section 4 details two applications of the methodology to oceanography and hydrology; the first of these relates to understanding the extremal dependence of significant wave heights over the North Sea and the second addresses questions on widespread risk of flooding raised by the UK Government's 2016 National Flood Resilience Review.

2. Existing methods

2.1. Univariate modelling

Underpinning the two main distributions of univariate extreme value theory are representational characterisations of max-stability and threshold-stability which uniquely define these distributions. Here we recap these features in the univariate case, as they provide the core structure for the existing spatial extremal theory.

Much classical extreme value theory is based on the property of max-stability that leads to the extremal types theorem of Fisher and Tippett (1928). For independent and identically distributed univariate random variables $\{Y_i; i = 1, \ldots, n\}$, with continuous but otherwise arbitrary distribution function F with upper endpoint y_F, let $M_n = \max\{Y_1, \ldots, Y_n\}$. If there are normalising sequences $a_n > 0$ and b_n such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x) \qquad (n \to \infty), \tag{2.1}$$

where G is a non-degenerate distribution function, then G is of the form

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

with parameters $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ corresponding to location, scale and shape parameters and $\{z\}_+ = \max\{0, z\}$. This is known as the generalised extreme value (GEV) distribution, and is denoted GEV(μ, σ, ξ). This class of distributions uniquely satisfies the max-stability property which says that for all $m \in \mathbb{N}$ and $x \in \mathbb{R}$, there are constants $A_m > 0$, B_m such that

$$\{G(A_m x + B_m)\}^m = G(x).$$

Thus the GEV is the only non-degenerate distribution that is closed to the operation of maximisation.

An alternative approach to modelling univariate extremes is to focus on the exceedances of a threshold u. Pickands (1975) showed that if there is a non-degenerate limit (2.1), then there exists a normalising function $c(u) > 0$ such that as $u \to y_F$,

$$\frac{Y - u}{c(u)} \mid Y > u \xrightarrow{d} V,$$

where convergence is in distribution and V is non-degenerate. Then V follows a generalised Pareto distribution, which we denote GPD(ψ, ξ), with distribution function

$$H(x) = 1 - \left(1 + \frac{\xi x}{\psi}\right)_+^{-\frac{1}{\xi}}, \qquad (x > 0), \tag{2.2}$$

with scale parameter $\psi > 0$ and shape parameter $\xi \in \mathbb{R}$.

The characterising property of the GPD is that of threshold stability (Davison and Smith, 1990), that is, for any $v > 0$, there exists a function $c(v) > 0$ such that

$$\left\{\frac{V - v}{c(v)}\right\} \mid V > v \overset{d}{=} V. \tag{2.3}$$

Thus scaled excesses of a higher threshold v by V have the same distribution as V. This is illustrated in Fig. 1. The GPD is the only distribution with this threshold-stability property.

Based on this asymptotic justification, we make the modelling assumption that the distribution of excesses of $Y(s)$ over a high threshold u follows the limiting distribution for excesses exactly, i.e.,

$$Y(s) - u \mid Y(s) > u \overset{d}{=} V(s) \mid V(s) > 0 \qquad (s \in \mathcal{S}).$$

Consequently, the margins of $Y(s)$ are GPD(ψ, ξ) distributed above the threshold u, where ψ and ξ do not depend on $s \in \mathcal{S}$ as the $Y(s)$ process is stationary. Since the above assumption provides

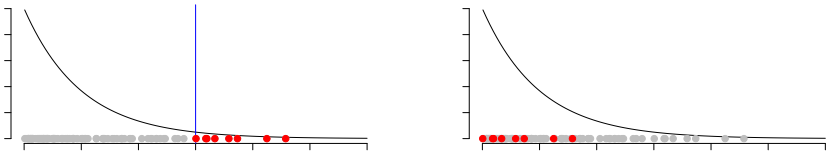


Fig. 1. Illustration of threshold stability property described by relationship (2.3). The left panel shows a sample from $V \sim \text{GPD}(\psi, \xi)$ with the vertical line representing the threshold v and the red points the exceedances of v; the right panel shows these same exceedances (shown as excesses in red) after scaling (here the GPD has parameters $(\psi, \xi) = (1, 0)$, and so $c_v = 1$). These scaled excesses are compared against a new sample (in grey) from the original distribution of V, we note that these two samples have the same distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

no information on the marginal behaviour below u, the empirical distribution is used below this threshold (Coles and Tawn, 1991). The resulting model for the marginal distribution function is

$$F(x) = \begin{cases} \tilde{F}(x) & \text{if } x \le u \\ 1 - [1 - \tilde{F}(u)]\left[1 + \dfrac{\xi(x-u)}{\psi}\right]_+^{-\frac{1}{\xi}} & \text{if } x > u, \end{cases}$$

where $\tilde{F}(x)$ is the empirical distribution function of all of the data at all sites. Due to stationarity of the process, data at all locations can be used to estimate F.

The study of dependence structure is typically undertaken via copulas (Nelsen, 2006), which requires the marginal distributions to be identical and uniformly distributed. Although we have identical margins, we prefer to transform them to non-uniform margins, via the pointwise transformation

$$X_t(s) = K^{-1}\{F(Y_t(s))\} \qquad (s \in \mathcal{S}, t = 1, \ldots, n),$$

so that $X_t(s)$ is a spatial process, independent over time, and with marginal distribution function K. We perform this transformation as the extremal dependence properties of $X_t(s)$ are more simply expressed for some non-uniform marginal choices.

The most convenient choice of K depends on the context: the Fréchet or Pareto distributions are typically assumed for max-stable distributions (Resnick, 1987, 2013); for conditional extremes, Heffernan and Tawn (2004) use Gumbel margins; for joint tail modelling, Wadsworth and Tawn (2012) used exponential margins while Keef et al. (2013) showed that Laplace margins allow negative dependence to be incorporated the most parsimoniously. Critically, Gumbel, Exponential and Laplace distributions all have exponential upper tails, so if negative dependence is avoided (which is reasonable in most spatial extremes applications) they are essentially identical approaches for our purposes. Here we take $X_t(s)$ to have Gumbel marginals, so that $K(x) = \exp\{-\exp(-x)\}$, as this gives the clearest link to the max-stable results; since $\exp\{X_t(s)\}$ has Fréchet margins. Thus, results in Fréchet margins translate to results in Gumbel margins via a log transformation.

We now have that $\{X_t(s); s \in \mathcal{S}\}$ is a stationary spatial process with Gumbel margins. Although the copula/dependence structure of this process is restricted by the stationarity of the process, the range of choice of models is nonetheless vast. We saw, in the univariate case, that looking at the extremes of the variable reduced the class of possible continuous distributions to either the GEV or GPD depending on the extremal feature that is studied. For the dependence structure similar simplifications arise by imposing max-stability and threshold stability in spatial contexts. We explore these two strategies in Sections 2.2 and 2.3 respectively.

2.2. Max-stable processes

Given that $\{X_t(s); s \in \mathcal{S}\}$ has Gumbel margins, it follows from (1.1) and (2.1) that we can take $a_n = 1$ and $b_n = \log n$ which gives $Z(s)$, defined by

$$Z(s) \overset{d}{=} \lim_{n\to\infty} \left\{ \max_{t=1,\ldots,n} X_t(s) - \log n \right\} \qquad (s \in \mathcal{S}),$$

to be a max-stable process with Gumbel margins. As a consequence of the $Z(s)$ process being max-stable, for any d sites $\{s_1, \ldots, s_d\}$ in $\mathcal{S}$ then $\{Z(s_1), \ldots, Z(s_d)\}$ with distribution function G is max-stable, i.e., for all $m \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^d$,

$$\{G(\mathbf{x} + \log m)\}^m = G(\mathbf{x}),$$

so the joint distribution is stable with respect to taking componentwise maxima. From the characterisation of de Haan (1984) and Schlather (2002), the max-stable process $Z(\cdot)$ takes the form

$$Z(s) = \max_{i \geq 1}\{R_i + W_i(s)\} \qquad (s \in \mathcal{S}), \tag{2.4}$$

where $\{R_i, i \in \mathbb{N}\}$ are the points of a Poisson process on $\mathbb{R}$ with intensity $\exp(-x)dx$ and the $W_i(s)$ over i are independent and identically distributed stochastic processes with continuous sample paths such that

$$\mathbb{E}[\exp\{W_i(s)\}] = 1 \qquad (i \in \mathbb{N}, s \in \mathcal{S}).$$

Note that the additive structure is identical to the usual product structure, with the difference arising due the change in choice of marginal distributions. When $W(\cdot)$ is a Gaussian process with a particular moment structure, this gives the Brown–Resnick process for $Z(\cdot)$ (Brown and Resnick, 1977; Davison et al., 2012). A weakness with this model is that G can only be specified via a series of evaluations of the multivariate normal distribution function (Genton et al., 2011), though reductions in the numerical difficulties can be achieved using methods of Wadsworth and Tawn (2012) that require additional information about which segments of $Z(s)$ arise from the same $Y_t(s)$ process.

2.3. Pareto processes

An alternative asymptotic characterisation for spatial extremes is to use the threshold exceedance analogue of max-stable processes, namely generalised Pareto processes (Ferreira and de Haan, 2014). The strategy behind this development is a spatial extension of the argument that led to the GPD in the univariate case, i.e., we condition on an extreme event occurring and then study the properties of this extreme event as the threshold that determines the extreme event tends to a limiting value. Specifically, define the process $T(s)$ by

$$\{T(s); s \in \mathcal{S}\} :\overset{d}{=} \lim_{u \to \infty} \left[\{X(s) - u; s \in \mathcal{S}\} \mid \sup_{s \in \mathcal{S}} X(s) > u \right].$$

Then $T(s)$ is a Pareto process, with the property that $\sup_{s \in \mathcal{S}} T(s)$ is distributed as a standard exponential random variable but that $T(s)$ can be negative for some values of $s \in \mathcal{S}$. Critically, for all $v > 0$, $T(s)$ then satisfies

$$\{T(s) - v \mid \sup_{s \in \mathcal{S}} T(s) > v\} \overset{d}{=} T(s),$$

so that $T(\cdot)$ satisfies the threshold-stability property. Pareto processes are the only such processes that possess this property. This property is illustrated in Fig. 2, which shows a set of realisations of the process $X(s)$ in black with a subset (indicated in red) corresponding to realisations with $\sup_{s \in \mathcal{S}} X(s) > u$. Thus each of the red realisations is approximately a Pareto process, i.e., $u + T(s)$.

To help study Pareto processes it is helpful to draw on the max-stable characterisation (2.4) of Ferreira and de Haan (2014). A Pareto process is simply one of the latent processes that underpin the $Z(s)$ process. It follows that we can represent the Pareto process $T(s)$ by

$$T(s) = R + W(s), \tag{2.5}$$

where R is a standard exponential random variable which is independent of a stochastic process $W(\cdot)$, satisfying $\sup_{s \in \mathcal{S}} W(s) = 0$. A common choice for this is to set $W(\cdot)$ to be a Gaussian process, such as the Gaussian process family used for Brown–Resnick processes (Brown and Resnick, 1977). In this case, $W(\cdot)$ is a conditional Gaussian process, conditional on $\sup_{s \in \mathcal{S}} W(s) = 0$. A benefit of working with Pareto processes over max-stable processes is that the process is derived from a single realisation of

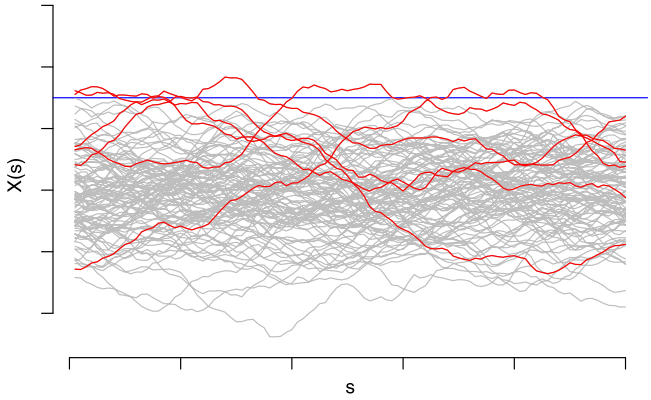


Fig. 2. Illustration of a Pareto process, showing realisations of a process $X(s)$ (grey lines), where for some chosen threshold u (blue line), with the realisations where $\sup_{s\in\mathcal{S}} X(s) > u$ (red lines) being approximately distributed as $u + T(s)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$W(\cdot)$ and R. Therefore, conditionally on R, the $T(s)$ process is a conditional Gaussian process which is massive simplification of inference relative to max-stable processes. However, the conditioning for $W(\cdot)$ is complex as it applies over all $s \in \mathcal{S}$, which makes computation non-trivial.

2.4. Weakness of Pareto processes

Assuming that the process $X(s)$, when it exceeds a threshold u, is exactly a Pareto process means that for large u, $X(s) = u + T(s)$. Hence, for some s_0, $s \in \mathcal{S}$, we have

$$X(s_0) = u + R + W(s_0) \text{ and } X(s) = u + R + W(s),$$

where R is a standard Exponential random variable and $W(s)$ is independent of R, so that when $X(s_0)$ is large,

$$X(s) = X(s_0) + \{W(s) - W(s_0)\}.$$

Then $X(s_0)$ is interpretable as the size of the event and $\{W(s) - W(s_0)\}$ as the spatial profile of the event. Critically, the shape and size of these extreme events are independent for Pareto processes. Thus events are equally likely to retain the same type of spatial profile whatever their size at a point s_0. An illustration of this is shown in the top row of panels in Fig. 3, with the profile of the events unchanged as the size of events increases (left to right panels). As a consequence, Pareto processes are asymptotically dependent at all lags, as

$$\lim_{x\to\infty} \mathbb{P}(X(s) > x | X(s_0) > x) > 0 \qquad (s_0, s \in \mathcal{S}).$$

However, in practice we almost never observe such processes. Instead, we often see events becoming more localised, as seen in the bottom row of panels in Fig. 3. Here we see events of the small initial magnitude and profile as in the top row become more spatially localised around the maximum value as the maximum value of the field increases. For this type of process, which include Gaussian processes,

$$\lim_{x\to\infty} \mathbb{P}(X(s) > x | X(s_0) > x) = 0 \qquad (s_0, s \in \mathcal{S}, s \neq s_0),$$

so the process is asymptotically independent at all lags.

It may be that both of these formulations are too simplistic and the process is asymptotically dependent up to a certain lag h_{AD}, then asymptotically independent when the lag exceeds h_{AD}, such as in the models of Bacro et al. (2016). Consequently, we want an inference method which does not

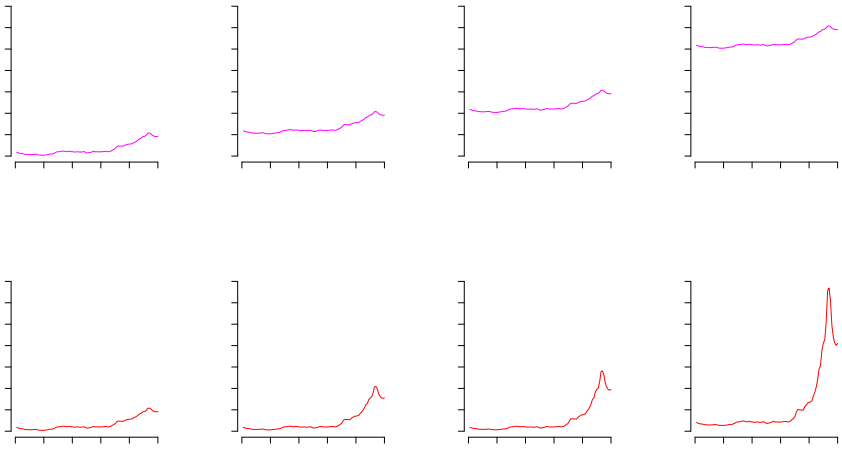


Fig. 3. Illustration of types of extremal spatial behaviour. The top row shows a process which retains the same spatial profile as the event becomes more extreme, corresponding to asymptotic dependence. The bottom row depicts the extreme event becoming more localised as its magnitude increases, commonly seen in practice and corresponding to asymptotic independence.

pre-determine that the process is asymptotically dependent at all lags, so that $h_{AD} = \infty$ (like max-stable and Pareto processes), or asymptotically independent at all lags with $h_{AD} = 0$ (like Gaussian processes). In particular, we would like to have the flexibility to determine the lag h_{AD} at which this transition occurs. The models introduced in Section 3 do precisely that.

3. Conditional extremes

3.1. Asymptotics for conditional multivariate extremes

Consider a vector random variable $\mathbf{X} = (X_1, \ldots, X_d)$ with Gumbel marginals; for $i < j$, we shall use the notation $\mathbf{X}_{i:j} = (X_i, \ldots, X_j)$. For simplicity, we will assume that all the variables are non-negatively dependent and that $\mathbf{X}$ has a joint density.

Heffernan and Tawn (2004) propose an asymptotically justified conditional multivariate extremes approach for modelling the extremes of a vector $\mathbf{X}$ given X_1 is large. To explore the conditional distribution $\mathbb{P}\{\mathbf{X} \leq \mathbf{x} \mid X_1 > u\}$ for large u, we use an asymptotically justified form for this distribution as $u \to \infty$. If $\mathbf{x}$ is fixed, in general the limit distribution will be a degenerate distribution. Hence $\mathbf{X}$ needs to be normalised appropriately so that the limiting conditional distribution is non-degenerate as $u \to \infty$. Heffernan and Resnick (2007) propose that $\mathbf{X}_{2:d}$ is linearly normalised as a function of either X_1 or u. Normalising by X_1 leads to simpler limit models, thus we use the approach of Heffernan and Tawn (2004) and carry out this normalisation.

Heffernan and Tawn (2004) assume that there exist functions $a\colon \mathbb{R} \to \mathbb{R}^{d-1}$ and $b\colon \mathbb{R} \to \mathbb{R}_+^{d-1}$, such that for $x > 0$,

$$\mathbb{P}\left(\frac{\mathbf{X}_{2:d} - a(X_1)}{b(X_1)} \leq \mathbf{z}_{2:d}, X_1 - u > x \mid X_1 > u\right) \to G_{2:d}(\mathbf{z}_{2:d}) \exp(-x), \tag{3.1}$$

as $u \to \infty$ with $\mathbf{z}_{2:d} \in \mathbb{R}^{d-1}$ and where $G_{2:d}$ is a joint distribution function that is non-degenerate in each margin. A key property of the limit (3.1) is that the limiting distribution factorises, corresponding to large values of X_1 being independent of the associated normalised $\mathbf{X}_{2:d}$.

Under weak assumptions on the joint distribution of $\mathbf{X}$, Heffernan and Resnick (2007) show that, componentwise, a and b must be regularly varying functions satisfying certain constraints, which for Gumbel margins corresponds to each of the components of a (respectively b) being regularly varying

functions of index 1 (respectively less than 1). Within this structure (Heffernan and Tawn, 2004) found that a simple form for a and b holds for a very broad range of copulas. In particular, they assume that

$$a(x) = \boldsymbol{\alpha}_{2:d} x \text{ and } b(x) = x^{\boldsymbol{\beta}_{2:d}}$$

where $\boldsymbol{\alpha}_{2:d} = (\alpha_2, \ldots, \alpha_d) \in [0, 1]^{d-1}$ and $\boldsymbol{\beta}_{2:d} = (\beta_2, \ldots, \beta_d) \in [0, 1)^{d-1}$. This canonical parametric subfamily of a and b provides a parsimonious, yet flexible, family for statistical modelling.

Different types of extremal dependence lead to different values of the extremal dependence parameters $\boldsymbol{\alpha}_{2:d}$ and $\boldsymbol{\beta}_{2:d}$. For $2 \leq j \leq d$, when $\alpha_j = 1$ and $\beta_j = 0$ the variables (X_1, X_j) are asymptotically dependent; when $\alpha_j < 1$, these variables are asymptotically independent. Within the asymptotic independence case a further resolution of the dependence structure is possible, with $0 < \alpha_j < 1$ or $\alpha_j = 0$ and $\beta_j > 0$ corresponding to positive dependence, and near independence when $\alpha_j = \beta_j = 0$. When there is a multivariate normal copula (with $\rho_{ij} > 0$ corresponding to the correlation parameter between variables i and j), then $\alpha_j = (\rho_{1j})^2$, $\beta_j = 1/2$ and $G_{2:d}$ is the joint distribution function of a multivariate Normal distribution which has mean vector $\mathbf{0}$, variance (for the jth variable) of $2\rho_{1j}^2(1 - \rho_{1j}^2)$ and a correlation between variables i and j of $(\rho_{ij} - \rho_{1i}\rho_{1j})/[(1 - \rho_{1i}^2)(1 - \rho_{1j}^2)]^{1/2}$; see Heffernan and Tawn (2004).

Unfortunately there is no finite parametric form for $G_{2:d}$ or its marginal distributions, so a range of approaches have been taken. Heffernan and Tawn (2004) use empirical estimates for $G_{2:d}$; Lugrin et al. (2016) utilise a mixture of Gaussian distributions, while Towe et al. (2016) use a Gaussian copula with kernel smoothed marginal distributions. Here, we make the assumption that $G_{2:d}$ is multivariate normal with margins $N(\mu_j, \sigma_j^2)$ for $j = 2, \ldots, d$. Under this assumption,

$$X_j \mid \{X_1 = x\} \sim N\left(\alpha_j x + \mu_j x^{\beta_j}, \sigma_j^2 x^{2\beta_j}\right) \qquad (x > u, j = 2, \ldots, d), \tag{3.2}$$

with parameters $\boldsymbol{\alpha}_{2:d}$, $\boldsymbol{\beta}_{2:d}$, $\boldsymbol{\mu}_{2:d} = (\mu_2, \ldots, \mu_d)$ and $\boldsymbol{\sigma}_{2:d} = (\sigma_2, \ldots, \sigma_d)$.

3.1.1. Inference

In order to estimate the dependence parameters $\boldsymbol{\alpha}_{2:d}$ and $\boldsymbol{\beta}_{2:d}$, a pseudo-likelihood is constructed with $\mathbf{X}_{2:d} \mid X_1 = x$ (for $x > u$) treated as independent with marginals of the joint conditional distribution stated in Eq. (3.2). The estimation of these dependence parameters is performed through maximum pseudo-likelihood for the n_u pairs for which $X_1 > u$. The likelihood is then

$$L\left(\boldsymbol{\alpha}_{2:d}, \boldsymbol{\beta}_{2:d}, \boldsymbol{\mu}_{2:d}, \boldsymbol{\sigma}_{2:d}\right) \propto \prod_{i=2}^{d} \prod_{j=1}^{n_u} \frac{1}{x_{ij}^{\beta_i} \sigma_i} \exp\left\{ -\frac{\left(x_{ij} - \left[\alpha_i x_{1j} + \mu_i x_{1j}^{\beta_i}\right]\right)^2}{2 x_{ij}^{2\beta_i} \sigma_i^2} \right\},$$

for $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $-1 \leq \alpha_i \leq 1$, and $-\infty < \beta_i < 1$ for $i = 2, \ldots, d$, and where x_{ij} denotes component i for the jth exceedance of u by X_1. The maximum pseudo-likelihood estimates are denoted by $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\sigma}}$. Then realisations of $\mathbf{Z}_{2:d} \sim G_{2:d}$ are given by

$$\mathbf{z}_{2:d}^{(j)} = \left(\frac{x_{ij} - \hat{\alpha}_i x_{1j}}{(x_{1j})^{\hat{\beta}_i}}, i = 2, \ldots, d \right) \text{ for } j = 1, \ldots, n_u \tag{3.3}$$

where $x_{1j} > u$ for each j. This sample of $\mathbf{Z}_{2:d}$ is used to obtain an empirical estimate of the joint distribution function $G_{2:d}$. Consequently, we have a model for the joint tail behaviour of $\mathbf{X}$, when X_1 is large. This enables us to make inferences beyond the range of the observed data with large X_1; for more details of fitting these models over different conditioning variables and methods for simulating jointly rare events see Heffernan and Tawn (2004) and Keef et al. (2013).

A limitation of the inference for models in the conditional multivariate extremes approach is that self-consistency of the different conditional distributions is not ensured. This may lead to inconsistencies when calculating joint exceedance probabilities such as

$$\begin{aligned} \mathbb{P}(X_1 > u, X_2 > u) &= \mathbb{P}(X_1 > u | X_2 > u) \cdot \mathbb{P}(X_2 > u) \\ &= \mathbb{P}(X_2 > u | X_1 > u) \cdot \mathbb{P}(X_1 > u), \end{aligned}$$

since the models for $X_1|X_2 > u$ and $X_2|X_1 > u$ are not necessarily equal. Liu and Tawn (2014) discussed this problem, making a range of proposals to reduce this problem. One proposal which removes the issue is to assume that (X_1, X_2) are exchangeable, which implies for that the associated parameters and distributions are equal for each conditional distribution. For non-exchangeable pairs though, whilst removing the self-consistency problems, this induces biased inference.

3.2. Models for conditional spatial extremes

This section gives an indication only of how some aspects of the multivariate conditional extremes methods could be extended to the spatial setting. For simplicity, it is assumed that $X(s)$ is isotropic as well as stationary and with Gumbel marginals, and let $h = |s - s_0|$ be the distance between two sites $s_0, s \in \mathcal{S}$. A consequence of these standard spatial statistics assumptions is that the joint distribution of pairs $\{X(s_1), X(s_2)\}$ are exchangeable variables, for all pairs $s_1, s_2 \in \mathcal{S}$, and hence there are none of the issues of self-consistency that are present in multivariate cases.

The natural spatial extension of the Heffernan and Tawn (2004) conditional multivariate extremes representation to the spatial context assumes that there exist normalisation functions $\alpha(h) \in [0, 1]$ and $\beta(h) \in [0, 1)$ for all $h > 0$, with $\alpha(0) = 1,\ \beta(0) = 0$, such that as $u \to \infty$,

$$\left\{ \frac{X(s) - \alpha(h)X(s_0)}{X(s_0)^{\beta(h)}} : s \in \mathcal{S}, X(s_0) - u > x \right\} \mid X(s_0) > u \xrightarrow{d} \{\mu(h) + \sigma(h)Z(s) : s \in \mathcal{S}, E\},$$

where, $\mu(\cdot)$ and $\sigma(\cdot)$ are deterministic functions with $\sigma(h) > 0$ for $h \neq 0$ and $\mu(0) = \sigma(0) = 0$; $Z(\cdot)$ is a random process with $\mathbb{E}[Z(s)] = 0$ and $\mathrm{Var}[Z(s)] = 1$ for all $s \in \mathcal{S}$ and E is a standard Exponential random variable that is independent of the process $Z(\cdot)$.

Assuming that this limit result holds exactly for a large choice of threshold u gives a model structure

$$X(s)|\{X(s_0) > u\} = \alpha(h)X(s_0) + X(s_0)^{\beta(h)}W(s - s_0) \qquad (s \in \mathcal{S}), \tag{3.4}$$

where $\{X(s_0) - u\}|X(s_0) > u$ follows a standard exponential distribution and is independent of $W(\cdot)$, where $W(s) := \mu(h) + \sigma(h)Z(s)$ is a spatial isotropic process with $W(0) = 0$, marginal mean $\mu(h)$, marginal variance $\sigma^2(h)$ and correlation function $\rho(\cdot)$. As in the multivariate conditional extremes case, we will make a modelling assumption that $W(\cdot)$ is a Gaussian process with a correlation structure to be estimated. This Gaussian assumption may appear to be a very strong assumption but it is the assumed process for all Brown–Resnick max-stable processes (Davison et al., 2012), for the type of processes given in Engelke et al. (2015) and in a conditional form for Pareto processes (Ferreira and de Haan, 2014).

The key is then to make inference on $\alpha(h)$, $\beta(h)$, $\mu(h)$, $\sigma(h)$ and the correlation structure of $W(\cdot)$ so that inference can be drawn on the process (1.3) (after back transformation from $X(s)$ to $Y(s)$). There are some interesting special cases of this model:

Pareto type process If $\alpha(h) = 1$ and $\beta(h) = 0$ for all $h \geq 0$, then model (3.4) is exactly that given by the process of Engelke et al. (2015) and is strongly related to the Pareto process, given by expression (2.5), as it is essentially the same process but subject to different conditioning constraints. It is asymptotically dependent at all lags.

Gaussian process From results in Section 3.1 on multivariate normal copulas, $\{\alpha(h)\}^{1/2}$ satisfies the properties of a valid spatial correlation function and $\beta(h) = 1/2$ for $h > 0$, then model (3.4) is exactly the limiting conditional extremal process of a Gaussian process; it is asymptotically independent for all positive lags.

Mixture process If $(\alpha(h), \beta(h)) = (1, 0)$ for all $h \leq h_{AD}$ but $\alpha(h) < 1$ for $h > h_{AD}$ then the process is asymptotically dependent up to lag h_{AD} and asymptotically independent otherwise.

The aim therefore is to identify if any of these structures is present in an application. To help give insight into these three different sub-classes of model (3.4), in Fig. 4 we show repeated simulations of a 1-dimensional process with $X(0)$ equal to the marginal 99.995% quantile, thus all simulations are

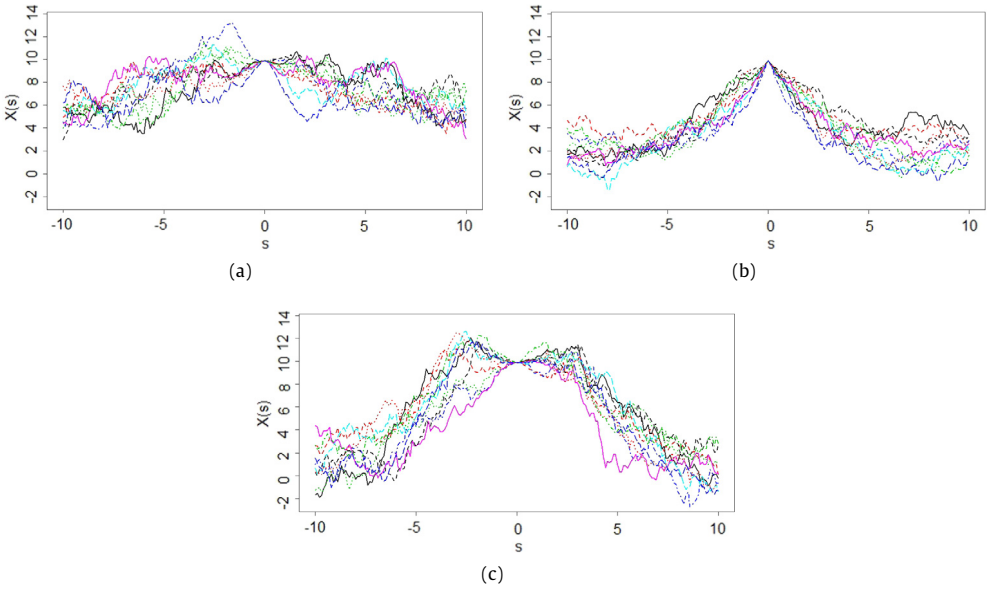


Fig. 4. Illustrations of Pareto type, Gaussian and mixture extremal processes on a space $\mathcal{S} = (-10, 10)$. In all cases $X(0)$ is in an extreme state (equal to the 99.995% marginal quantile), and the latent Gaussian process $Z(s)$ has mean and standard deviation of $\mu(h) = \mu_c$ and $\sigma(h) = \sigma_c$ for $h > 0$ and correlation function $\rho(h) = \exp(-h/3)$. Illustration is as follows: (a) Pareto type process with $\mu_c = -0.4, \sigma_c^2 = 1.3$; (b) Gaussian process $\alpha(h) = \exp(-h/3), \mu_c = 0.06, \sigma_c^2 = 0.6$; (c) mixture process with $h_{AD} = 3$; $\alpha(h) = \exp(-|h - h_{AD}|/3)$ for $h > 3$, $\beta(h) = 0$, $\mu_c = -0.05, \sigma_c^2 = 1.3$.

equal for $s = 0$. Firstly, we can see that the three types of process behave differently from one another in the location of a large event, with all replications for a given process type having broadly similar behaviour. Secondly, note that if $X(0)$ was more or less extreme the only effect would be a vertical shift of the process when the process is in on-extreme states.

Pareto type processes remain of the same order of magnitude over the space $\mathcal{S}$. Specifically, it has a mean negative drift away from an extreme level, with here, due to the choice of correlation function and the Gaussian process for $Z(s)$, in the neighbourhood of $s = 0$ the extremal process is a Brownian motion with negative drift in distance $|s|$ from the extreme event. Consequently there is a positive probability of $X(\tau)$ being large given $X(0)$ is large for all $s \in \mathcal{S}$, hence the process is asymptotically dependent for all lags τ as defined by definition (1.4). In contrast, for the extremal Gaussian process events decay much more rapidly, essentially geometrically, until the process returns to a non-extremal state. Thus, it can be seen that the process is asymptotically independent for all lags τ, but with the rate of convergence of the non-limit probability in definition (1.4) to 0 is dependent on τ. The mixture type processes behave like Pareto type processes up to lag h_{AD} from the extreme event at $s = 0$, but then decay more rapidly to until the process returns to a non-extremal state. Hence the mixture process is seen to be asymptotically dependent up to lag h_{AD} and asymptotically independent for larger lags.

4. Applications

4.1. Offshore risk from waves

4.1.1. Background

The accurate modelling of extreme wave heights is of key importance in the design of offshore structures. Such structures must be constructed adhering to strict guidelines, which themselves rely

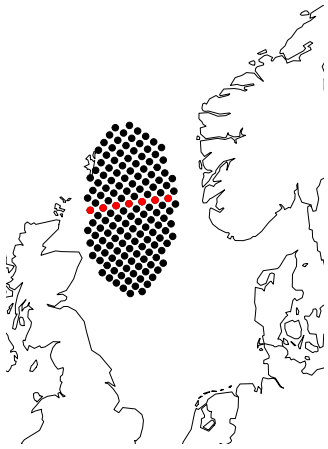


Fig. 5. Map of sampling locations in the North Sea from which the data are collected, with the particular transect used for model fitting highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on the assessment of how often extreme events occur. Methods for spatial extremes are useful for enabling the likelihood over sites to be constructed for improved marginal parameter inference and for spatial risk assessment over a network of offshore structures. For the former, we need a reliable spatial dependence model to ensure valid inferences are made for the smoothly varying marginal parameter models (Randell et al., 2015). For the latter, companies with offshore interests often have more than one asset to insure and so having a joint risk assessment that gives the probability than none of the assets will be affected in their lifetime is required.

The aim of our analysis is to test the viability of the conditional spatial extremes methods set out in Section 3.2 for application to significant wave data (defined as four times the standard deviation of the sea-surface) in the North Sea region shown in Fig. 5. The data come from a numerical model driven by observational wind data but have been filtered and transformed to give one observation per storm event and to have the marginal wave directional effects removed. This leaves 1680 storm events where the event is extreme for at least one of the 150 locations on the grid. A description of the data and pre-processing is given in Randell et al. (2016) with these data representing for Shell Research their test-bed for spatial analysis methods.

Directionality of the waves is found to be present in the spatial dependence structure, so for simplicity we perform our spatial inference on a directional transect through the grid, reducing the field to approximately 1 dimension. The transect used is orientated east–west in the centre of the grid and consists of 7 sites; this is highlighted in Fig. 5. The use of transects for this ocean basin is similar to that as used in Ross et al. (2017), though max-stable processes are fitted in that case.

4.1.2. Methods

We apply the multivariate conditional extremes model of Section 3.1 to identify the potential structure for the spatial functions $\alpha(h)$, $\beta(h)$, $\mu(h)$ and $\sigma(h)$. For illustrative purposes, we only condition on the west-most site in this transect and then fit the model to the other locations in the transect. This is not necessary, however, and more information can be extracted by suitably combining the different conditional distributions. Similar studies using other transects are expected to give weaker levels of extremal dependence as our selected transect direction aligns with most major storm tracks.

To obtain estimates for the model, some assumptions are made for the form of $G_{2:7}$ in limit (3.1). Specifically, to correspond to the Gaussian process formulation in Section 3.2, we take $G_{2:7}$ to be the distribution function of a multivariate normal with mean and standard deviation vectors $(\mu_2, \ldots, \mu_7)$ and $(\sigma_2, \ldots, \sigma_7)$ and with correlation function at lag h taken to be ρ^h. This model is fitted jointly over

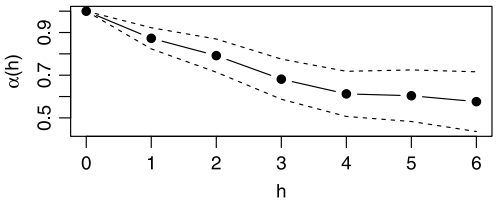


Fig. 6. Pointwise estimates of $\alpha(h)$ from the multivariate conditional extremes fit, conditioned on the west-most location in the transect. Lag $h = 0$ corresponds to the conditioning site, with $h = 6$ being the parameter estimate at the most easterly site. Estimates are for integer values of h and these are shown to be linearly interpolated to show we know that the function is continuous. The dotted lines show 95% confidence intervals for the pointwise estimates.

sites, with a multivariate normal likelihood, unlike in all previous applications of Heffernan and Tawn (2004) which use pseudo-likelihood in Section 3.1.1. For each fitted parameter θ, we set $\theta(i) = \theta_{i+1}$ for $i = 1, \ldots, 6$ so that, for example, $\alpha(1) = \alpha_2$.

In fitting the conditional extremes model, the 0.8 quantile of $X(s)$ has been selected as the conditioning threshold u. This value was chosen for u as this seemed to satisfy the required approximate independence property of limit (3.1) both for that level and that it holds for all higher threshold choices. In practice, the threshold choice is a compromise between being sufficiently low to utilise enough data whilst being suitably high so that the asymptotic argument in (3.1) provides a good approximation.

4.1.3. Results

Exploratory analysis using the model described in Section 4.1.2 showed that there was no evidence for $\beta(h)$ to vary with $h > 0$, and so we take $\beta(h) = \beta_c$, where $0 \leq \beta_c < 1$ is some constant, for $h > 0$; our estimated model gives $\hat{\beta}_c = 0.17$. Also, we found $\hat{\rho} = 0.9$. The corresponding $\alpha(i), i = 1, \ldots, 6$, estimates are shown in Fig. 6, with the values presented here as pointwise estimates of the function $\alpha(h)$. The estimates are consistent with the physical characteristics that may be expected from extreme waves. For $0 \leq h < h_{AD}$ such that $\alpha(h) = 1$ the process is asymptotically dependent, then it would be anticipated that a nearby location is likely to experience an extreme wave of the same order of magnitude if the conditioning site has observed such an event. We see that if this holds then $0 \leq h_{AD} < 1$ based on the 95% confidence intervals for the pointwise estimates. We also see that the degree of dependence is estimated to decrease as the distance between sites increases, which is physically realistic. The decay of the pointwise estimates for $\alpha(h)$, for $h > h_{AD}$, seems smooth and the analysis suggests a simple parametric form for $\alpha(h)$ of the form

$$\alpha(h) = \begin{cases} 1 & \text{if } h < h_{AD} \\ \exp\{-\gamma(h - h_{AD})\} & \text{if } h \geq h_{AD}. \end{cases}$$

Previous spatial modelling of significant wave heights has utilised models of max-stable processes, see Section 2.2. However, these are asymptotically dependent, i.e., $\alpha(h) = 1$ for all h. We can see from Fig. 6 that this is not a good model for $h \geq 1$ for these wave data.

Next, consider the estimated mean and standard deviation functions of the limit process $W(\cdot)$. Pointwise estimates for $\mu(h)$ and $\sigma(h)$ are given in Fig. 7. Both functions behave very similarly; as the distance between the two sites increases, the limit process increases in mean and standard deviation but with decreasing rate for larger distances. This form of $\sigma(h)$ is as expected since the unpredicted variability is likely to increase as the extremal dependence weakens, but the former is a feature that justifies investigation in future research to understand why this property arises. On this initial analysis, however, it appears that $\mu(h) \propto \sigma(h)$ would form a good spatial model.

To assess whether the estimates of $\alpha(h)$ and $\beta(h)$ are reasonable, we simulate using our fitted model realisation of $\{X(s), X(s + h)\}$, for $h = 1, 3, 6$, where $X(s)$ is the standardised (to Gumbel margins) wave height at the most westerly site of the transect and is above the modelling threshold u. The observed data (black), 1680 points from these joint distributions with Gumbel margins, together with

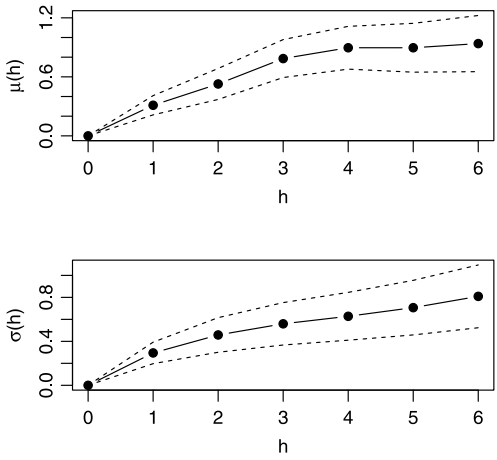


Fig. 7. Pointwise estimates of $\mu(h)$ and $\sigma(h)$ with properties shown identical to Fig. 6.

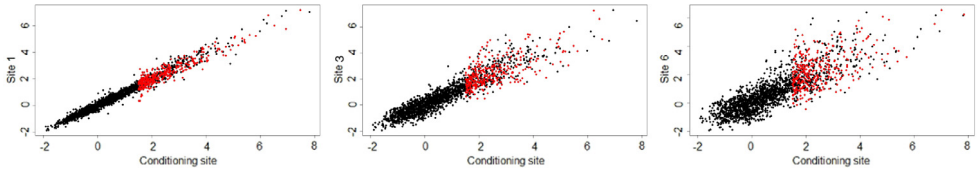


Fig. 8. Simulations from the fitted multivariate conditional extremes model; black points are the data on Gumbel margins, whilst red points are simulated data from the fitted model: left, centre and right panels plot shows these data when $h = 1, 3, 6$ respectively. In each case, the x-axis is the standardised wave height at the conditioning site (the most westerly in the transect), with the y-axis being the standardised wave height at the other site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

336 simulated points with $X(s) > u$ (red) are shown in Fig. 8. It appears from these simulations that the fitted model provides a reasonable fit to the data; for each pair of sites, the distribution of extreme wave data appears to have been captured well. Hence, the model appears to be appropriate for modelling significant wave height in the North Sea on this particular transect. More work will be undertaken to establish if this is the case for further transects in this ocean basin and also to determine how to pool information across transects to estimate the functions $\alpha(\cdot)$, $\beta(\cdot)$, $\mu(\cdot)$, $\sigma(\cdot)$ that change smoothly over distance or separation depending on whether isotropy is found to hold for extreme wave events.

4.2. Understanding widespread flood risk

4.2.1. Background

Understanding flood risk is an important issue for insurance companies, the government, as well as local communities. Previous events have shown that flood events can affect large spatial areas and have devastating impacts on transport and infrastructure (Shaw et al., 2010). Therefore, it is of paramount interest to understand the features of these events and plan future defences to be able to withstand physically plausible events that we have not yet observed.

Flooding is a continuous spatial process but restricted to the river network; however as is common with environmental problems we only have access to observations at a finite number of locations. Therefore, we want to be able to make predictions from these pointwise locations that are consistent with the underlying spatial process (Davison et al., 2012). Furthermore, the dependence structure of

measurements of river flow is highly complex; this is because river flow gauges considered spatially distant through standard metrics such as Euclidean distance can in fact be similar because they lie within the same catchment (Asadi et al., 2015; Shaw et al., 2010). Previous studies such as Asadi et al. (2015) have used the max-stable processes (see Section 2.2), however this approach does not suit large scale studies. Other approaches such as Keef et al. (2009), Lamb et al. (2010) and Towe et al. (2016) have adopted the conditional multivariate extremes model stated in Section 3.1 to understand widespread flood risk.

4.2.2. National flood resilience review

During winter 2015, consecutive storms Desmond, Eva and Frank hit the UK causing widespread flooding across large regions of northern England. These storms required significant responses from the emergency services and in some cases the army to help with the protection of property as well as infrastructure (Lamb et al., 2015). Due to the unprecedented effect of these storms and often the rapid response required, the UK government set up the National Flood Resilience Review (NFRR). The aim of the NFRR was to gain a better understanding of the drivers of flooding in the UK as well as the current methods to deal with the associated risks and damages caused by flooding (Government, 2016).

In particular, the scientific advisory group of the NFRR wanted to understand more about the likelihood of flooding in the UK and move towards thinking about risks at a national scale rather than location by location. To better understand the risk of widespread flooding, a comprehensive analysis of UK river flow gauges was required. As we are interested in understanding the characteristics of widespread flooding in the UK, a flexible spatial extreme value model that is able to accommodate the known features within the data is required. For example, this needs to model that flood events can be both localised as well as national and not all sites are likely to be extreme concurrently. The Heffernan and Tawn (2004) conditional multivariate extreme value model, stated in Section 3, satisfies both of these modelling requirements.

4.2.3. Methods

Observations of river flow gauges were obtained from the National River Flow Archive maintained by the Centre of Ecology and Hydrology, as well as from Environment Agency records. Before any statistical modelling was undertaken, a quality assurance of the data was performed. This quality assurance required the data to have at least 20 years of observations with a relatively small percentage of missing values, this requirement enabled robust estimation of the parameters of the associated statistical models (see Sections 2.1 and 3). Furthermore, gauges were removed from the analysis if unnatural changes in the time series were observed, for example if a dam was installed further upstream. This results in unnatural changes of the time series at downstream gauge being present in the time series (Shaw et al., 2010). This quality assurance process resulted in 916 suitable gauging records. To maintain consistency with previous studies of UK flooding, an event was defined to last for a period of time of up to 7 days (Keef et al., 2009). The statistical analysis includes extensions to the Heffernan and Tawn methodology as stated in Section 3.1 such as the handling of missing values as well as efficient simulation techniques for high dimensional data sets and methods to model the rate of the number of extreme events per year (Keef et al., 2013). These aspects are key when modelling spatial river flow data sets with more details of these methods found in Keef et al. (2013). In order to assess the validity of the statistical models, comparisons such as those shown in Fig. 8 were made. From the statistical analysis, 10 000 years worth of events were simulated in Gumbel margins, we denote these by

$$\{\tilde{X}_t(s_i);\ i = 1, \ldots, 916,\ t = 1, \ldots, 10000 n_y\}, \tag{4.1}$$

where n_y is the average number of events in the region per year. This simulated event set includes events that are larger than those observed in the data for at least one site but with the dependence structure of these events being consistent with the features from the observed extreme events (Keef et al., 2009). This simulated event set then allows us to estimate a number of summary statistics for a range of severities of events to help us characterise the behaviour of flooding across the UK.

4.2.4. Conditional probability calculation

In order to test the validity of simulations from the conditional extreme value model, we compare the calculation of conditional probabilities from both the observed and simulated data sets. For all return levels, the non-limit conditional probability in Eq. (1.4) is calculated relative to a conditioning gauge, which in this case is situated on the river Severn. For the empirical data, the conditional probability was calculated relative to the 99th percentile (approximately a 5 month level) as well as to a level equivalent to the one year return level, the estimates of this can be seen in Fig. 9(a) and (b) respectively. For the Heffernan and Tawn (2004) model estimates of the conditional probability the empirical conditional probability from the simulated data set was evaluated for both a 10 and 100 year return level, see Fig. 9(c) and (d) respectively.

In both cases, the strongest dependence is seen with nearby gauges as well as those that lie within the river Severn catchment. However, the spatial dependence is not stationary, as distant gauges can still have strong extremal dependence, which is larger than those gauges nearby. This feature is due to the similarity of their catchments with the catchment of the conditioning gauge. Focusing on Fig. 9(a) and (b), when we consider higher levels the conditional probability decreases, this suggests that as events become more severe, they are also becoming more localised. Higher conditional probabilities from the observed data sets cannot be considered as there is insufficient data to produce stable estimates. This decaying conditional probability characteristic though is also observed for the higher levels considered in Fig. 9(c) and (d), which show our model-based estimates. There is also a smooth transition in Fig. 9 between the estimates of the conditional probabilities from the observed and simulated data sets.

If the statistical model had assumed asymptotic dependence between river flow gauges, the conditional probabilities shown in Fig. 9 would be estimated as invariant to conditioning return level. Therefore, if the 99% quantile was used to fit the model, comparing Fig. 9(a) and (d) shows that this leads to an error in spatial extremal dependence estimation. In this particular case, there would be massive over-estimation of the spatial extremal dependence between river flow gauges. These comparisons confirm that the conditional extreme value model of Heffernan and Tawn (2004) is accurately capturing the extremal dependence observed in spatially extreme river flows.

4.2.5. Scenario evaluation for the national flood resilience review

The analysis of the observed and simulated data sets in Section 4.2.4 confirmed that the features of the observed data set are being captured in the models represented by the simulated event set. As a result, we are able to use the simulated event set as a proxy for a long observational record to answer fundamental questions for flood risk management posed by the NFRR such as:

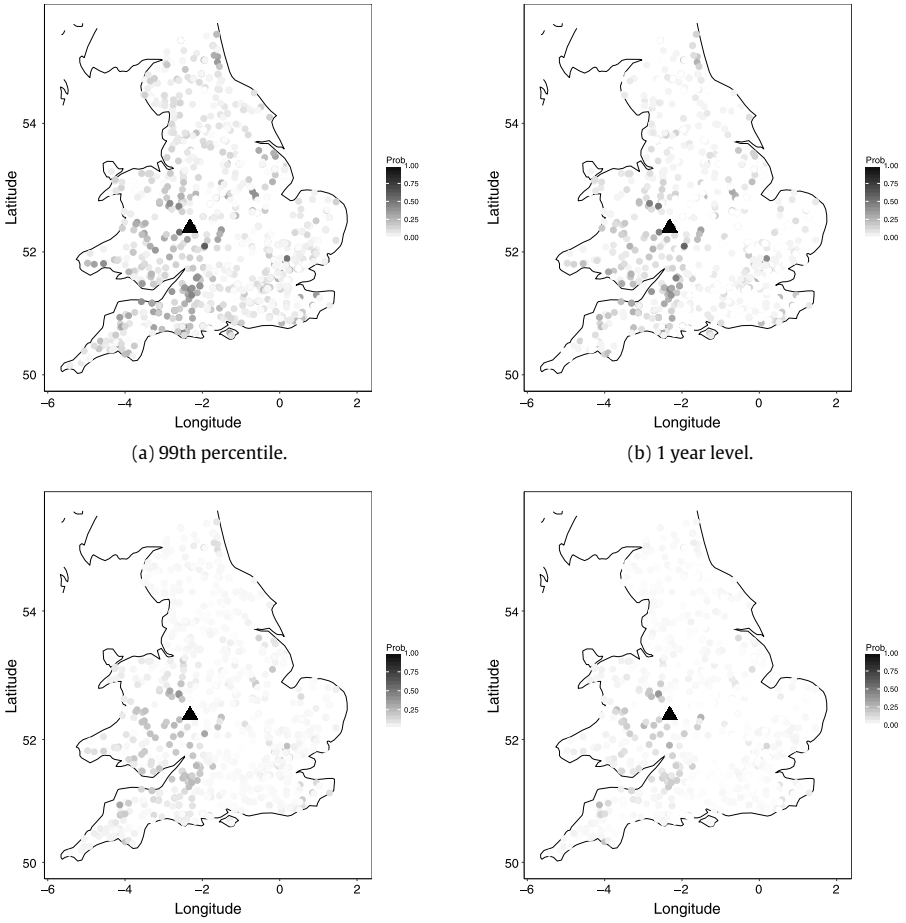
> *What is the chance of an extreme river flow occurring at one or more gauges across England and Wales, somewhere within the national river gauge network in any one year?*

To frame this question in terms of our notation, we need, for an arbitrary year t, to estimate $1 - \mathbb{P}\left(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \ldots, 916\right)$, where $M_{Y(s),t}$ is the annual maximum in year t for the river flow in site s and $y_{s,T}$ is the T year return level at site s. This probability is identical to $1 - \mathbb{P}\left(M_{X(s_i),t} < x_T; i = 1, \ldots, 916\right)$, where x_T is the T year return level on Gumbel margins. We estimate the second term in this probability using the simulated sample (4.1) as

$$\hat{\mathbb{P}}\left(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \ldots, 916\right) = \frac{1}{k}\sum_{j=1}^{k} \mathbb{1}\left(\max_{i=1,\ldots,916} M_{\tilde{X}(s_i),j} < x_T\right), \tag{4.2}$$

where $k = 10000n_y$ and $\mathbb{1}(A)$ is the indicator function of event A.

The estimates of $1 - \mathbb{P}\left(M_{Y(s_i),t} < y_{s_i,T}; i = 1, \ldots, 916\right)$ are shown as T varies in Fig. 10 using the modelled dependence with estimator (4.2) and under the two limiting cases that assume all of the 916 gauges are either completely independent or completely dependent. Here the complete independence case assumes that there is no association between when flooding occurs at each of the 916 gauges, whereas the complete dependence assumes that each of the 916 gauges behave identically. The benefit of the conditional extremes approach is that we are able to estimate the probability whatever T, i.e., even for events with return periods that are greater than the severity of the events captured



(a) 99th percentile.

(b) 1 year level.

(c) 10 year level.

(d) 100 year level.

Fig. 9. Comparisons of the non-limit conditional probability (1.4) for (a) the 99th percentile and (b) the one year return period from the observed data set; (c) and (d) show this conditional probability estimated using our model for the 10 and 100 year return periods respectively. The triangle symbol represents the conditioning gauge for the estimate, this gauge is situated in the river Severn catchment.

in the observed data set. For the NFRR, the key feature of this analysis was that the probability of observing a 1 in 100 year event at any of the 916 gauging stations in any given year is 0.78, so its very likely a 100 year event occurs somewhere in this region.

This analysis considered only those locations where there are gauges with river flow measurements; current research is addressing how this question can be answered for every place along the river network, i.e., to estimate $1 - \mathbb{P}\left(M_{Y(s),t} < y_{s,T};\ \text{for } s \in \mathcal{S}\right)$. It should be also noted that our study focusses on England and Wales, reflecting the scope of the NFRR (flood risk management is a devolved matter in the United Kingdom, with separate arrangements in place in Scotland).

> *What is the chance of an extreme river flow occurring in one or more Local Resilience Forums, somewhere within the national river gauge network in any one year?*

The analysis shown in Fig. 10 considered the probability of observing a flood event at any gauge across the river network. However, for emergency planning purposes, interest lies in determining the

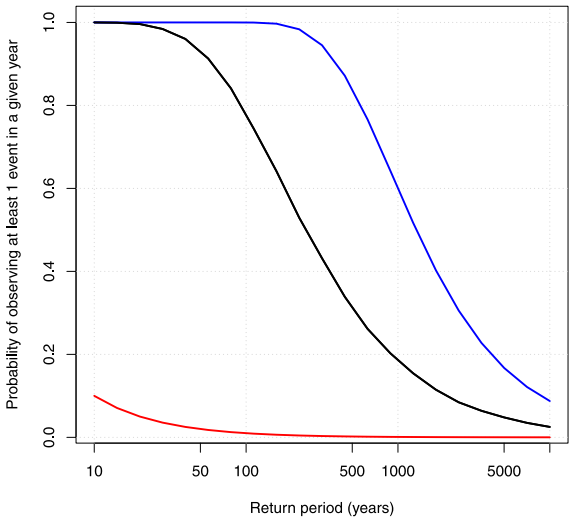


Fig. 10. Comparison of the three dependence models used to estimate probability of observing at least one event above a T-year return period for a given year: our model for the dependence (black), under a complete independence model (blue) and under complete dependence (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spatial extent of potential events. Within the England and Wales, responses to natural hazards are managed through 42 Local Resilience Forums (LRFs), which we denote by $\{L_p;\ p = 1, \ldots, R = 42\}$. Therefore, it seems natural to define events in accordance to the number of LRFs that receive a T year event at some gauge. Let $M_{X,t}(L_p) = \max_{i \in L_p} M_{X(s_i),t}$, i.e., it is the maximum level, on Gumbel scale, over the pth LRF and let $M_{X,t}(L_{(r)})$ be the r largest value of $M_{X,t}(L_p)$, $p = 1, \ldots, R$ in year t, so $M_{X,t}(L_{(1)}) > \cdots > M_{X,t}(L_{(R)})$. To understand the regional extent of spatial flood events, we are interested in whether in an arbitrary year t, at least r LRFs have exceedances of the marginal T return level, i.e., the $\{M_{X,t}(L_{(r)}) > x_T\}$. We estimate this probability using the simulated sample (4.1) by

$$\hat{\mathbb{P}}\left(M_{X,t}(L_{(r)}) > x_T\right) = \frac{1}{k} \sum_{j=1}^{k} \mathbb{1}\left\{M_{\tilde{X},j}(L_{(r)}) > x_T\right\},$$

where $k = 10000 n_y$.

Estimates of the probability for $r = 1, \ldots, 4$ are shown in Fig. 11. As expected the estimates for at least $r = 1$ region being above a T-year return period in any given year is consistent with the analysis shown in Fig. 10. Most interesting is that in any given year there is 0.35 probability of at least a 1 in 100 year event occurring in at least four LRFs.

Both of the questions proposed by the NFRR highlighted that flooding is more common than one might expect. The typical communication of return period is a single site measure. The conditional spatial extreme value model of Heffernan and Tawn (2004) allows us to provide robust answers to these national scale questions through carefully capturing the complex dependence structure of a high dimensional set of river flow gauges. The uncertainty around the estimates of the conditional probability as well as the point estimates shown in Figs. 10 and 11 from the NFRR can easily be assessed by bootstrap methods.

The questions proposed by the NFRR were answered by modelling the spatial dependence of gauges on the river network. However, ultimate interest lies in estimating the chance of observing a flood in a given year at any location along the river network. Answering this question is an ongoing research question, which involves exploiting information about the river network as well as modelling the joint dependence of river flow with that of the process of rainfall.

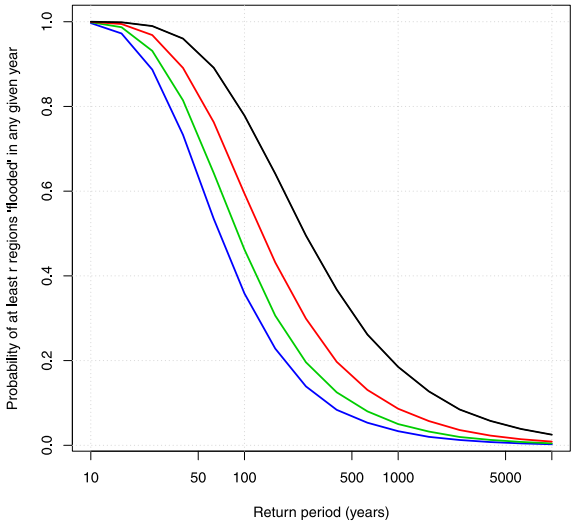


Fig. 11. Estimated probability of observing at least r LRF regions above a T-year return period in any given year. The black, red, green and blue curves show the cases for when $r = 1, 2, 3, 4$ respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgements

We would like to thank Jennifer Wadsworth for her help in producing the figures shown in Section 2. R. Towe's research was supported by Jeremy Benn Associates Ltd. and Innovate UK KTP009454 and EP/P002285/1 (The Role of Digital Technology in Understanding, Mitigating and Adapting to Environmental Change). R. Shooter's research was supported by EP/L015692/1 (STOR-i Centre for Doctoral Training) and Shell Research. The statistical analysis of these 916 river flow gauges stated in Section 4 was conducted by using JBA Risk Management's Stochastic Hazard Event Estimation Programme tool JSheep. We thank Shell Research for providing the North Sea significant wave height data and for helpful discussions.

References

Asadi, P., Davison, A., Engelke, S., 2015. Extremes on river networks. Ann. Appl. Stat. 9, 2023–2050.
Bacro, J., Gaetan, C., Toulemonde, G., 2016. A flexible dependence model for spatial extremes. J. Statist. Plann. Inference 172, 36–52.
Brown, B.M., Resnick, S.I., 1977. Extreme values of independent stochastic processes. J. Appl. Probab. 14, 732–739.
Coles, S.G., Heffernan, J.E., Tawn, J.A., 1999. Dependence measures for extreme value analyses. Extremes 2, 339–365.
Coles, S.G., Tawn, J.A., 1991. Modelling extreme multivariate events. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 377–392.
Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modeling of spatial extremes. Statist. Sci. 27, 161–186.
Davison, A., Smith, R., 1990. Models for exceedances over high thresholds (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 52, 393–442.
de Haan, L., 1984. A spectral representation for max-stable processes. Ann. Probab. 12, 1194–1204.
Dombry, C., Ribatet, M., 2015. Functional regular variations, Pareto processes and peaks over threshold. Statist. Interface 8 (1), 9–17.
Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M., 2015. Estimation of Hüsler-Reiss distributions and Brown–Resnick processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (1), 239–265.
Ferreira, A., de Haan, L., 2014. The generalized Pareto process; with a view towards application and simulation. Bernoulli 20, 1717–1737.
Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math. Proc. Camb. Phil. Soc. 24, 180–190.
Genton, M., Ma, Y., Sang, H., 2011. On the likelihood function of Gaussian max-stable processes. Biometrika 98, 481–488.
Government, H.M., 08 September 2016. National Flood Resilience Review, Tech. rep.

Heffernan, J.E., Resnick, S.I., 2007. Limit laws for random vectors with an extreme component. Ann. Appl. Probab. 17, 537–571.
Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 66, 497–546.
Jonathan, P., Ewans, K., Randell, D., 2013. Joint modelling of extreme ocean environments incorporating covariate effects. Coast. Eng. 79, 22–31.
Keef, C., Papastathopoulos, I., Tawn, J.A., 2013. Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. J. Multivariate Anal. 115, 396–404.
Keef, C., Svensson, C., Tawn, J.A., 2009. Spatial dependence in extreme river flows and precipitation for Great Britain. J. Hydrol. 378, 240–252.
Lamb, R., Keef, C., Tawn, J.A., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., Batstone, C., 2010. A new method to assess the risk of local and widespread flooding on rivers and coasts. J. Flood Risk Manag. 3, 323–336.
Lamb, R., Szönyi, M., May, P., 2015. Flooding after Storm Desmond, JBA Trust.
Liu, Y., Tawn, J., 2014. Self-consistent estimation of conditional multivariate extreme value distributions. J. Multivariate Anal. 127, 19–35.
Lugrin, T., Davison, A.C., Tawn, J.A., 2016. Bayesian uncertainty management in temporal dependence of extremes. Extremes 19, 491–515.
Nelsen, R.B., 2006. An Introduction to Copulas, second ed.. In: Springer Series in Statistics, Springer, New York.
Padoan, S.A., Ribatet, M., Sisson, S.A., 2010. Likelihood-based inference for max-stable processes. J. Amer. Statist. Assoc. 105, 263–277.
Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.
Randell, D., Feld, G., Ewans, K., Jonathan, P., 2015. Distributions of return values for ocean wave characteristics in the South China Sea using directional-seasonal extreme value analysis. Environmetrics 26, 442–450.
Randell, D., Turnbull, K., Ewans, K., Jonathan, P., 2016. Bayesian inference for nonstationary marginal extremes. Environmetrics 27, 439–450.
Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes. In: Applied Probability, vol. 4, Springer-Verlag, New York.
Resnick, S.I., 2013. Extreme Values, Regular Variation and Point Processes. Springer.
Ross, E., Kereszturi, M., van Nee, M., Randell, D., Jonathan, P., 2017. On the spatial dependence of extreme ocean storm seas. Ocean Eng. 145, 359–372.
Schlather, M., 2002. Models for stationary max-stable random fields. Extremes 5, 33–44.
Shaw, E.M., Beven, K.J., Chappell, N.A., Lamb, R., 2010. Hydrology in Practice. CRC Press.
Smith, R.L., (1990) 1990. Max-stable processes and spatial extremes, Unpublished manuscript.
Towe, R.P., Eastoe, E.F., Tawn, J.A., Jonathan, P., 2017. Statistical downscaling for future extreme wave heights in the North Sea. Ann. Appl. Stat. 11 (4), 2375–2403.
Towe, R.P., Tawn, J.A., Lamb, R., Sherlock, C.G., Liu, Y., 2016. Improving statistical models for flood risk assessment. In: E3S Web of Conferences, Vol. 7, EDP Sciences, p. 01011.
Wadsworth, J.L., Tawn, J.A., 2012. Dependence modelling for spatial extremes. Biometrika 101 (1), 1–15.
Winter, H.C., Tawn, J.A., 2016. Modelling heatwaves in central France: a case-study in extremal dependence. J. R. Stat. Soc. Ser. C. Appl. Stat. 65, 345–365.
Winter, H.C., Tawn, J.A., 2017. kth-order Markov extremal models for assessing heatwave risks. Extremes 20, 393–415.
Winter, H.C., Tawn, J.A., Brown, S.J., 2016. Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia. Ann. Appl. Stat. 10, 2075–2101.