Multidimensional covariate effects in spatial and joint extremes

P. Jonathan, K. Ewans and D. Randell

Shell Projects & Technology

Corresponding author: philip.jonathan@shell.com

Abstract

Careful modelling of non-stationarity is critical to reliable specification of marine and coastal design criteria. We present a spline based methodology to incorporate spatial, directional, temporal and other covariate effects in extreme value models for environmental variables such as storm severity. For storm peak significant wave height events, the approach uses quantile regression to estimate a suitable extremal threshold, a Poisson process model for the rate of occurrence of threshold exceedances, and a generalised Pareto model for size of threshold exceedances. Multidimensional covariate effects are incorporated at each stage using penalised (tensor products of) B-splines to give smooth model parameter variation as a function of multiple covariates. Optimal smoothing penalties are selected using cross-validation, and model uncertainty is quantified using a bootstrap re-sampling procedure. The method is applied to estimate return values for large spatial neighbourhoods of locations, incorporating spatial and directional effects. Extensions to joint modelling of multivariate extremes, incorporating extremal spatial dependence (using max-stable processes) or more general extremal dependence (using the conditional extremes approach) are outlined.

1   Introduction

Availability of comprehensive met–ocean data allows the effect of heterogeneity (or non-stationarity) of extremes with respect to direction, season and location to be accommodated in estimation of design criteria. Jonathan and Ewans [2013] review statistical modelling of extremes for marine design.

Capturing covariate effects in extreme sea states is important when developing design criteria. In previous work (e.g Jonathan and Ewans [2007a], Ewans and Jonathan [2008]) it has been shown that omni-directional design criteria derived from a model that adequately incorporates directional covariate effects can be materially different from a model which ignores those effects(e.g. Jonathan et al. 2008). Directional storm peaks H_{S100} derived from a directional model can be heavier tailed than that derived from a direction-independent approach, indicating that large values of storm peak H_S are more likely than we might anticipate were we to base our beliefs on estimates which ignore directionality. Similar effects have been demonstrated for seasonal covariates (e.g.Anderson et al. 2001, Jonathan et al. 2008).

There is a large body of statistical literature regarding modelling of covariate effects in extreme value analysis; for example, Davison and Smith [1990] or Robinson and Tawn [1997]. The case for adopting an extreme value model incorporating covariate effects is clear, unless it can be demonstrated statistically that a model ignoring covariate effects is no less appropriate. Chavez-Demoulin and Davison [2005] and Coles [2001] provide straight-forward descriptions of a non-homogeneous Poisson model in which occurrence rates and extremal properties are modelled as functions of covariates. Scotto and Guedes-Soares [2000] describe modelling using non-linear thresholds. A Bayesian approach is adopted Coles and Powell [1996] using data from multiple locations, and by Scotto and Guedes-Soares [2007]. Spatial models for extremes (Coles and Casson [1998], Casson and Coles [1999]) have also been used, as have models (Coles and Tawn [1996, 2005]) for estimation of predictive distributions, which incorporate uncertainties in model parameters. Ledford and Tawn [1997] and Heffernan and Tawn [2004] discuss the modelling of dependent

joint extremes. Chavez-Demoulin and Davison [2005] also describe the application of a block bootstrap approach to estimate parameter uncertainty and the precision of extreme quantile estimates, applicable when dependent data from neighbouring locations are used. Jonathan and Ewans [2007b] use block bootstrapping to evaluate uncertainties associated with extremes in storm peak significant wave heights in the Gulf of Mexico.

Guedes-Soares and Scotto [2001] discuss the estimation of quantile uncertainty. Eastoe [2007] and Eastoe and Tawn [2009] illustrate an approach to removing covariate effects from a sample of extremes prior to model estimation.

One of the first examinations of spatial characteristics of extreme wave heights was reported by Haring and Heideman [1978] for the Gulf of Mexico. They performed extremal analysis of the ODGP hurricane hindcast data base (Ward et al. [1978]) at a number of continental shelf locations from Mexico to Florida, and concluded that there was no practical difference between the sites, but they did observe a gradual reduction in extreme wave heights with decreasing water depth. Chouinard et al. [1997] took the opportunity to re-examine the spatial behaviour of extremes in the Gulf of Mexico, when the GUMSHOE hindcast data base became available. Jonathan and Ewans [2011] used thin–plate splines to model the spatial characteristics of events in the Gulf of Mexico. Extending the thin–plate spline formulism to include other (possibly periodic) covariates is difficult; instead, the sample is typically pre-processed to remove the influence of all covariates other than the (2–D) spatial, prior to model estimation using thin–plate splines. Models estimated in this way suffer from the fact that interactions between the various modelling steps (and the parameters estimated therein) cannot be easily quantified.

Characterising the joint structure of extremes for different environmental variables is also important for improved understanding of those environments. Yet many applications of multivariate extreme value analysis adopt models that assume a particular form of extremal dependence between variables without justification, or restrict attention to regions in which all variables are extreme. The conditional extremes model of Heffernan and Tawn [2004] provides one approach avoiding these particular restrictions. Extremal dependence characteristics of environmental variables also typically vary with covariates. Reliable descriptions of extreme environments should also therefore characterise any non-stationarity. Jonathan et al. [2013] extends the conditional extremes model of Heffernan and Tawn to include covariate effects, using Fourier representations of model parameters for single periodic covariates.

The last decade has seen the emergence of useable statistical models for spatial extremes based on max–stable processes, at least in academia. The application of max–stable processes is complicated due to unavailability of the full multivariate density function. Padoan et al. [2010] develops inferentially practical, likelihood-based methods for fitting max–stable processes derived from a composite likelihood approach. The procedure is sufficiently reliable and versatile to permit the simultaneous modelling of marginal and dependence parameters in the spatial context at a moderate computational cost. Davison and Gholamrezaee [2012] describes an approach to flexible modelling for maxima observed at sites in a spatial domain, based on fitting of max–stable processes derived from underlying Gaussian random process models. Generalised extreme value (GEV) margins as assumed throughout the spatial domain, and models incorporate standard geo–statistical correlation functions. Estimation and fitting are performed through composite likelihood inference applied to observations from pairs of sites. Davison et al. [2012] also provides a good introduction and review. Erhardt and Smith [2011] uses approximate Bayesian computation to circumvent the need for a joint likelihood function by instead relying on simulations from the (unavailable) likelihood avoiding the need to construct composite likelihoods at higher computational cost.

In this work, we apply a marginal model for spatio–directional extremes to a sample of data for storm severity on the north west continental shelf of Western Australia. The model (developed in Section 2) adopts a penalised B-splines formulation to characterise smooth variation of extreme value parameters spatially and directionally. The North West Shelf application is then presented in Section 3. In Section 4, we discuss model extension to incorporate appropriate spatial extremal dependence, and also outline a non–stationary extension of the conditional extremes model of Heffernan and Tawn [2004]).

2   Model

The objective is to estimate design criteria for individual locations within a spatial neighbourhood, accounting for spatial and storm directional variability of extremal characteristics.

Model components

Following the work of Jonathan and Ewans [2008] and Jonathan and Ewans [2011], summarised in Jonathan and Ewans [2013], we model storm peak significant wave height, namely the largest value of significant wave height observed at each location during the period of a storm event. We assume that each storm event is observed at all locations within the neighbourhood under consideration. For a sample $\{\dot{z}_i\}_{i=1}^{\dot{n}}$ of $\dot{n}$ storm peak significant wave heights (henceforth H_S) observed at locations $\{\dot{x}_i, \dot{y}_i\}_{i=1}^{\dot{n}}$ with dominant wave directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$ at storm peak H_S (henceforth "storm directions"), we proceed using the peaks over threshold approach as follows.

Threshold: We first estimate a threshold function ϕ above which observations $\dot{z}$ are assumed to be extreme. The threshold varies smoothly as a function of covariates ($\phi \stackrel{\triangle}{=} \phi(\theta, x, y)$) and is estimated using quantile regression. We retain the set of n threshold exceedances $\{z_i\}_{i=1}^{n}$ observed at locations $\{x_i, y_i\}_{i=1}^{n}$ with storm peak directions $\{\theta_i\}_{i=1}^{n}$ for further modelling.

Rate of occurrence of threshold exceedance: We next estimate the rate of occurrence ρ of threshold exceedance using a Poisson process model with Poisson rate $\rho(\stackrel{\triangle}{=} \rho(\theta, x, y))$.

Size of occurrence of threshold exceedance: We estimate the size of occurrence of threshold exceedance using a generalised Pareto (henceforth GP) model. The GP shape and scale parameters ξ and σ are also assumed to vary smoothly as functions of covariates.

This approach to extreme value modelling follows that of Chavez-Demoulin and Davison [2005] and is equivalent to direct estimation of a non-homogeneous Poisson point process model (e.g., Dixon et al. 1998, Jonathan and Ewans [2013]).

Parameter estimation

For quantile regression, we seek a smooth function ϕ of covariates corresponding to non-exceedance probability τ of H_S for any combination of θ, x, y. We might estimate ϕ by minimising the quantile regression lack of fit criterion

$$\ell_\phi = \{\tau \sum_{i, r_i \geq 0}^{n} |r_i| + (1 - \tau) \sum_{i, r_i < 0}^{n} |r_i|\}$$

for residuals $r_i = x_i - \phi(\theta_i, x_i, y_i; \tau)$. We regulate the smoothness of the quantile function by penalising lack of fit for parameter roughness R_ϕ (with respect to all covariates), by minimising the revised penalised criterion

$$\ell_\phi^* = \ell_\phi + \lambda_\phi R_\phi$$

where the value of roughness coefficient λ_ϕ is selected using cross-validation to provide good predictive performance.

For Poisson modelling, we use penalised likelihood estimation. The rate ρ of threshold exceedance is estimated by minimising the roughness–penalised (negative log) likelihood

$$\ell_\rho^* = \ell_\rho + \lambda_\rho R_\rho$$

where R_ρ is parameter roughness with respect to all covariates, λ_ρ is again evaluated using cross-validation. The Poisson (negative log) likelihood is given by

$$\ell_\rho = - \sum_{i=1}^{n} \log \rho(\theta_i, x_i, y_i) + \int \rho(\theta, x, y) d\theta dx dy$$

which is approximated by

$$\hat{\ell}_\rho = - \sum_{j=1}^{m} c_j \log \rho(j\Delta) + \Delta \sum_{j=1}^{m} \rho(j\Delta)$$

where $\{c_j\}_{j=1}^{m}$ are counts of numbers of threshold exceedances per degree longitude, latitude and storm direction, per annum on an index set of m ($>> 1$) bins on a regular lattice partitioning the covariate domain into intervals of constant volume Δ.

The generalised Pareto model of size of threshold exceedance is estimated in a similar manner by minimising the roughness penalised (negative log) GP likelihood

$$\ell^*_{\xi,\sigma} = \ell_{\xi,\sigma} + \lambda_\xi R_\xi + \lambda_\sigma R_\sigma$$

where R_ξ and R_σ are parameter roughnesses for GP shape and scale with respect to all covariates, and roughness coefficients λ_ξ and λ_σ are evaluated using cross-validation. The GP (negative log) likelihood is given by

$$\ell_{\xi,\sigma} = \sum_{i=1}^{n} \log \sigma_i + \frac{1}{\xi_i} \log(1 + \frac{\xi_i}{\sigma_i}(z_i - \phi_i))$$

where $\phi_i = \phi(\theta_i, x_i, y_i)$, $\xi_i = \xi(\theta_i, x_i, y_i)$ and $\sigma_i = \sigma(\theta_i, x_i, y_i)$, and a similar expression is used when $\xi_i = 0$ (see Jonathan and Ewans 2013).

Return values

Return value z_T of storm peak significant wave height corresponding to some return period T, expressed in years, can be evaluated in terms of estimates for model parameters ϕ, ρ, ξ and σ. For any choice of covariates θ, x, y, the return value is given by

$$z_T = \phi - \frac{\sigma}{\xi}(1 + \frac{1}{\rho}(\log(1 - \frac{1}{T}))^{-\xi})$$

where all of ϕ, ρ, ξ and σ are understood to be functions of θ, x, y, and ρ is expressed as an annual rate of threshold exceedance per location per 1-degree storm direction. Thus, z_{100} corresponds to the 100-year return value, often denoted by H_{S100}. Interpretation of the value of z_T should be undertaken with considerable care. For example, in the current spatio–directional case, the value of $z_{100}(\theta, x, y)$ corresponds to the 100-year return value for storm peak significant wave height at location (x, y) for storm directions within a 1-degree directional sector centred on direction θ.

For estimation of directional design values per location, simulation under the fitted model (incorporating the effects of storm directional dissipation if required) is a better alternative. In the current work, we report omni–directional design values, and design values for directional octants of equal size, centred on cardinal and inter–cardinal storm directions (see Section 3), estimated from simulation. The procedure for simulation, in outline, is given in Algorithm 1 below.

---

input: ϕ, ρ, ξ and σ defined on index set of covariate combinations; ρ in units of number of occurrences per location per 1-degree storm direction per annum; return period of interest, T; number of realisations, R

output: per location: median and 95% uncertainty for z_T omni–directionally and for 8 directional sectors of equal size centred on cardinal and semi–cardinal directions

foreach *location in turn*
    estimate the expected total (omni-directional) number of threshold exceedances in return period (using ρ and T);
    foreach *realisation in turn*
        sample an actual number of threshold exceedances for return period (using the expected total);
        sample storm directions for threshold exceedances (using ρ);
        sample sizes for threshold exceedances (using ϕ, ξ and σ);
        foreach *directional sector in turn*
            store maximum size of threshold exceedance, incorporating dissipation if required;
        end
    end
    store median and 95% uncertainty for z_T band (over realisations) per directional sector per location;
end

Algorithm 1: Procedure for estimation of return values by simulation

Parameter smoothness

Physical considerations suggest we should consider parameters ϕ, ρ, ξ and σ to be smooth functions of covariates θ, x, y. For estimation, this can be achieved by expressing each parameter in terms of an appropriate basis for the domain D of covariates, where $D = D_\theta \times D_x \times D_y$. $D_\theta = [0, 360)$ is the (marginal) domain of storm peak directions, and D_x, D_y are the domains of x- and y-values (e.g. longitudes and latitudes) under consideration.

For each covariate (and marginal domain) in turn, we first calculate a B-spline basis matrix for an index set (of size $<< n$) of covariate values; potentially we could calculate the basis matrix for each of the n observations, but usually avoid this for computation efficiency. Specifically, for D_θ, we calculate basis matrix B_θ ($m_\theta \times p_\theta$) such that the value of any function at each of the m_θ points in the index set for storm direction can be expressed as $B_\theta \beta_\theta$ for some vector β_θ ($p_\theta \times 1$) of basis coefficients. Note that periodic marginal bases can be specified if appropriate (e.g. for D_θ).

Then we define a basis matrix for the three-dimensional domain D using Kronecker products of marginal basis matrices. Thus

$$B = B_\theta \otimes B_x \otimes B_y$$

provides a ($m \times p$) basis matrix (where $m = m_\theta m_x m_y$, and $p = p_\theta p_x p_y$) for modelling each of ϕ, ρ, ξ and σ on the corresponding "spatio–directional" index set (of size m). Any of ϕ, ρ, ξ and σ (η, say, for brevity) can then be expressed in the form $\eta = B\beta$ for some ($p \times 1$) vector β of basis coefficients. Model estimation therefore reduces to estimating appropriate sets of basis coefficients for each of ϕ, ρ, ξ and σ. The values of p_θ, p_x, p_y are functions of the number of spline knots for each marginal domain, and also depend on whether spline bases are specified to be periodic (e.g D_θ) or not (e.g D_x and D_y).

The roughness R of any function can be easily evaluated on the index set (at which $\eta = B\beta$). Following the approach of Eilers and Marx (e.g. Eilers and Marx 2010), writing the vector of differences of consecutive values of β as $\Delta\beta$, and vectors of second and higher order differences using $\Delta^k \beta = \Delta(\Delta^{k-1}\beta)$, $k > 1$, the roughness R of β is given by

$$R = \beta' P \beta$$

where $P = (\Delta^k)'(\Delta^k)$ for differences of order k. We use $k = 1$ throughout this work. With this choice of k, heavy roughness penalisation results in stationarity of parameters with respect to periodic and aperiodic covariates.

Computational considerations

Quantile regression estimation is performed by direct minimisation of the criterion ℓ_ϕ^* from a good starting solution using a linear programming approach (e.g. Koenker 2005 and Bollaerts 2009). A starting solution is estimated by fitting a smoothing spline to estimates of the spatio–directional quantile with non-exceedance probability τ at each of the m covariate combinations on the index set. Poisson and generalised Pareto estimation is achieved using iterative back-fitting (e.g., Davison 2003). Good starting solutions are found to be essential for GP minimisation in particular. These are obtained by estimating local GP models at each of the m members of the index set (or combinations of neighbours thereof to increase sample size), then fitting smoothing spline models for each of GP shape ξ and scale σ. Using algorithms developed for generalised linear array models (Currie et al. 2006), direct computation of Kronecker products of the form $B_\theta \otimes B_x \otimes B_y$ (and some other computationally-demanding operations) can be avoided, providing large reductions in computer memory requirements and execution times. For larger problems, it is also computationally advantageous to adopt an appropriate model cost-complexity criterion (such as AIC) as an alternative to cross-validation, thereby avoiding the need for repeated model estimation.

3 Application

Data and regional climatology

The application sample corresponds to storm peak H_S and dominant wave direction (at storm peak H_S) for 6156 hindcast storm events at 1089 locations on a 33 x 33 regular grid over the North West Shelf of Western Australia for the period 1970-2007.
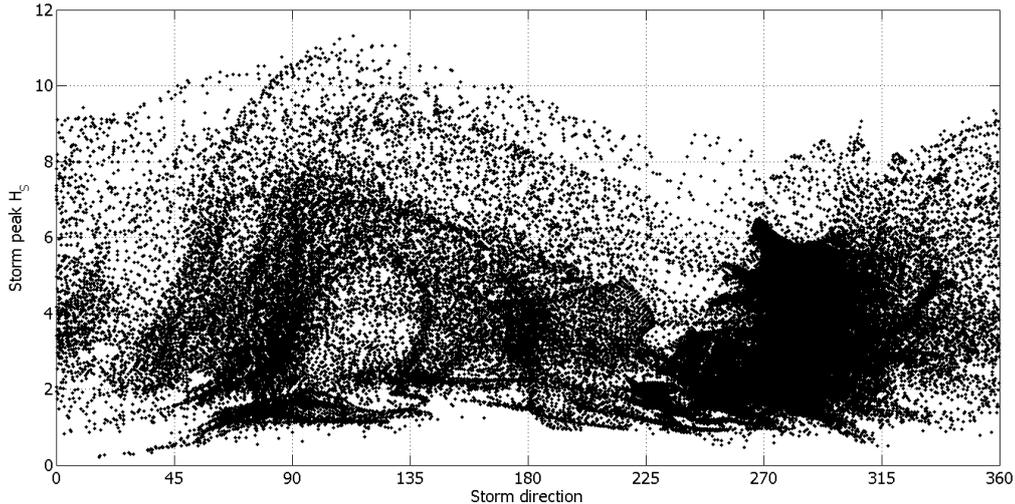
Figure 1: H_S against direction for all 1089 locations. Clear variability with direction, largest H_S from $90^o - 130^o$. Large number of small events from $260^o - 310^o$.

The regional climate is monsoonal, displaying two distinct seasons, "winter" from April to September and "summer" from October to March, with very rapid transitions between these, generally in April and September/October. The winter "dry" is the result of a steady air flow from the east (South East Trade Winds) originating from the Australian mainland, propagating over the Timor Sea. The summer "wet" is the result of the North West Monsoon, a steady, moist flow of air predominantly from the west to south west (and to a lesser extent from the north west). Tropical cyclones occur during summer months and are clearly the most important for extreme met–ocean criteria. Tropical cyclones originate from south of the equator in the eastern Indian Ocean and Timor and Arafura Seas. The most severe cyclones often occur in December and March-April, when sea surface temperatures are highest. In the region under consideration, most storms are tropical lows or developing storms, but can be very severe nevertheless, as exemplified by tropical cyclones Thelma (1998) and Neville (1992). Most storms passing through the region head in a westerly or south westerly direction before turning southwards. The prevailing wave climate comprises contributions from Indian Ocean swell, winter easterly swell, westerly monsoonal swell, tropical cyclone swell, and locally generated wind-sea. Indian Ocean swell is a perennial feature typically, propagating from the south-west through north-west. The largest sea states are wind generated sea states associated with tropical cyclones. Figure 1 shows a scatter plot of H_S against direction, pooled over locations; H_S is non–stationary with respect to storm direction. There are a large number of small events from $260 - 310^o$, seen also in the rose plots of Figure 2; the rate of occurrence of large H_S is also non–stationary with respect to storm direction.

3.1 Spline parameterisation

We fit the spatio–directional spline model assuming an index set of 32 directional bins x 33 longitude bins x 33 latitude bins. Cubic splines are used for each dimension, the directional spline basis being the only periodic. Longitude and latitude domains are characterised by 15 evenly spaced knots, and the directional domain by 32 evenly spaced knots. Therefore, a total of $32 \times 18 \times 18$ spline parameters need to be estimated.

3.2 Extreme value threshold

We estimate threshold ϕ, using spline quantile regression, above which observations $\dot{z}$ of H_S are assumed to be extreme. A number of different quantile non-exceedance thresholds were examined; a 50% threshold was adopted. Figure 3 shows a spatio–directional plot for ϕ. The 8 right–hand plots show ϕ per location corresponding to the 8 cardinal and inter–cardinal directions. Spatially, smallest thresholds occur in the south east, nearest land, whereas
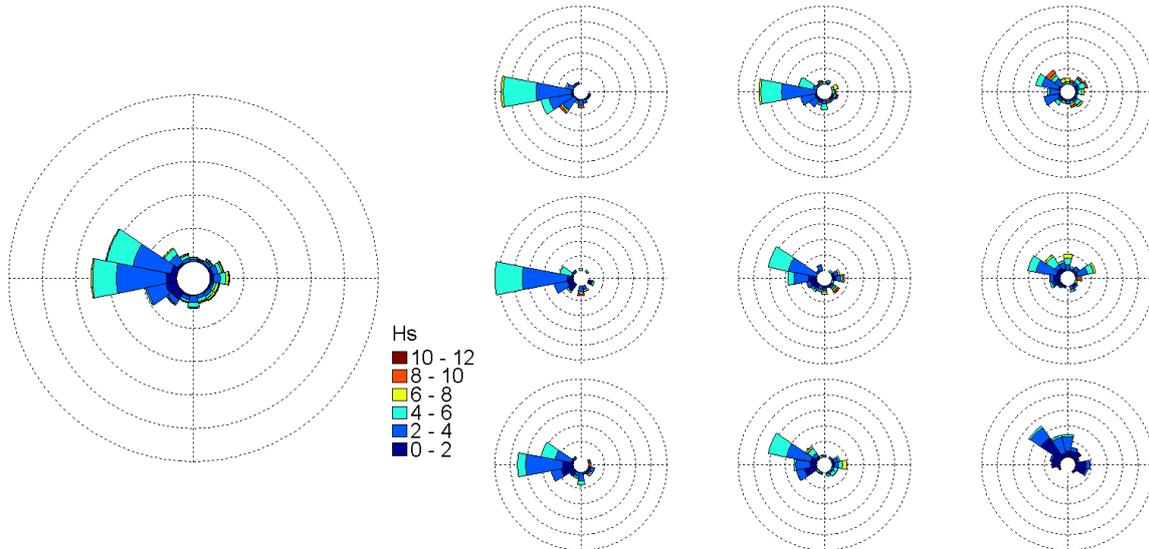
Figure 2: Rose histogram plots of H_S. Size and colour of bin shows proportion of interval of H_S values. Left-hand plot shows pooled rose for all 1089 locations. Right-hand plots show direction histograms for individual locations in the (from left to right, top to bottom) NW, N, NE, W, Central, E, SW, S and SE locations respectively. All right-hand plots (except "central") correspond to locations at the boundary of the spatial domain in the given direction; the "central" plot corresponds to the centre of the spatial domain. In left hand plot, large number of smaller storms from west, smaller number of larger storms from east north-east.

larger thresholds are seen in the north and east in more open ocean. The left–hand plot shows the direction from which ϕ is largest per location. Largest thresholds correspond to the north or south east.

3.3 Rate of threshold exceedance

Next we estimate the rate of occurrence ρ of threshold exceedances using a Poisson process model with rate $\rho(\overset{\triangle}{=} \rho(\theta, x, y))$. Figure 4 shows a spatio–directional plot of estimated ρ. Rate ρ is seen to be relatively constant spatially, but directionally ρ is higher for events from the west (250°-290°). This is consistent with the directional histograms in Figure 3.

3.4 Size of threshold exceedance:

Now we model the sizes of threshold exceedance using a GP model. Figure 5 shows a spatio–directional plot for GP shape parameter, ξ, seen to be largest in general for events from the north east and south west. Figure 6 shows the corresponding plot for GP scale, σ, seen to be higher in general for events with directions from the east. Because of the dependence between estimates of ξ and σ, care should be taken not to over-interpret these plots; inspection of estimates for return values (below) is preferred.

3.5 Return value estimation

Figure 7 shows 100-year return values, H_{S100}, estimated by simulation under the fitted model. Both omni–directional return values and return values corresponding to directional sectors centred on the cardinal and inter–cardinal directions are given. Directional dissipation (which plays no role in omni-directional estimation) is not included in the octant return values. Generally, lower return values are seen in the south east corner nearest land. The highest return values at most locations are generally from the north to north east. However the most north easterly locations have higher return values for storms from the south east to east. Return values are generally lower for storms from the west.

7
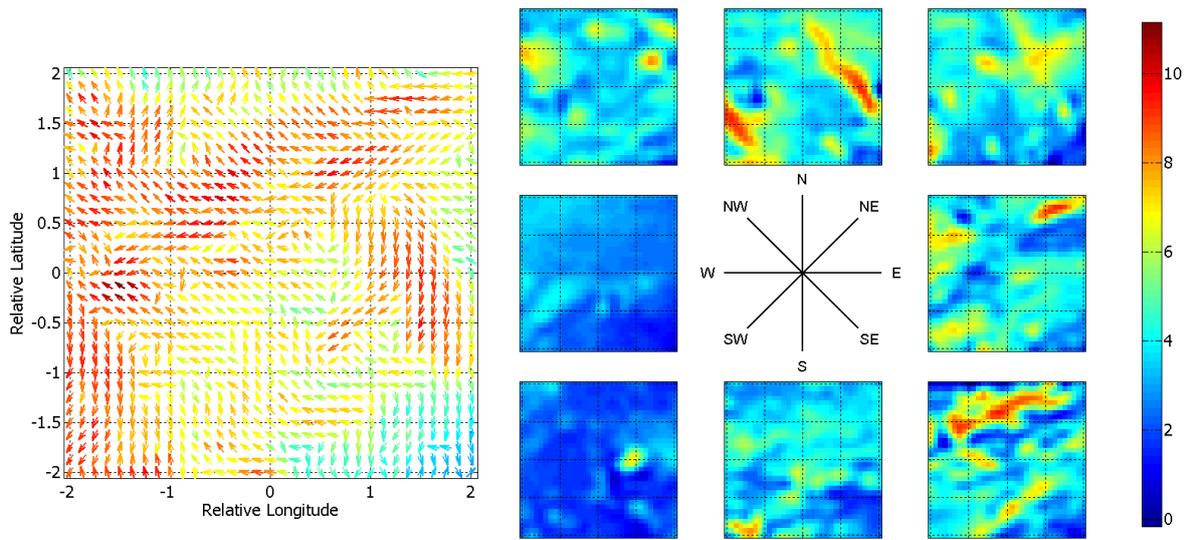
Figure 3: Spatio–directional plot for exceedance threshold, ϕ. The 8 right–hand plots show the 50% threshold values of ϕ for each location for 8 storm directions (from left to right, top to bottom: storms from NW, N, and NE; W and E; SW, S and SE respectively). The left–hand plot shows the direction from which ϕ is largest, for each location. Colour scale in metres.
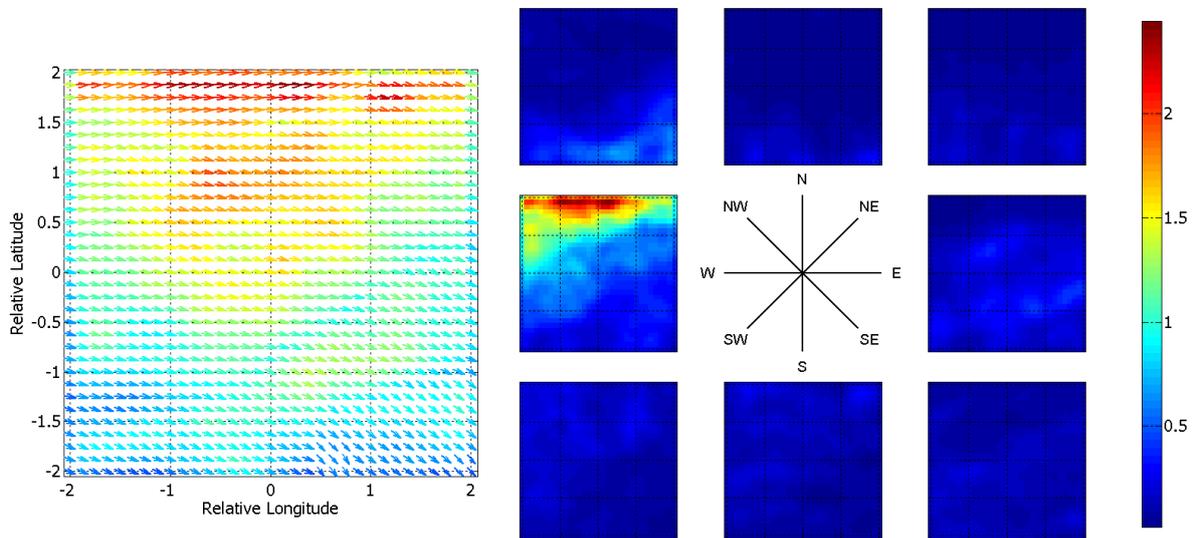
Figure 4: Spatio–directional plot for Poisson rate of threshold exceedance ρ. The 8 right–hand plots show ρ for each location for 8 directions (from left to right, top to bottom: storms from NW, N, and NE; W and E; SW, S and SE respectively). The left–hand plot shows the direction from which ρ is largest for each location. Units of ρ are number of threshold exceedances per degree longitude, latitude and direction, per annum.
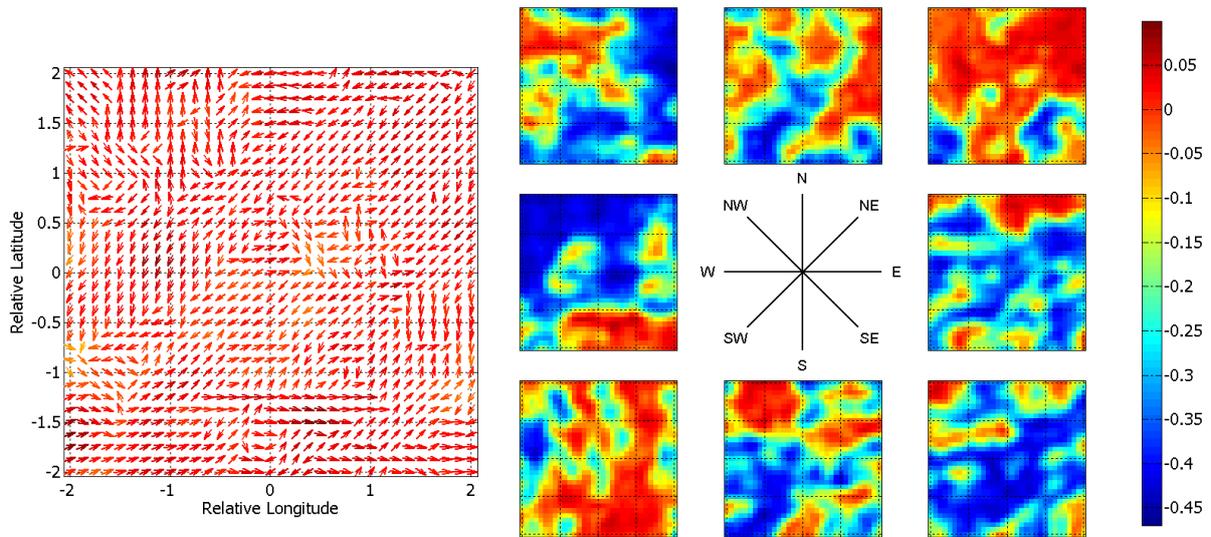
Figure 5: Spatio–directional plot for estimates of GP shape, ξ. The 8 right–hand plots show ξ for each location for 8 directions (from left to right, top to bottom: storms from NW, N, and NE; W and E; SW, S and SE respectively). The left–hand plot shows the direction from which ξ is largest, for each location.
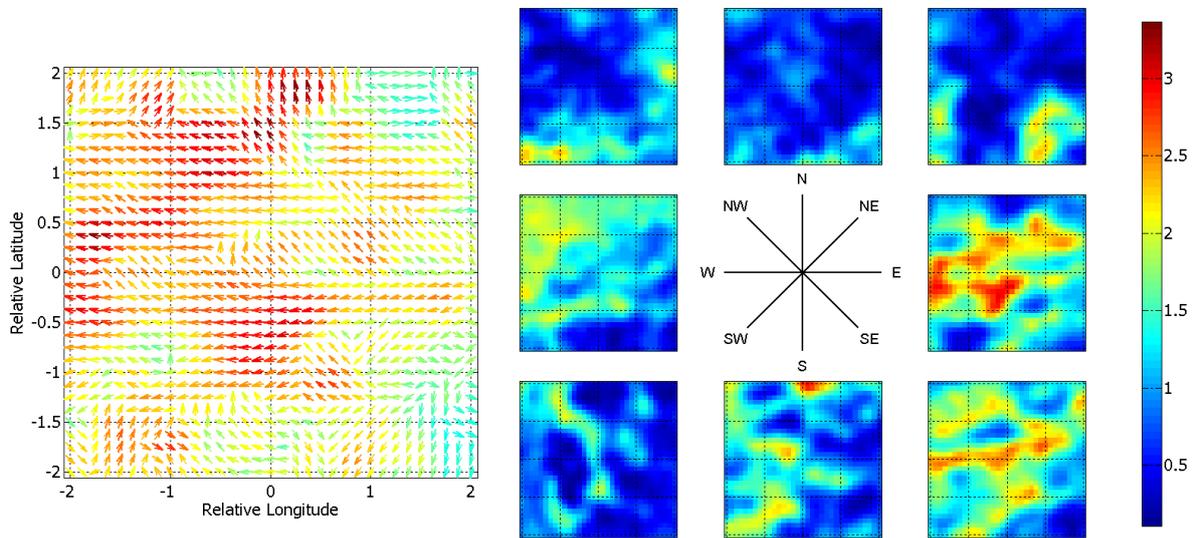
Figure 6: Spatio–directional plot for estimates of GP scale, σ. The 8 right–hand plots show σ for each location for 8 directions (from left to right, top to bottom: storms from NW, N, and NE; W and E; SW, S and SE respectively). The left–hand plot shows the direction from which σ is largest, for each location.

9
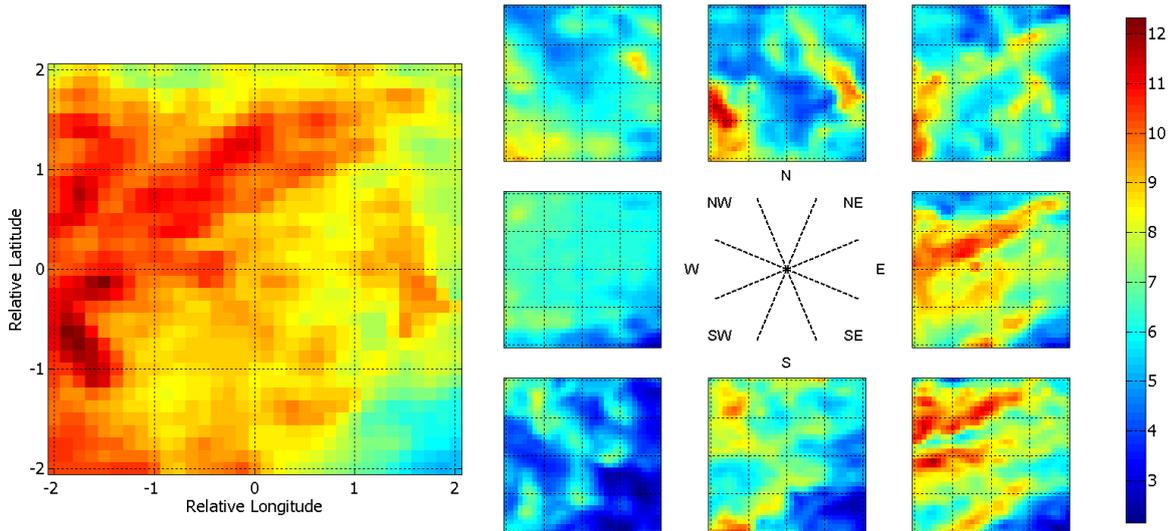
Figure 7: 100-year return values, H_{S100} from simulation. The left–hand plot shows median omni–directional return values for each of the 1089 locations. The 8 right–hand plots show median directional sector H_{S100} per location for 8 directional octants centred (from left to right, top to bottom) on storms from NW, N, and NE; W and E; SW, S and SE respectively. Colour scale in metres.

4   Discussion

In the paper, we introduce a marginal spatio–directional model for extreme storm peak significant wave height, applied to estimation of spatio–directional design values for a neighbourhood of locations off the North West Shelf of Australia. The model uses the peaks over threshold approach, incorporating estimation of an extreme value threshold and the rate and sizes of threshold exceedance. Model parameters are smooth spatio–directional functions. Cross–validation is used to estimate appropriate parameter smoothness in each case. Re–sampling techniques such as bootstrapping can be used to estimate the uncertainty of model parameters and estimates of return values and other structure variables. The model yields parameter estimates and design values consistent with physical intuition and previous estimates.

The main advantage of the approach is that marginal spatial and directional variation of extreme value characteristics is incorporated in a rational, consistent, scalable and computationally-efficient manner eliminating the need for ad-hoc procedures such as site pooling. In isolating storm peak events, we also estimate the directional dissipation (e.g. Jonathan and Ewans 2007a) of storms across locations. This allows us also to estimate design criteria for arbitrary directional sectors for a given location together with the omni-directional estimate, in a consistent manner.

4.1   Conditional extremes

Spline representations are also useful in non–stationary conditional extremes modelling based on the approach of Heffernan and Tawn [2004]. Jonathan et al. [2014] introduces a general–purpose approach, common to all inference steps in conditional extremes inference. Non-crossing quantile regression estimates appropriate non–stationary marginal quantiles simultaneously (for a range of non–exceedance probabilities) as functions of covariate; these are necessary as thresholds for extreme value modelling, and for standardisation of marginal distributions prior to application of the conditional extremes model. Marginal extreme value and conditional extremes modelling is performed within a roughness–penalised likelihood framework, with cross–validation to estimate suitable model parameter roughness. A bootstrap re–sampling procedure, encompassing all inferences, quantifies uncertainties in, and dependence structure of, parameter estimates and estimates of conditional extremes of one variate given large values of another. The approach is validated using simulations from known joint distributions, the extremal dependence structures of which change with covariate. The approach is illustrated in application to joint modelling of storm peak significant wave

10

height and associated storm peak period for extra-tropical storms at northern North Sea and South Atlantic Ocean locations, with storm direction as covariate.

4.2 Spatial extremes

We are currently extending the marginal model presented here to incorporate multivariate spatial dependence, using composite likelihood and censored likelihood methods, so that joint characteristics of extremes of storm peak significant wave height across multiple locations can also be estimated and studied. At present, however, spatial extremes methods suffer from a number of restrictions which may cause biased inferences. Firstly, models are typically developed using block maxima rather than threshold exceedances, since spatial extremes theory is motivated by consideration of component-wise maxima, with full likelihoods replaced with approximations constructed as weighted sums of pairwise likelihoods. Censored composite likelihood approaches permit more efficient analysis of threshold exceedances (e.g. Huser and Davison 2012, draft). Secondly, spatial extremes models only admit certain types of extremal dependence structure (e.g. Ledford and Tawn [1997]), namely perfect independence or asymptotic dependence; they do not admit asymptotic independence in particular. An extension to incorporate asymptotic independence has recently been proposed by Wadsworth and Tawn [2012]. The generalised Pareto process is another promising emerging description for spatial extremes (Ferreira and de Haan 2014). In general, the extremal dependence structure of spatial extremes will itself be non–stationary; estimation of covariate effects in the dependence model will therefore also be necessary, using the spline approach outlined above. For example, the dependence structure of the simplest max-stable process for two spatial locations, known as the Smith process (Smith 1990), is parameterised in terms of a bivariate Gaussian covariance matrix (e.g. Jonathan and Ewans 2013), therefore requiring the estimation of three parameters (two variances and one covariance), all of which in principle may vary spatio-directionally. Other more realistic max-stable processes typically have larger numbers of dependence parameters (e.g., Davison et al. 2012).

Spatial extremes methods are nevertheless potentially of great value in met–ocean design, since they provide a framework within which extremal behaviour of complete ocean basins can be modelled, incorporating appropriate marginal and dependence structure and avoiding the need for site pooling in particular. It might be possible that in future, only one extreme modelling task would be necessary per hindcast. That model, for the whole ocean basin, could then be interrogated routinely to estimate design values for one location, or joint design values for an arbitrary number of arbitrary locations. The ocean engineer would then no longer in principle need to perform further site–specific extreme value analysis.

5 Acknowledgement

References

C.W. Anderson, D.J.T. Carter, and P.D. Cotton. *Wave climate variability and impact on offshore design extremes.* Report commissioned from the University of Sheffield and Satellite Observing Systems for Shell International, 2001.

Kaatje Bollaerts. *Statistical models in epidemiology and quantitative microbial risk assessment applied to salmonella in pork.* PhD thesis, Hasselt University, Belgium, 2009.

E. Casson and S. G. Coles. Spatial regression models for extremes. *Extremes*, 1:449–468, 1999.

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207, 2005.

L. E. Chouinard, C. Liu, and C. K. Cooper. Model for severity of hurricanes in Gulf of Mexico. *J. Wtrwy., Harb., and Coast. Engrg.*, May/June:120–129, 1997.

S. Coles. *An introduction to statistical modelling of extreme values*. Springer, 2001.

S. G. Coles and E. Casson. Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20:283–296, 1998.

S. G. Coles and E. A. Powell. Bayesian methods in extreme value modelling: a review and new developments. *International Statistics Review*, 64:119–136, 1996.

S. G. Coles and J. A. Tawn. A Bayesian analysis of extreme rainfall data. *Applied Statistics*, 45:463–478, 1996.

S. G. Coles and J. A. Tawn. Bayesian modelling of extreme sea surges on the uk east coast. *Phil. Trans. R. Soc. A*, 363:1387–1406, 2005.

I. D. Currie, M. Durban, and P. H. C. Eilers. Generalized linear array models with applications to multidimensional smoothing. *J. Roy. Statist. Soc. B*, 68:259–280, 2006.

A. C. Davison. *Statistical models*. Cambridge University Press, 2003.

A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. *Proc. R. Soc. A*, 468:581–608, 2012.

A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statistical Science*, 27: 161–186, 2012.

A.C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393, 1990.

J. M. Dixon, J. A. Tawn, and J. M. Vassie. Spatial modelling of extreme sea-levels. *Environmetrics*, 9:283–301, 1998.

E.F. Eastoe. *Statistical models for dependent and non-stationary extreme events*. Ph.D. Thesis, University of Lancaster, U.K., 2007.

E.F. Eastoe and J.A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Appl. Statist.*, 58:22–45, 2009.

P H C Eilers and B D Marx. Splines, knots and penalties. *Wiley Interscience Reviews: Computational Statistics*, 2: 637–653, 2010.

R. J. Erhardt and R. L. Smith. Approximate Bayesian computing for spatial extremes. *Available from http://arxiv.org/pdf/1109.4166.pdf*, 2011.

K. C. Ewans and P. Jonathan. The effect of directionality on Northern North Sea extreme wave design criteria. *J. Offshore Mechanics Arctic Engineering*, 130:10, 2008.

A. Ferreira and L. de Haan. The generalised Pareto process. *(Accepted for Bernoulli Journal, preprint at www.bernoulli-society.org)*, 2014.

C. Guedes-Soares and M. Scotto. Modelling uncertainty in long-term predictions of significant wave height. *Ocean Eng.*, 28:329, 2001.

R. R. Haring and J. C. Heideman. Gulf of Mexico rare wave return periods. *Offshore Technology Conference*, OTC3229, 1978.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66: 497, 2004.

R. Huser and A. C. Davison. Space-time modelling of extreme events. *Draft at arxiv.org/abs/1201.3245*, 2012, draft.

P. Jonathan and K. C. Ewans. The effect of directionality on extreme wave design criteria. *Ocean Eng.*, 34:1977–1994, 2007a.

P. Jonathan and K. C. Ewans. Uncertainties in extreme wave height estimates for hurricane dominated regions. *J. Offshore Mechanics Arctic Engineering*, 129:300–305, 2007b.

P. Jonathan and K. C. Ewans. On modelling seasonality of extreme waves. In *Proc. 27th International Conf. on Offshore Mechanics and Arctic Engineering, 4-8 June, Estoril, Portugal*, 2008.

P. Jonathan and K. C. Ewans. A spatiodirectional model for extreme waves in the Gulf of Mexico. *ASME J. Offshore Mech. Arct. Eng.*, 133:011601, 2011.

P. Jonathan and K. C. Ewans. Statistical modelling of extreme ocean environments with implications for marine design : a review. *Ocean Engineering*, 62:91–109, 2013.

P. Jonathan, K. C. Ewans, and G. Z. Forristall. Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects. *Ocean Eng.*, 35:1211–1225, 2008.

P. Jonathan, K. C. Ewans, and D. Randell. Joint modelling of environmental parameters for extreme sea states incorporating covariate effects. *Coastal Engineering*, 79:22–31, 2013.

P. Jonathan, K. C. Ewans, and D. Randell. Non-stationary conditional extremes. *(Submitted to Environmetrics August 2013, draft at www.lancs.ac.uk/∼jonathan)*, 2014.

R. Koenker. *Quantile regression*. Cambridge University Press, 2005.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. R. Statist. Soc. B*, 59:475–499, 1997.

S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *J. Am. Statist. Soc.*, 105:263–277, 2010.

M. E. Robinson and J. A. Tawn. Statistics for extreme sea currents. *Applied Statistics*, 46:183–205, 1997.

M.G. Scotto and C. Guedes-Soares. Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Eng.*, 40:313, 2000.

M.G. Scotto and C. Guedes-Soares. Bayesian inference for long-term prediction of significant wave height. *Coastal Eng.*, 54:393, 2007.

R. L. Smith. Max-stable processes and spatial extremes. *Unpublished article, available electronically from www.stat.unc.edu/postscript/rs/spatex.pdf*, 1990.

J.L. Wadsworth and J.A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99:253–272, 2012.

E. G. Ward, L. E. Borgman, and V. J. Cardone. Statistics of hurricane waves in the Gulf of Mexico. *Offshore Technology Conference*, 1978.