Ocean Engineering 145 (2017) 359–372

Contents lists available at ScienceDirect

Ocean Engineering

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

On the spatial dependence of extreme ocean storm seas

Emma Ross [a], Monika Kereszturi [b], Mirrelijn van Nee [c], David Randell [d], Philip Jonathan [a,*]

[a] Shell Projects and Technology, London SE1 7NA, United Kingdom
[b] Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YW, United Kingdom
[c] Faculty of Electrical Engineering, Mathematics and Computer Science, Technical University of Delft, 2628 CD Delft, The Netherlands
[d] Shell Projects and Technology, 1031 HW Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Extreme
Spatial
Dependence
Max-stable process
Composite likelihood
Pooling
North Sea

ABSTRACT

Contemporaneous occurrences of extreme seas at multiple locations in a neighbourhood can cause greater structural reliability and human safety concerns than extremes at a single location. Understanding spatial dependence of extreme seas is important therefore in metocean design, yet has received little rigorous attention in the offshore engineering literature. We characterise the spatial dependence of storm peak significant wave height using three models motivated by max-stable processes for locations in the northern North Sea. Models for marginal extremes per location, and dependence of extremes between locations, are estimated using Bayesian inference with composite spatial likelihoods. We show that, in addition to marginal directional non-stationarity of extreme seas per location, all three models indicate spatial anisotropy in extremal dependence quantified by the spatial covariance matrix of the corresponding max-stable process. Estimates suggest that extreme seas show greater extremal dependence from West to East than from North to South.

1. Introduction

Extreme value analysis is a framework to characterise and quantify extreme phenomena. Using extreme value analysis, we estimate the marginal tail distribution of a single random variable, or the joint tail distribution of two or more random variables. Compared with marginal analysis, multivariate extreme value analysis is more challenging, less developed theoretically, and less used in practice.

One approach to multivariate extreme value analysis of oceanographic and engineering interest uses spatial processes to describe the behaviour of spatially-distributed extremes. Consider significant wave height (H_S) from wind-driven sea states over a spatial lattice of locations for intervals of time corresponding to storm events. We observe maxima of H_S per location per storm, referred to as storm peaks, and assume these to be independent in time. We are interested in characterising the spatial distribution of storm peak H_S. If the lattice consists of locations $1, 2, \ldots, p$, and the continuous random variables and values observed are respectively X_j, x_j, for $j = 1, 2, \ldots, p$, then the joint spatial density of storm peak H_S might be written $f(x_1, x_2, \ldots, x_p) = \partial^p F / \partial x_1 \partial x_2 \ldots \partial x_p$ evaluated at $(x_1, x_2, \ldots, x_p)$, where $F(x_1, x_2, \ldots, x_p) = \Pr[X_1 \le x_1, X_2 \le x_2, \ldots, X_p \le x_p]$ is the joint cumulative distribution function. Assume now that H_S achieves a large maximum value (exceeding some threshold u_k) at some location k on the lattice. Then the conditional density

$$f(x_1, x_2, \ldots, x_p | X_k = x_k > u_k)$$

describes the "spatial shape" of dependence for a typical extreme storm. We might expect that spatial shape is dependent on characteristics of the environment: location (fetch, water depth, bathymetry) and wind field (central pressure, speed, direction, gradients, wind field spatial extent). We know from the literature (e.g. Mendez et al., 2008; Sartini et al., 2015) that marginal extreme value characteristics of H_S are non-stationary with respect to covariates such as wave direction and season. We might surmise therefore that storm shape varies with storm direction and season, as well as varying between ocean basins. When the data sample is sufficiently large, statistical models whose parameters are functions of covariates are generally necessary. In practical application, there may be insufficient evidence in the sample to identify and hence justify incorporating non-stationarity, especially for the spatial dependence structure. Ascertaining whether a typical North Sea hindcast sample shows evidence for non-stationarity of spatial dependence structure is the main objective of this work.

* Corresponding author.
E-mail address: philip.jonathan@shell.com (P. Jonathan).

http://dx.doi.org/10.1016/j.oceaneng.2017.08.051
Received 10 April 2017; Received in revised form 20 June 2017; Accepted 24 August 2017
Available online 26 September 2017

We can write the joint density $f(x_1, x_2, \ldots, x_p)$ as the product of marginal densities $f(x_1), f(x_2), \ldots, f(x_p)$ and a dependence function $\mathcal{C}(x_1, x_2, \ldots, x_p)$

$$f(x_1, x_2, \ldots, x_p) = [f(x_1)f(x_2)\ldots f(x_p)]\mathcal{C}(x_1, x_2, \ldots, x_p).$$

Estimating the marginal densities is familiar territory: for each k, the tail $X_k > u_k$ can be estimated using a marginal extreme value model, and the density below u_k estimated empirically. Estimating the dependence function $\mathcal{C}$ is more problematic, especially when the values of one or more variables is extreme: spatial extreme value methods are needed. Note that spatial dependence is characterised by $\mathcal{C}$ only: if we transform each variable X_j to Y_j, such that $\{Y_j\}$ has a common marginal distribution, the corresponding transformed dependence function $\mathcal{C}_Y(y_1, y_2, \ldots, y_p)$ is a copula function. There is a large number of copula functions available to describe multivariate distributions, but only so-called extreme value copulas are appropriate to describe multivariate extreme value distributions (see Section 3).

From an engineering perspective, improved quantification of spatial dependence of extreme storms would allow better estimation of uncertainties in return values from analysis of spatially-pooled data; this is particularly relevant for ocean basins where storm events (hurricanes, tropical cyclones) have relatively low rates of occurrence over a spatial neighbourhood. The work of Heideman and Mitchell (2009) provides a valuable introduction to approaches used by practising metocean engineers for hurricane-type applications, including site averaging, grid point pooling and track shifting. The usual approach to uncertainty quantification of return values from spatially-pooled data is to assume (wrongly) that data from different locations are mutually independent; this is called the "independence likelihood" assumption (e.g. Chandler and Bate, 2007) leading to a joint density

$$f(x_1, x_2, \ldots, x_p) \approx f(x_1)f(x_2)\ldots f(x_p),$$

which ignores dependence function $\mathcal{C}$. We effectively assume that there are more independent observations than is actually the case, and therefore underestimate uncertainties during maximum likelihood estimation. To correct this, we then need to inflate uncertainty bands using a spatial "block bootstrapping" scheme (e.g. Chavez-Demoulin and Davison, 2005). Adopting a sample likelihood which more adequately represents extremal dependence would avoid the need to make the independence likelihood assumption. An approximation for the sample likelihood with captures spatial dependence of extremes is essential if Bayesian inference is to be used, since bootstrapping makes little sense in a Bayesian context. A better description of spatial shape would also improve our ability to quantify the consequences of extreme seas impacting multiple locations at the same time.

Max-stable process (MSP) models (following from the work of de Haan and Resnick, 1977) represent the most reasonable statistical approach currently available to inference for spatial extremes. MSPs can be thought of as extensions of multivariate extreme value distributions to continuous space, as summarised in Section 3. Their finite p-dimensional distributions provide a description of dependence function $\mathcal{C}$ above for a lattice of locations. Ribatet (2013) provides a review of MSPs, outlining the so-called Smith, (1990), Schlather, (2002) and Brown-Resnick (Brown and Resnick, 1977) models which have found some application in the environmental science literature. These models are described further in Section 3.2 and the Appendix, and will be applied in Section 5. Ribatet (2013) also provides an overview of methods for simulating from MSPs. Since the multivariate likelihood characterising extremal dependence cannot be written in closed form except in the bivariate case, approximate likelihoods are necessary for inference using maximum likelihood estimation. Padoan et al. (2010) presents a composite likelihood-based approximation for fitting MSPs, evaluates its performance using simulated data, and applies it to spatial extremes of daily precipitation. The composite likelihood framework used in this work is outlined in Section 4.

As MSPs arise as the limit distribution of componentwise maxima, they should be applied to samples of (contemporaneous) maxima (per location per time interval) for a lattice of spatial locations. However, extreme value inference using temporal peaks over threshold exploits the sample more efficiently. Following the ideas of Smith et al. (1997), Huser and Davison (2014) presents an approximate censored likelihood scheme for spatial modelling of peaks over threshold, used in this work also, as outlined in Section 4.

Different forms of extremal dependence exist, as outlined by Eastoe et al. (2013). Models based on consideration of componentwise maxima typically assume a particular form of extremal dependence, known as asymptotic dependence, as discussed e.g. by Kereszturi et al. (2016) for a sample of significant wave heights similar to that used in this work. This amounts to the assumption $\lim_{y\to\infty} \Pr[Y_l > y | Y_k > y] > 0$ at all pairs k, l of locations with Y_k, Y_l on common marginal scale (unless Y_k, Y_l are perfectly independent in which case the limit is 0). However, it is usually very difficult if not impossible to identify the form of extremal dependence present in a typical sample of limited size, with covariate effects also in play. Kereszturi et al. (2016) seek to refine the way diagnostics for extremal dependence are used in practice to improve their interpretability. For relatively large samples of sea state H_S, they find some evidence for asymptotic independence of extreme values between locations; for smaller samples of storm peak H_S, diagnostics are inconclusive. For this reason, we choose here to report applications of asymptotic dependent models (Section 5), whilst referring to related work (Kereszturi, 2016) for models exhibiting asymptotic independence yielding very similar results.

Estimation of spatial extremes models for samples of non-stationary peaks over threshold is complicated by at least three effects. Firstly, spatial extremes models are usually defined for block maxima not peaks over threshold; yet statistical inference is more efficient using peaks over threshold, and inference using peaks over threshold is commonplace in ocean engineering. Fortunately, for exceedances of a high threshold, likelihoods for block maxima and peaks over threshold can be shown to be approximately equal, so that the spatial extremes model can also be applied to peaks over threshold. However, we also need to model whole samples; we therefore apply a censored likelihood argument to construct approximate whole sample likelihoods for spatial extremes of peaks over threshold. Secondly, full joint distributions for spatial extremes are not available in closed form, but expressions for bivariate cumulative distribution functions and densities typically are. Using these, we need to construct composite likelihood approximations to the full likelihood for parameter estimation. Thirdly, spatial extremes models are defined assuming common standard Fréchet marginal distributions. In reality, marginal distributions are non-stationary with respect to covariates. We therefore need to fit non-stationary marginal models and transform marginally to standard Fréchet scale under these models. The spatial extremes model may of course also be sensitive to covariates.

The MSP is by far the most popular approach in the statistics literature to characterise extreme spatial processes. However, the conditional extremes model of Heffernan and Tawn (2004) provides an alternative modelling framework motivated by extreme value theory, advantageous in that it admits both asymptotic dependence and asymptotic independence within the same model. It also provides a relatively straightforward means for estimation on spatial grids, as illustrated e.g. by Eastoe et al. (2013).

1.1. Outline of paper

The motivation for this work is to consider (a) the feasibility and (b) the usefulness of applying multivariate extreme value models to real-world ocean engineering design problems. We assess the extent to which there is evidence, in typical samples of (storm peak) significant wave height where marginal non-stationarity has already been