

Joint modelling of extreme ocean environments incorporating covariate effects

Philip Jonathan

Shell Research Ltd., Chester, CH1 3SH, UK.

Kevin Ewans

Sarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia.

David Randell

Shell Research Ltd., Chester, CH1 3SH, UK.

Published. *Coastal Engineering* 79 (2013) 22–31.

Abstract

Characterising the joint distribution of extremes of ocean environmental variables such as significant wave height (H_S) and spectral peak period (T_P) is important for understanding extreme ocean environments and in the design and assessment of marine and coastal structures. Many applications of multivariate extreme value analysis adopt models that assume a particular form of extremal dependence between variables without justification. Models are also typically restricted to joint regions in which all variables are extreme, but regions where only a subset of variables are extreme can be equally important for design. The conditional extremes model of Heffernan and Tawn (2004) provides one approach to overcoming these difficulties.

Here, we extend the conditional extremes model to incorporate covariate effects in all of threshold selection, marginal and dependence modelling. Quantile regression is used to select appropriate covariate-dependent extreme value thresholds. Marginal and dependence modelling of extremes is performed within a penalised likelihood framework, using a Fourier parameterisation of marginal and dependence model parameters, with cross-validation to estimate suitable model parameter roughness, and bootstrapping to estimate parameter uncertainty with respect to covariate.

We illustrate the approach in application to joint modelling of storm peak H_S and T_P at a Northern North Sea location with storm direction as covariate. We evaluate the impact of incorporating directional effects on estimates for return values, including those of a structure variable, similar to the structural response of a floating structure. We believe the approach offers the ocean engineer a straightforward procedure, based on sound statistics, to incorporate covariate effects in estimation of joint extreme environmental conditions.

Keywords: offshore design; joint extremes; conditional extremes; covariates;

*philip.jonathan@shell.com

1. Introduction

It is well known that the characteristics of extreme ocean environments vary with respect to a number of covariates. For example, extremes of H_S vary with wave direction and season (as demonstrated by, e.g., Ewans and Jonathan (2008) and Jonathan and Ewans (2011)). Incorporation of covariate effects is important to successful marginal modelling of environmental extremes (see, e.g., Jonathan et al. 2008). It would seem reasonable, therefore, to expect that covariate effects are important in modelling joint extremes also in general.

Marginal modelling of extremes with covariates has been performed for many years. Some authors follow the approach of Davison and Smith (1990), parameterising extreme value model parameters in terms of one or more covariates (see, e.g., Chavez-Demoulin and Davison 2005). Other authors prefer to transform (or *whiten*) the sample to remove the effects of covariates (see Eastoe and Tawn 2009) prior to extreme value analysis. Marginal (and dependence) modelling of extremes with covariates requires the specification of a threshold for extreme value modelling. A number of authors (see, e.g., Anderson et al. 2001) have commented on the importance of specifying a covariate-dependent threshold when covariate effects are suspected. One approach to covariate-dependent threshold specification is quantile regression (Koenker 2005), illustrated recently in environmental applications by Kyselý et al. (2010) and Northrop and Jonathan (2011).

Characterising the joint distribution of extremes of ocean environmental parameters is important in understanding extreme ocean environments and in the design and assessment of marine structures. The conditional extremes model of Heffernan and Tawn (2004) provides a straightforward procedure for modelling multivariate extremal dependence in the absence of covariates. Jonathan et al. (2010) illustrate the application of conditional extremes model to characterise the dependence structure of storm peak significant wave height (H_S) and wave spectral peak period (T_P) and estimate the return values of T_P conditional on extreme values of H_S . To estimate the conditional extremes model for bivariate extremes of random

variables \dot{X}_1 (H_S , say) and \dot{X}_2 (T_P , say), the following procedure is appropriate. (a) Select a range of appropriate thresholds for threshold exceedance modelling for each variable in turn. (b) Fit marginal generalised Pareto models to threshold exceedances of the sample data for each variable in turn for different threshold choices, plot the values of model parameter estimates as a function of threshold, and select the lowest threshold value per variable corresponding to approximately stable models. (c) Transform \dot{X}_1 and \dot{X}_2 in turn to Gumbel scale (to X_1 and X_2) using the probability integral transform. (d) Fit the conditional extremes model for $X_2|X_1$ (and $X_1|X_2$) in turn for various choices of threshold of the conditioning variate, retain the estimated model parameters and residuals, plot the values of model parameter estimates and examine residuals as a function of threshold, and select the lowest threshold per variable consistent with modelling assumptions. (e) Simulate joint extremes on the standard Gumbel scale under the model, and transform realisations to the original scale using the probability integral transform.

To our knowledge, there is no literature on incorporation of covariate effects within the conditional extremes model, the topic of this article. The layout of the paper is as follows. In section 2, we introduce a motivating application. Section 3 is a description of the marginal model (incorporating threshold modelling, generalised Pareto modelling of threshold exceedances and transformation to Gumbel scale), and the extended conditional extremes model incorporating covariate effects. Section 4 addresses the estimation of conditional extremes for the Northern North Sea location under consideration, followed by a more general discussion and conclusions in section 5.

2. Data

We motivate and illustrate the methodology by considering joint estimation of extreme values of storm peak H_S and T_P with directional covariate at a location in the northern North Sea. Data correspond to hindcast values of storm peak H_S over threshold, observed during periods of storm events, and associated values for T_P . The sample consists of 1163 pairs of values for the period October 1964 to August 1998.

Figure 1 shows the approximate North Sea location corresponding to the data, and the relatively long fetches for waves emanating from the Atlantic Ocean, the Norwegian Sea and the North Sea. With direction from which waves travel expressed in degrees clockwise with respect to north, Figure 2 gives scatter plots of H_S (horizontal) and T_P (vertical) for directional sectors corresponding approximately to those in Figure 1. It can be seen that $H_S - T_P$ dependence for Atlantic storms (directional sector [230, 280)) is quite different to that for storms emanating from the south (directional sector [140, 200)), suggesting that the dependence between H_S and T_P varies as a function of direction, and that this dependence should be accommodated in any joint modelling of extremes of H_S and T_P .

[Figure 1 about here.]

[Figure 2 about here.]

3. Model

The application introduced above suggests that we treat storm peak H_S and associated T_P as varying with storm direction in order to characterise their joint extremal behaviour. Consider therefore two random variables $\dot{X}_1(\theta), \dot{X}_2(\theta)$ of common covariate θ . We are interested in modelling their joint extremal structure for any particular value of covariate θ . We assume that the joint tail of $\dot{X}_1(\theta)$ and $\dot{X}_2(\theta)$ can be characterised adequately using the single covariate θ .

In this section we describe the extension of the conditional extremes modelling procedure, outlined in section 1, to incorporate covariates. Sections 3.1 and 3.2 outline the coupled marginal generalised Pareto modelling of threshold exceedances and quantile regression for threshold selection respectively. Section 3.3 discusses transformation to standard Gumbel scale, necessary for application of the conditional extremes model in section 3.4. Finally, the adoption of Fourier series representations for model parameter functions is outlined in section 3.5.

3.1. Generalised Pareto model for threshold exceedances

Marginally, for each of $\dot{X}_1(\theta)$, $\dot{X}_2(\theta)$ in turn for a given value of θ , we assume that, conditional on exceeding a large value, the corresponding random variables are generalised Pareto distributed:

$$\Pr(\dot{X}_j(\theta) > \dot{x} | \dot{X}_j(\theta) > \psi_j(\theta; \tau_{j*})) = \left(1 + \frac{\xi_j(\theta)}{\zeta_j(\theta)} (\dot{x} - \psi_j(\theta; \tau_{j*}))\right)^{-1/\xi_j(\theta)} \text{ for } j = 1, 2$$

for $\dot{x} > \psi_j(\theta; \tau_{j*})$, $(1 + \frac{\xi_j(\theta)}{\zeta_j(\theta)} (\dot{x} - \psi_j(\theta; \tau_{j*}))) > 0$ and $\zeta_j(\theta) > 0$. $\psi_j(\theta; \tau_{j*})$ is a pre-selected quantile threshold, assumed to be a smooth function of θ , associated with a non exceedance probability τ_{j*} :

$$\Pr(\dot{X}_j(\theta) \leq \psi_j(\theta; \tau_{j*})) = \tau_{j*} \quad .$$

Model parameters $\xi_j(\theta)$, $\zeta_j(\theta)$, respectively generalised Pareto shape and scale, are also assumed to be smooth functions of covariate. For a sample of values $\{\dot{x}_{ij}\}_{i=1}^n$, $j = 1, 2$, corresponding to set $\{\theta_i\}_{i=1}^n$ of known covariate values, and pre-specified threshold $\psi_j(\theta; \tau_{j*})$, estimates for the values of the functions $\xi_j(\theta)$ and $\zeta_j(\theta)$ at $\{\theta_i\}_{i=1}^n$ can be obtained in principle by maximum likelihood estimation, by minimising the negative log-likelihood:

$$\ell_{GP,j} = \sum_{i=1}^n \log \zeta_j(\theta_i) + \frac{1}{\xi_j(\theta_i)} \log \left(1 + \frac{\xi_j(\theta_i)}{\zeta_j(\theta_i)} (\dot{x}_{ij} - \psi_j(\theta_i; \tau_{j*}))\right) \text{ for } j = 1, 2 \quad .$$

Each of the parameter functions $\psi_j(\theta; \tau_{j*})$, $\xi_j(\theta)$ and $\zeta_j(\theta)$ can be specified as a linear combination of suitable basis functions, such as periodic splines and Fourier series for periodic covariates such as direction, as discussed in section 3.5. In this case, we regulate parameter smoothness with covariate using a penalised likelihood fitting criterion:

$$\ell_{GP,j}^* = \ell_{GP,j} + \lambda_{\xi_j} R_{\xi_j} + \lambda_{\zeta_j} R_{\zeta_j} \text{ for } j = 1, 2$$

for roughness coefficients λ_{ξ_j} , λ_{ζ_j} , and parameter roughnesses R_{ξ_j} , R_{ζ_j} which are easily evaluated (see Section 3.5) for suitable choice of basis. The values of roughness coefficients are selected using cross-validation to maximise the predictive performance of the model. 10-fold cross-validation was used throughout this work.

3.2. Quantile regression model for thresholds

For each random variable in turn, the quantile threshold $\psi_j(\theta; \tau)$ corresponding to quantile probability τ is estimated using quantile regression, by minimising the roughness penalised loss criterion:

$$\ell_{QR,j}^* = \left\{ \tau \sum_{i,r_{ij} \geq 0}^n |r_{ij}| + (1 - \tau) \sum_{i,r_{ij} < 0}^n |r_{ij}| \right\} + \lambda_{\psi_j} R_{\psi_j} \text{ for } j = 1, 2$$

and residuals $r_{ij} = \hat{x}_{ij} - \psi_j(\theta_i; \tau)$. The terms in parentheses correspond to the unpenalised quantile regression loss criterion. Parameter roughness R_{ψ_j} can be evaluated in closed form (see Section 3.5) for efficient estimation. The value of roughness coefficient λ_{ψ_j} is selected using cross-validation to maximise the predictive performance of the quantile regression model.

In practice, quantile regression thresholds are estimated for an increasing sequence of D quantile probabilities $0 < \tau_1 < \tau_2 < \dots < \tau_d < \dots < \tau_D < 1$. For each choice of τ_d , standard diagnostic plots for generalised Pareto fitting (such as the variation of the estimated shape parameter or some extreme quantile estimate with threshold) are examined (see, e.g., Scarrott and MacDonald 2012). The lowest value of quantile probability consistent with an adequate generalised Pareto fit is selected as τ_{j*} . The quantile regression threshold estimates for quantile probabilities $\leq \tau_{j*}$ are useful for marginal transformation to standard Gumbel scale, discussed in the next section.

3.3. Marginal transformation to standard Gumbel scale

The conditional extremes model is applied to random variables with standard Gumbel marginal distributions. For each random variable in turn, the quantile regression models for different quantile probabilities, and the marginal generalised Pareto model for threshold exceedances, provide a means to transform from sample $\{\dot{x}_{ij}\}_{i=1}^n$ corresponding to random variable $\dot{X}_j(\theta)$ at the set $\{\theta_i\}_{i=1}^n$ of known covariate values to an equivalent sample $\{x_{ij}\}_{i=1}^n$ corresponding to random variable $X_j(\theta)$ with approximately a standard Gumbel distribution for any θ .

Above the threshold $\psi_j(\theta; \tau_{j*})$, the unconditional cumulative distribution function for threshold exceedances $\dot{x} > \psi_j(\theta; \tau_{j*})$, for any value of θ , is given by:

$$\Pr(\dot{X}_j(\theta) \leq \dot{x}) = 1 - (1 - \tau_{j*}) \Pr(\dot{X}_j(\theta) > \dot{x} | \dot{X}_j(\theta) > \psi_j(\theta; \tau_{j*})) \text{ for } j = 1, 2 \quad .$$

Below the threshold, in the absence of a parametric form for the cumulative distribution function, we approximate it using:

$$\Pr(\dot{X}_j(\theta) \leq \dot{x}) \approx \tau_d + (\tau_d - \tau_{d-1}) \frac{(\dot{x} - \psi_j(\theta; \tau_{d-1}))}{(\psi_j(\theta; \tau_d) - \psi_j(\theta; \tau_{d-1}))} \text{ for } j = 1, 2$$

where $\psi_j(\theta; \tau_{d-1}) \leq \dot{x} < \psi_j(\theta; \tau_d)$ for the sequence of quantile probabilities τ_d such that $\tau_d \leq \tau_{j*}$.

Using the probability integral transform, we can transform from original to standard Gumbel scales, since:

$$\Pr(X_j(\theta) \leq x) = \exp(-\exp(-x)) = \Pr(\dot{X}_j(\theta) \leq \dot{x}) \text{ for } j = 1, 2$$

so that the individuals in the transformed sample $\{x_{ij}\}_{i=1}^n$, $j = 1, 2$ are given by:

$$x_{ij} = -\log(-\log(\Pr(\dot{X}_j(\theta_i) \leq \dot{x}_{ij}))) \text{ for } i = 1, 2, \dots, n, \text{ and } j = 1, 2$$

now assumed to be marginally stationary.

3.4. Conditional extremes model

For positively dependent random variables $X_1(\theta)$, $X_2(\theta)$ with standard Gumbel marginal distributions for any θ , we extend the asymptotic argument of Heffernan and Tawn (2004) for the form of the conditional distribution of $X_{j^c}(\theta)$, $j^c = 1, 2$ given the value of $X_j(\theta)$, $j = 1, 2, j \neq j^c$ for any value of covariate θ :

$$(X_{j^c}(\theta)|X_j(\theta) = x) = \alpha_j(\theta)x + x^{\beta_j(\theta)}W_j(\theta) \text{ for } j, j^c = 1, 2, j^c \neq j$$

and $x > \phi_j(\kappa_{j^*})$ where $\phi_j(\kappa_{j^*})$ is a threshold with non-exceedance probability κ_{j^*} above which the conditional extremes model fits well. The parameter functions $\alpha_j(\theta) \in [0, 1]$, $\beta_j(\theta) \in (-\infty, 1]$ vary smoothly with covariate θ . $W_j(\theta)$ is a random variable drawn from an unknown distribution. We assume that the standardised variable $Z_j = (W_j(\theta) - \mu_j(\theta))/\sigma_j(\theta)$ follows a common distribution G_j , independent of covariate, for smooth location and scale parameter functions $\mu_j(\theta)$, $\sigma_j(\theta) > 0$. We write, for any value of θ :

$$(X_{j^c}(\theta)|X_j(\theta) = x) = \alpha_j(\theta)x + x^{\beta_j(\theta)}(\mu_j(\theta) + \sigma_j(\theta)Z_j) \text{ for } j, j^c = 1, 2, j^c \neq j \quad .$$

For potentially negatively dependent variables, extended forms of the equations above are available in the covariate-free case (see Heffernan and Tawn 2004). (Recently, Keef et al. (2013) proposed marginal transformation to the Laplace distribution rather than Gumbel. The former has exponential tails on both sides and symmetry, capturing the exponential upper tail of the Gumbel required for modelling positive dependence but the symmetry also allows for negatively associated variables to be incorporated into the model parsimoniously.) To estimate the parameter functions $\alpha_j(\theta)$, $\beta_j(\theta)$, $\mu_j(\theta)$ and $\sigma_j(\theta)$, we follow Heffernan and Tawn (2004) in assuming that G_j is the standard normal distribution. The corresponding

negative log likelihood for pairs $\{x_{i1}, x_{i2}\}$ from the original sample for which $x_{ij} > \phi_j(\kappa_{j*})$, conditioned on $X_j(\theta)$ is:

$$\ell_{CE,j} = \sum_{i, x_{ij} > \phi_j(\theta_i; \kappa_{j*})} \log s_{ij} + \frac{(x_{ij^c} - m_{ij})^2}{2s_{ij}^2} \text{ for } j, j^c = 1, 2, j^c \neq j$$

where $m_{ij} = \alpha_j(\theta_i)x_{ij} + \mu_j(\theta_i)x_{ij}^{\beta_j(\theta_i)}$ and $s_{ij} = \sigma_{ij}x_{ij}^{\beta_j(\theta_i)}$. Adopting a penalisation procedure to regulate parameter roughness, the penalised negative log likelihood is:

$$\ell_{CE,j}^* = \ell_{CE,j} + \lambda_{\alpha_j} R_{\alpha_j} + \lambda_{\beta_j} R_{\beta_j} + \lambda_{\mu_j} R_{\mu_j} + \lambda_{\sigma_j} R_{\sigma_j} \text{ for } j = 1, 2$$

where parameter roughnesses $R_{\alpha_j}, R_{\beta_j}, R_{\mu_j}, R_{\sigma_j}$ are easily evaluated (see Section 3.5), and roughness coefficients $\lambda_{\alpha_j}, \lambda_{\beta_j}, \lambda_{\mu_j}, \lambda_{\sigma_j}$ are estimated using cross-validation. To reduce computational burden, we choose to fix the relative size of the roughness coefficients. Residuals:

$$r_{ij} = \frac{1}{\hat{\sigma}_j(\theta_i)} ((x_{ij^c} - \hat{\alpha}_j(\theta_i)x_{ij})x_{ij}^{-\hat{\beta}_j(\theta_i)} - \hat{\mu}_j(\theta_i)) \text{ for } j, j^c = 1, 2, j^c \neq j$$

evaluated for $x_{ij} > \phi_j(\kappa_{j*})$ are inspected to confirm reasonable model fit, as discussed in section 4. The set of residuals is also used as a random sample of values for Z_j from the unknown distribution G_j for simulation to estimate extremes quantiles in section 5.

3.5. Parameter functional forms

Motivated by the application in section 2, we assume that the common covariate θ is periodic on $[0, 360)$. It is then natural to represent parameter functions $\psi_j(\theta), \xi_j(\theta), \zeta_j(\theta), \alpha_j(\theta), \beta_j(\theta), \mu_j(\theta)$ and $\sigma_j(\theta)$ using Fourier series. Adopting the notation $\eta(\theta)$ for a typical parameter function:

$$\eta(\theta) = \sum_{k=0}^p a_{\eta k} \cos(k\theta) + b_{\eta k} \sin(k\theta)$$

for Fourier coefficients $a_{\eta k}, b_{\eta k}$ with $b_{\eta 0} \triangleq 0$, and Fourier order p . The roughness of $\eta(\theta)$ with respect to θ can then be evaluated as:

$$R_\theta = \int_0^{360} (\eta''(\theta))^2 d\theta = \sum_{k=0}^p k^4 (a_{\eta k}^2 + b_{\eta k}^2)$$

where $\eta''(\theta)$ is the second derivative of $\eta(\theta)$ with respect to θ . This simple form facilitates efficient parameter estimation using roughness penalisation described above for quantile regression, generalised Pareto modelling and conditional extremes modelling.

Various other choices of basis are available. For periodic covariates, a basis of periodic splines (see, e.g., Eilers and Marx 2010) provides a good alternative.

4. Application

Marginal quantile regression thresholds $\psi_j(\theta; \tau_d)$ for quantile probabilities τ_d of 0.1, 0.2, ..., 0.9 were estimated for storm peak H_S ($j = 1$) and T_P ($j = 2$), and used in turn for generalised Pareto modelling. For quantile regression only, we found it advantageous to use evenly-spread transformed covariate values $\{\theta_i^*\}_{i=1}^n$, with:

$$\theta_i^* = \frac{360}{n} (r(\theta_i) - 1) \text{ for } i = 1, 2, \dots, n$$

where $r(\theta_i)$ is the rank of θ_i in the set of covariates, namely the position of θ_i in the set of covariate values sorted in ascending order. The set $\{\theta_i^*\}_{i=1}^n$ is uniformly distributed on $[0, 360)$ by design, stabilising quantile regression estimation on the transformed θ^* scale. Interpreted on the original θ scale, the transformation imposes greater smoothness on quantile thresholds in directional sectors less frequently observed, and allows greater threshold flexibility in more frequently observed sectors, in a natural way according to the rate of occurrence of events from different directions.

The marginal directional quantiles for storm peak $H_S(\theta)$ and $T_P(\theta)$ are illustrated in Figure 3, showing a clear directional dependence in each case. For H_S , longer tails are evident for directions corresponding to longer fetches (e.g. the North Sea, the Atlantic Ocean and Norwegian Sea). For T_P , a longer tail is evident for the Atlantic sector, reflecting the occurrence of long-period swell events and the larger extreme wind seas for that sector. The greater smoothness of quantile thresholds for sectors corresponding to land shadows (e.g. Norway, achieved by transforming directional values to approximately uniform scale as described above) is also evident and intuitively reasonable.

[Figure 3 about here.]

In the case that one of the random variables is regarded as the dominant design variable, e.g. $H_S(\theta)$ for a fixed marine structure, there is usually interest in estimating the distribution of the associated random variable, e.g. $T_P(\theta)$, conditioned on an extreme value of the conditioning dominant design variate for one or more choices of covariate θ . For this reason we focus here on modelling $T_P(\theta)$ given large values of $H_S(\theta)$ as a function of θ . The analogous procedure can be used of course to model $H_S(\theta)$ given large $T_P(\theta)$.

Following transformation to Gumbel scale, the conditional extremes model for $T_P(\theta)$ given $H_S(\theta)$ was estimated using penalised likelihood with $\kappa_{1*} = 0.8$. Estimates for parameters α , β , μ and σ (all functions of θ) are given in Figure 4 as solid lines, together with 2.5%, 50% and 97.5% percentiles from a bootstrapping analysis (using 1000 resamples). There is good agreement between the median bootstrap and point estimates. Uncertainty bands are also reasonable for all parameters except β , which is difficult to identify for α close to unity (see Keef et al. 2013 for a constrained solution). The influence of the directional covariate is again clear, particularly for α and σ . For α , this suggests that the dependence between H_S and T_P is greatest in the North Sea sector; we also note that dependence is positive for all directions. For σ , the estimates suggest that the variability in the dependence between H_S and T_P is greater in the Atlantic and Norwegian sectors. Corresponding residuals (see Section 3.4) are plotted against direction and conditioning

variate in Figures 9 and 10 in the Appendix; no obvious inconsistencies with modelling assumptions (see, e.g., Jonathan et al. 2010) are observed. For comparison, however, Figure 11 in the Appendix illustrates residuals from a covariate-free conditional extremes model for the same starting sample. The dependence of residual on direction is apparent, suggesting that this model would not be suitable.

[Figure 4 about here.]

5. Estimation of extreme quantiles

In this section we use simulation under the fitted model to estimate the conditional distributions of conditioned variates and structure variables given large values of the conditioning variate, as a function of covariate. We start by outlining the simulation procedure.

5.1. Simulation procedure

To simulate a realisation of an exceedance of a high quantile of the conditioning variate $X_j(\theta)$ ($j = 1, 2$) and a corresponding value of the conditioned variate $X_{j^c}(\theta)$ ($j^c = 1, 2, j^c \neq j$), for some θ , we proceed as follows:

1. Draw a value of covariate θ_s (e.g. at random from the original sample or from some estimate of its distribution estimated from the sample),
2. Draw a value of residual r_{sj} from the set of residuals obtained during model fitting (see Section 3.4),
3. Draw a value x_{sj} of the conditioning variate from its standard Gumbel distribution,
4. If the value x_{sj} exceeds $\phi_j(\kappa_{j*})$ continue, else resample x_{sj} ,
5. Estimate the value of the conditioned variate x_{sj^c} using:

$$x_{sj^c} = \hat{\alpha}_j(\theta_s)x_{sj} + x_{sj}^{\hat{\beta}_j(\theta_s)}(\hat{\mu}_j(\theta_s) + \hat{\sigma}_j(\theta_s)r_{sj})$$

where the obvious notation is used for the estimated values of model parameters (see Section 3.4),

6. Transform the pair x_{sj}, x_{sj^c} in turn to the original scale (to $\hat{x}_{sj}, \hat{x}_{sj^c}$) using the probability integral transform (see Section 3.3).

Note that this procedure can be extended to include realisations for which $x_{sj} \leq \phi_j(\kappa_{j*})$, by drawing a pair of values (for the conditioning and conditioned variates) at random from the subset of the Gumbel-transformed original sample (for which the conditioning variate is $\leq \phi_j(\kappa_{j*})$) at step 4.

5.2. Covariate-dependent conditional distributions

Using the simulation procedure above, the conditional distribution of $T_P(\theta)$ given large values of $H_S(\theta)$ exceeding its directional quantile with non-exceedance probability 0.99 (in 34 years) was estimated, and is illustrated in Figure 5. A solid black line represents the median value of $T_P(\theta)$, and dashed black lines give the 2.5% and 97.5% percentiles of the distribution. Also shown (as dark grey straight lines) are the corresponding values estimated using a covariate-free conditional extremes model. For further comparison, the marginal $T_P(\theta)$ curve with non-exceedance probability 0.99 is also shown (in light grey).

[Figure 5 about here.]

From the figure it can be seen that the conditional values of T_P follow a similar trend to marginal T_P with direction, except that the conditional values are smaller (due to the imperfect dependence between H_S and T_P , resulting from swell events in the data). The uncertainty of the conditional T_P is greatest in the Atlantic sector, reflecting the original sample (see Figure 2). In contrast, the fluctuation of conditional T_P with direction is not captured by the covariate-free conditional extremes model, which in particular underestimates conditional T_P for the Atlantic and Norwegian Sea sectors. Nonetheless, the location and uncertainty of estimates from the covariate-free model are generally consistent with those from the directional model. These estimates are illustrated alongside those for marginal $H_S(\theta)$ with quantile non-

exceedance probability of 0.99 (in 34 years) in Figure 6. The directional variation of return values for H_S and conditional T_P are consistent with physical understanding of fetch and land shadow effects.

[Figure 6 about here.]

5.3. Covariate-dependent conditional distribution of structure variable

Simulations under the estimated conditional extremes model can also be useful to estimate return values for structure variables defined in terms of the conditioning and conditioned variates. For illustration, the roll or heave response $R(\theta)$ of a floating structure with direction θ can be described in terms of $H_S(\theta)$ and $T_P(\theta)$ using a functional form similar to:

$$R(\theta) = \frac{AH_S(\theta)}{1 + B/T_P(\theta)^2}$$

for $A, B > 0$, both potentially also functions of covariate. For the response $R(\theta)$ illustrated in Figure 7 (in terms of R/H_S against θ , for constant A, B), the variation of median conditional response (given that H_S exceeds its quantile threshold with non-exceedance probability 0.99 in 34 years) with direction is shown in Figure 8 (as a solid black curve). Also shown are 2.5% and 97.5% percentiles of the distribution of conditional distribution (as dashed black curves). For comparison, Figure 8 also gives the corresponding (straight) lines corresponding to a covariate-free conditional extremes model. Conditional response is underestimated by the covariate-free model for North Sea, Atlantic and Norwegian Sea sectors. Again, however, the general location and spread of covariate-free estimates are in agreement with estimates from the directional conditional extremes model.

[Figure 7 about here.]

[Figure 8 about here.]

6. Discussion

In this article we demonstrate that the joint tail of storm peak significant wave height and associated peak period shows storm directional dependence for hindcast data from one Northern North Sea location. We extend the conditional extremes model of Heffernan and Tawn (2004) to incorporate covariate effects. We show that estimates of conditional extremes of peak period ignoring directional variability are different to those which allow the characteristics of the joint tail region to vary smoothly with storm direction. We conclude that neglecting directional covariate effects in joint tail modelling can lead to misleading estimates of return values for the variables concerned. We believe therefore that the conditional extremes model incorporating covariates is a useful complement to existing methods for extreme value analysis in ocean design.

The conditional extremes model estimates the dependence between random variables independently of their marginal characteristics (see, e.g., Jonathan and Ewans 2013, Heffernan and Tawn 2004). Moreover, it adopts appropriate model forms (known from asymptotic extreme value theory) for both marginal (e.g. generalised Pareto for peaks over threshold) and dependence models (e.g. the Heffernan and Tawn model for variables with Gumbel marginal distributions) of extreme values. Jonathan et al. (2009) illustrate the conditional extremes approach in the absence of covariate effects, for estimation of joint extremes of storm peak H_S and associated T_P , and compare the approach with that of Haver (1985). The latter assumes that large values of H_S follow a Weibull distribution, and that conditional values of T_P given H_S follow a log-normal distribution. The conditional extremes model is shown to perform better than the Haver model for simulated samples with known extremal characteristics. The main reason for this is that there is no prescribed model form for extrapolation of the parameter estimates of the Haver (and similar empirical) models beyond the domain of the data. We would therefore also expect the conditional extremes model to also provide more realistic estimates of characteristic structure variables.

A fuller examination of the conditions required for the (original covariate-free) conditional extremes model

to be valid is given by Heffernan and Resnick (2007). An interesting discussion of the merits of the model is included in the discussion of the original paper of Heffernan and Tawn (2004). The limit assumption underlying the extended conditional extremes model is itself an extension of that made by Heffernan and Tawn (2004). Informally, in the notation of Section 3, we assume that for positively dependent random variables $X_1(\theta)$ and $X_2(\theta)$ of common covariate θ with standard Gumbel marginal distributions for any value of θ , the standardised variable:

$$Z_j = \sigma_j(\theta)^{-1} \left(\frac{X_{j^c}(\theta) - \alpha_j(\theta)}{x_j^{\beta_j(\theta)}} - \mu_j(\theta) \right) \text{ for } j, j^c = 1, 2, j^c \neq j$$

is such that:

$$\Pr(Z_j \leq z | X_j = x_j, \theta) \rightarrow G(z) \text{ as } z \rightarrow \infty$$

for some non-degenerate distribution G independent of covariate θ , where α , β , μ and σ are smooth functions of θ . For the current application, this would seem to be a reasonable assumption (from physical considerations and inspection of model diagnostics, e.g., Figures 9 and 10 in the Appendix).

Many design recipes specify a combination such as 100-year wave, 100-wind and 10-year current to provide a conservative structural load for design purposes. The veracity of this assumption is difficult to assess in general. However, using the estimated conditional extremes model, simulations of extreme environments, structural loads and responses can be performed efficiently. In principle therefore, combinations of environmental design conditions, incorporating covariate effects, corresponding to a specified level of risk or reliability can be estimated.

The conditional extremes model is applicable to the joint tail of any distribution of random variables, for example representing peaks over threshold or block (e.g. annual) maxima. It is also applicable to joint modelling of serially correlated variables characterising consecutive sea states, with care in interpretation

of inferences for parameter uncertainty and return values in the presence of serial dependence.

The adoption of quantile regression for threshold modelling with covariates together with subsequent marginal and conditional extremes modelling provides a rational and scalable framework for joint tail modelling. In the current work, directional thresholds corresponding to a common non exceedance probability are used for marginal and dependence modelling. In other applications, adopting different non exceedance probabilities for each marginal and dependence model has been found, based on inspection of model fit diagnostics, to be beneficial. In principle, other criteria for threshold selection could be used, but we find the use of constant non exceedance probability with covariate to be physically appealing and practically useful.

Incorporation of multiple (and multivariate) covariates is possible in principle, and may be justifiable and necessary in future. Computationally, inference for one covariate is relatively straightforward, notwithstanding the need to estimate multiple quantile thresholds, appropriate parameter roughnesses in marginal and conditional models (using cross-validation), and uncertainties (using bootstrapping). Bondell et al. (2010) propose simultaneous estimation of non-crossing quantiles. Extensions of the method to multiple covariates will require access to good computational resources. Extension to conditional modelling of three or more random variables, and potentially even to include different covariates for different subsets of variables are possible. Spatial, spatio-directional or spatio-temporal covariates are attractive from an oceanographic perspective.

The size of estimated uncertainties of conditional extremes model parameters is influenced by many factors, including the form of the conditional extremes model and sample size. For some values of parameters (e.g. $\alpha_j(\theta)$ and $\beta_j(\theta)$ near unity, see Section 3.4), model identification is problematic. Small samples typically yield large uncertainties. The effect of parameter uncertainty on estimated design values can be easily quantified by including the estimation of design values within the bootstrapping analysis. Bayesian specification might prove advantageous especially when prior judgements regarding realistic smoothness of parameters with covariates are possible from physical understanding.

Acknowledgement

The authors acknowledge discussions with Yanyun Wu at Shell, Jonathan Tawn of Lancaster University, UK, and useful comments from two anonymous reviewers.

Appendix

This appendix gives supporting diagnostic plots for the conditional extremes model for associated T_P given storm peak H_S discussed in Section 4 of the main text.

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

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.
- H. D. Bondell, B. J. Reich, and H. Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97:825–838, 2010.
- V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207, 2005.
- A.C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393, 1990.
- 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.
- 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.
- S. Haver. Wave climate off northern Norway. *Applied Ocean Research*, 7:85–92, 1985.
- J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17:537–571, 2007.
- J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497, 2004.

- P. Jonathan and K. C. Ewans. Modelling the seasonality of extreme waves in the Gulf of Mexico. *ASME J. Offshore Mech. Arct. Eng.*, 133:021104, 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, J. Flynn, and K. C. Ewans. Joint modelling of wave spectral parameters for extreme sea states. In *Proc. 11th International Workshop on Wave Hindcasting and Forecasting, Halifax, Nova Scotia, Canada*, 2009.
- P. Jonathan, J. Flynn, and K. C. Ewans. Joint modelling of wave spectral parameters for extreme sea states. *Ocean Eng.*, 37:1070–1080, 2010.
- C. Keef, I. Papastathopoulos, and J. A. Tawn. Estimation of the conditional distribution of a vector variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Mult. Anal.*, 115:396–404, 2013.
- R. Koenker. *Quantile regression*. Cambridge University Press, 2005.
- J. Kyselý, J. Pícek, and R. Beranová. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72:55–68, 2010.
- P. Northrop and P. Jonathan. Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809, 2011.
- C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10:33–60, 2012.

List of Figures

1	Figure 1: North Sea location. Directional sectors corresponding to long fetches associated with the Atlantic Ocean, Norwegian Sea and North Sea typically yield more severe storm events. Sectors corresponding to Norway and the United Kingdom are fetch limited. Storm direction is the direction from which the storm emanates, and is measured clockwise from North.	25
2	Plots of storm peak H_S (in metres, horizontal) versus associated T_P (in seconds, vertical) for the most severe (20% of) storms emanating from 6 directional sectors (ordered, clockwise from 20°). The characteristics of dependence between T_P and H_S varies from sector to sector. For example, for storms emanating from the south (directional sector $[140, 200)$), T_P is highly dependent on H_S in contrast to Atlantic storms (from directional sector $[230, 280)$).	26
3	Polar plot of marginal directional quantile estimates for storm peak H_S (in metres, on left hand side) and T_P (seconds) for deciles with probabilities 0.1 to 0.8 together with the sample (grey dots). Transformed directions were used for quantile regression. Directional dependence is apparent.	27
4	Estimates for parameters α , β , μ and σ and their uncertainties as functions of θ using penalised likelihood with bootstrap resampling. Median bootstrap estimate is given in solid black, with a 95% bootstrap uncertainty band in dashed black. The estimate obtained using the original sample is given in solid grey. Dependence term α is largest for storms emanating from the North Sea sector as expected from inspection of sample (see Figure 2). Directional effects are apparent in α and σ in particular. The bootstrap uncertainty in β is particularly large. Note that parameter uncertainty is also large in the interval $[0, 135)$ (not shown), due to small sample size.	28
5	Marginal storm peak T_P (seconds) corresponding to non-exceedance probability of 0.99 (in 34 years), incorporating covariate effects (solid light grey). Median conditional storm peak T_P given exceedances of storm peak H_S (metres) with exceedance probability 0.01, incorporating (solid black) and ignoring (solid dark grey) covariate effects, together with 2.5%- and 97.5% percentiles (dashed).	29
6	Return values of storm peak H_S (metres) and associated conditional values of T_P (seconds). Inner dot-dashed lines (on common scale): storm peak H_S with probability non-exceedance probability 0.99 (in 34 years), with (grey) and without (black) directional effects. Outer solid lines (on common scale): median associated T_P with (black) and without (grey) directional covariate effects; outer dashed lines give corresponding 2.5%- and 97.5% percentile values for associated T_P	30
7	Simple ratio of response R to storm peak H_S as a function of storm peak T_P (seconds) for any direction.	31
8	Median conditional structural response (in metres) for storm peak H_S (metres) exceeding its directional quantile non-exceedance probability of 0.99 incorporating (solid black) and ignoring (solid grey) directional covariate effect, with corresponding 2.5%- and 97.5% percentiles (dashed).	32
9	Values for residuals Z from conditional extremes model against direction θ . There is no obvious directional dependence.	33
10	Values for residuals Z from conditional extremes model against conditioning variate H_S (on Gumbel scale). There is no obvious dependence.	34

11	Values for residuals Z from a covariate-free conditional extremes model against direction θ . Residuals appear to be dependent on the value of covariate.	35
----	---	----

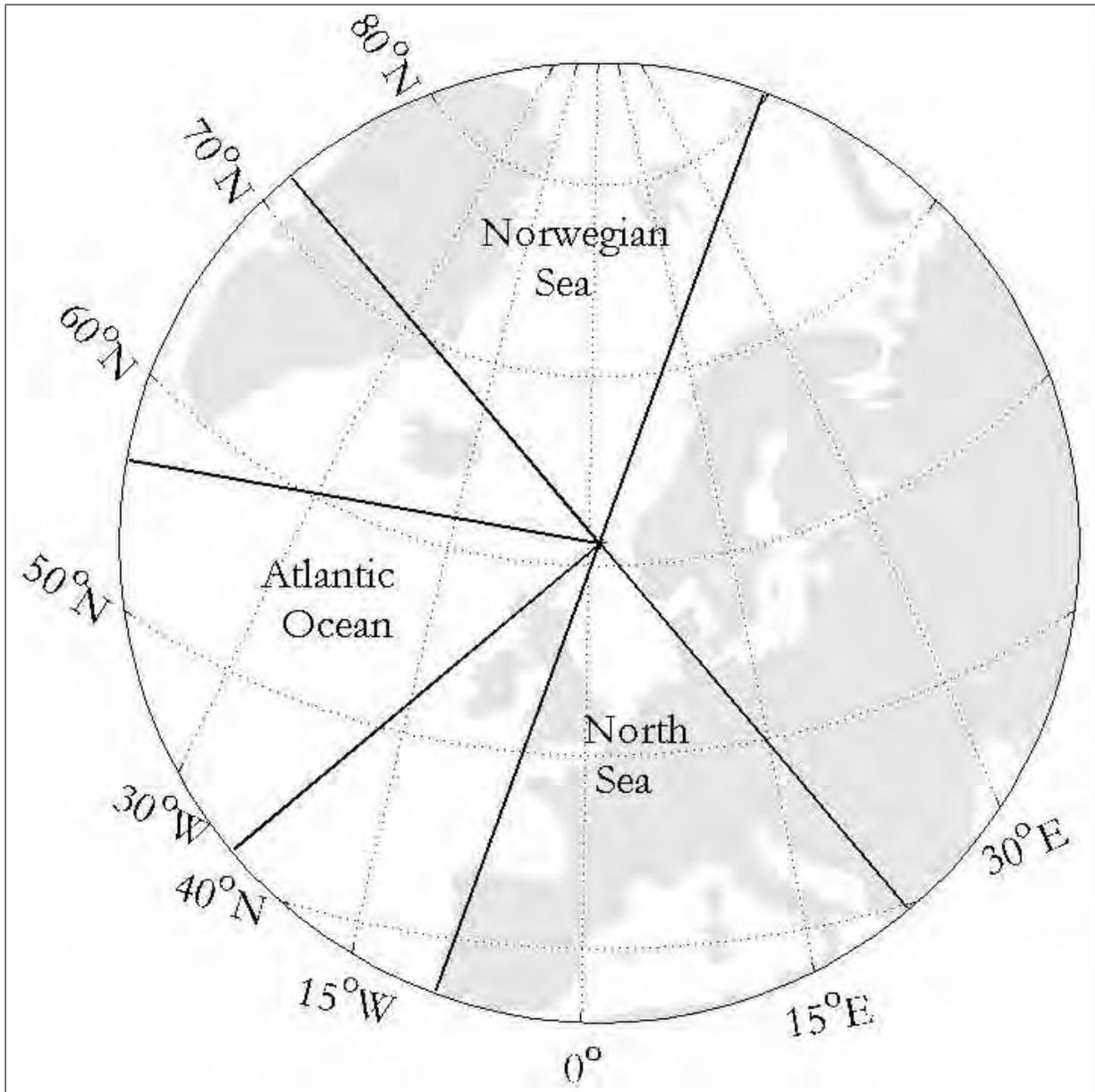


Figure 1: Figure 1: North Sea location. Directional sectors corresponding to long fetches associated with the Atlantic Ocean, Norwegian Sea and North Sea typically yield more severe storm events. Sectors corresponding to Norway and the United Kingdom are fetch limited. Storm direction is the direction from which the storm emanates, and is measured clockwise from North.

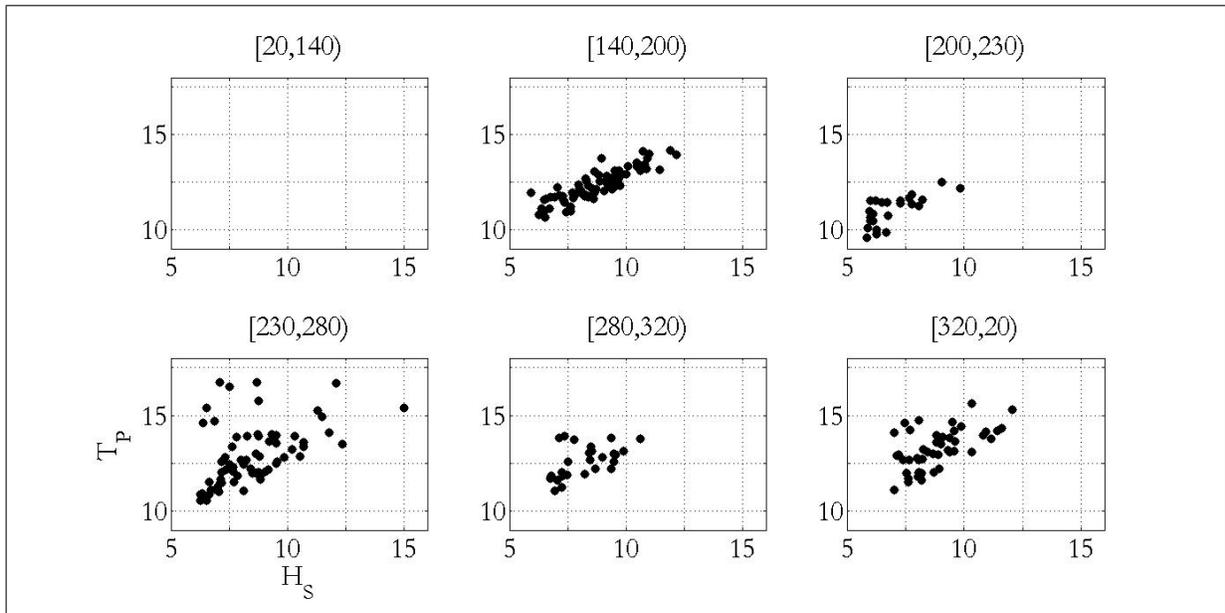


Figure 2: Plots of storm peak H_S (in metres, horizontal) versus associated T_P (in seconds, vertical) for the most severe (20% of) storms emanating from 6 directional sectors (ordered, clockwise from 20°). The characteristics of dependence between T_P and H_S varies from sector to sector. For example, for storms emanating from the south (directional sector $[140, 200)$), T_P is highly dependent on H_S in contrast to Atlantic storms (from directional sector $[230, 280)$).

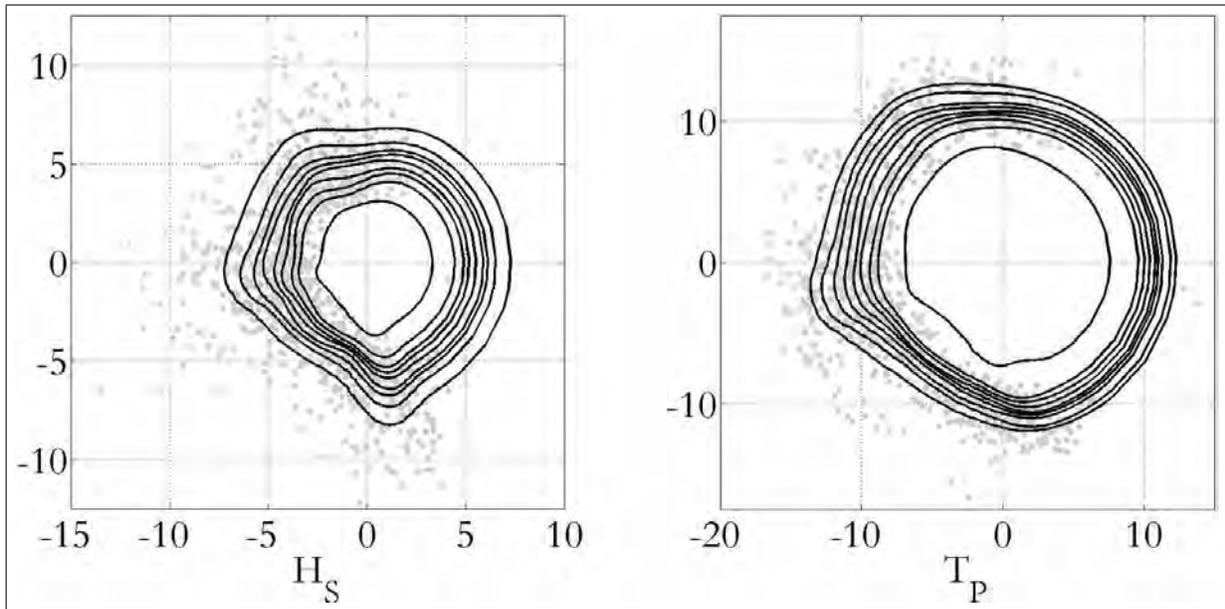


Figure 3: Polar plot of marginal directional quantile estimates for storm peak H_S (in metres, on left hand side) and T_P (seconds) for deciles with probabilities 0.1 to 0.8 together with the sample (grey dots). Transformed directions were used for quantile regression. Directional dependence is apparent.

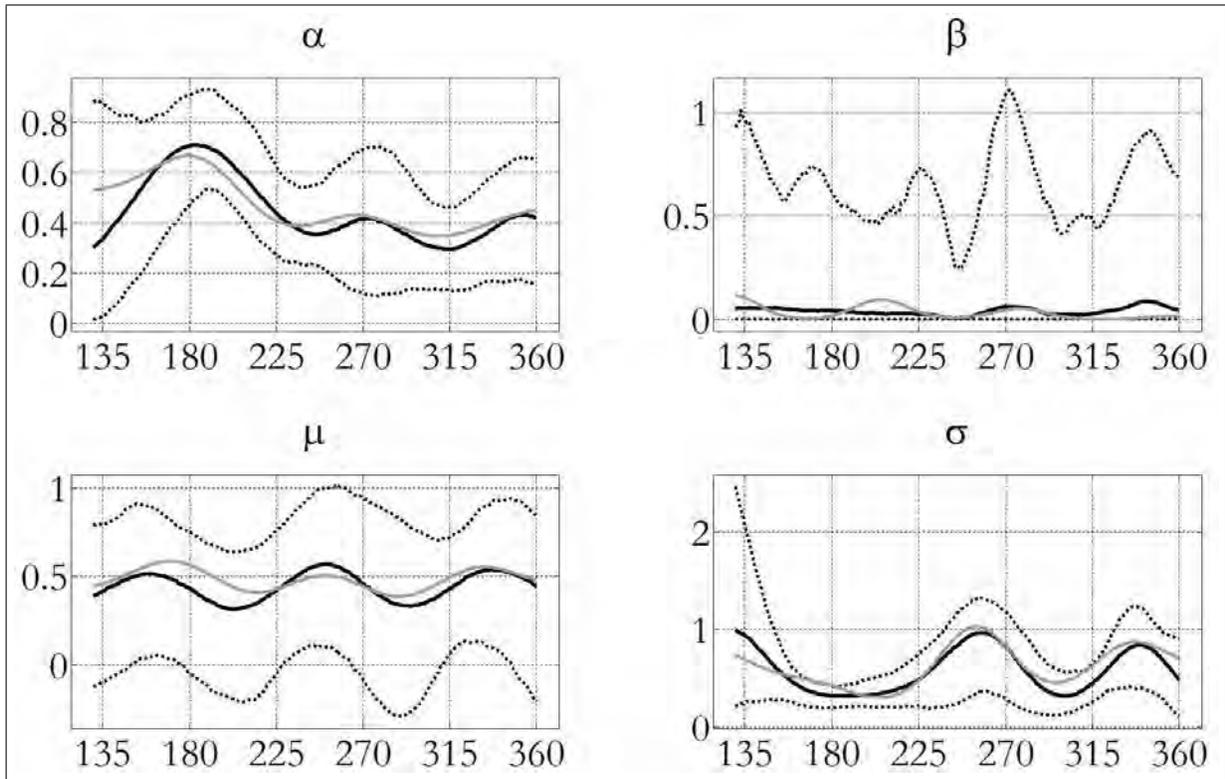


Figure 4: Estimates for parameters α , β , μ and σ and their uncertainties as functions of θ using penalised likelihood with bootstrap resampling. Median bootstrap estimate is given in solid black, with a 95% bootstrap uncertainty band in dashed black. The estimate obtained using the original sample is given in solid grey. Dependence term α is largest for storms emanating from the North Sea sector as expected from inspection of sample (see Figure 2). Directional effects are apparent in α and σ in particular. The bootstrap uncertainty in β is particularly large. Note that parameter uncertainty is also large in the interval $[0,135)$ (not shown), due to small sample size.

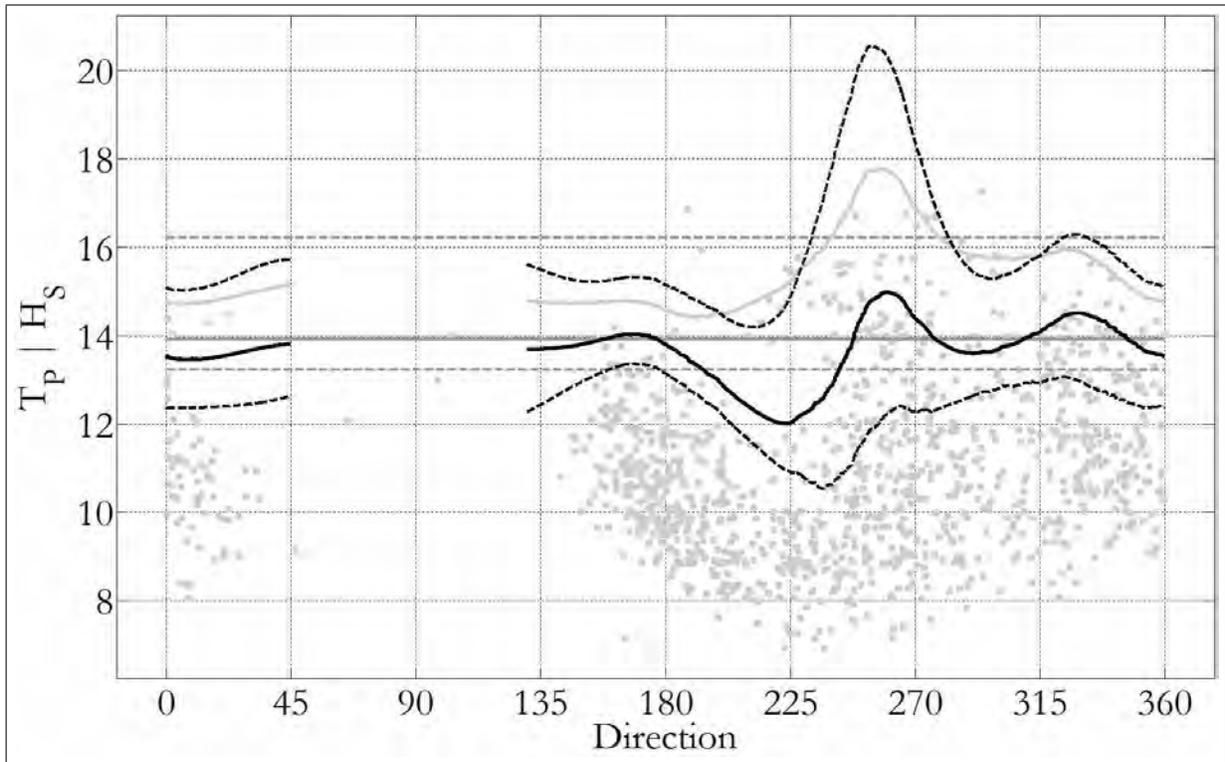


Figure 5: Marginal storm peak T_P (seconds) corresponding to non-exceedance probability of 0.99 (in 34 years), incorporating covariate effects (solid light grey). Median conditional storm peak T_P given exceedances of storm peak H_S (metres) with exceedance probability 0.01, incorporating (solid black) and ignoring (solid dark grey) covariate effects, together with 2.5%- and 97.5% percentiles (dashed).

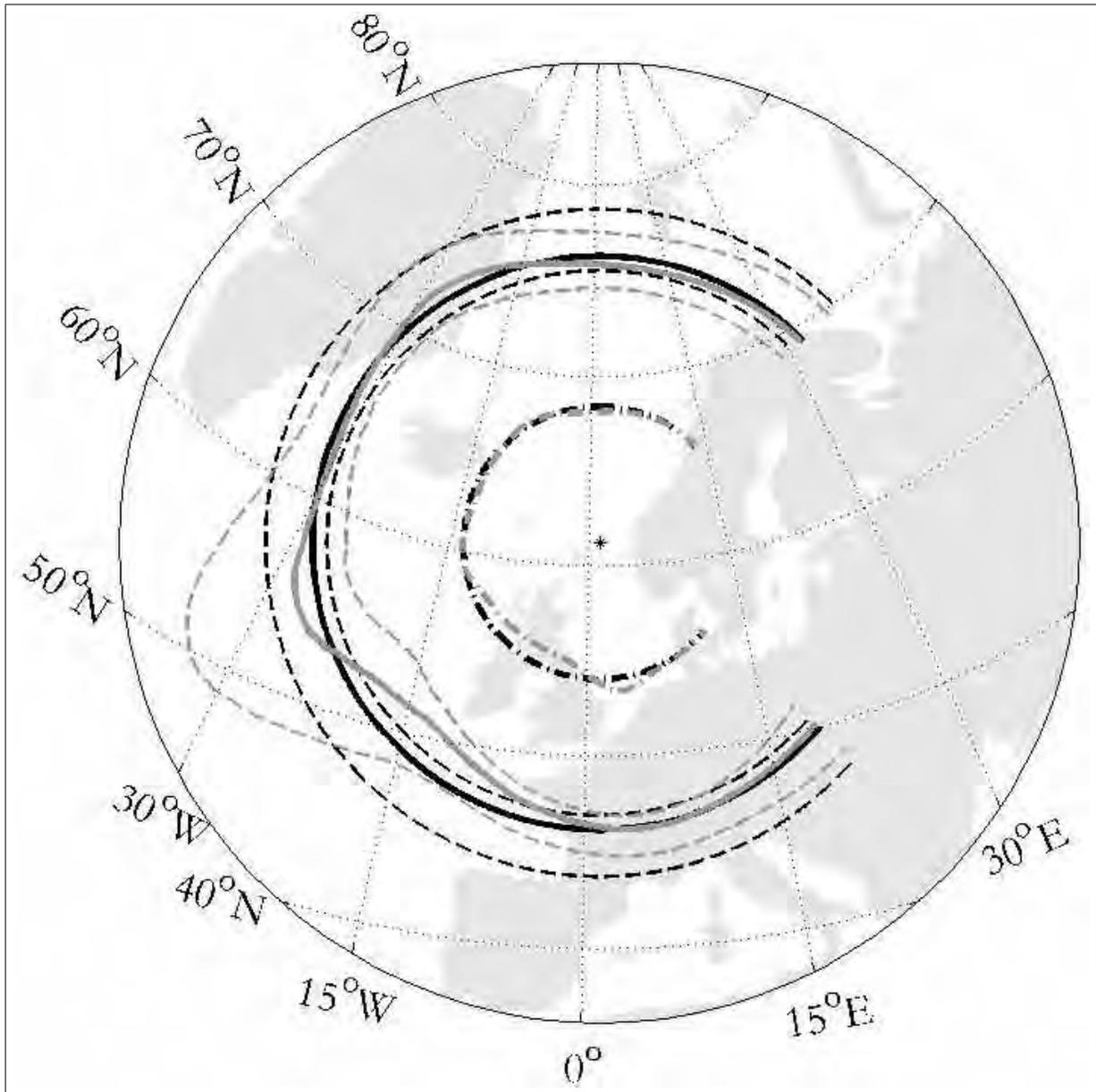


Figure 6: Return values of storm peak H_S (metres) and associated conditional values of T_P (seconds). Inner dot-dashed lines (on common scale): storm peak H_S with probability non-exceedance probability 0.99 (in 34 years), with (grey) and without (black) directional effects. Outer solid lines (on common scale): median associated T_P with (black) and without (grey) directional covariate effects; outer dashed lines give corresponding 2.5%- and 97.5% percentile values for associated T_P .

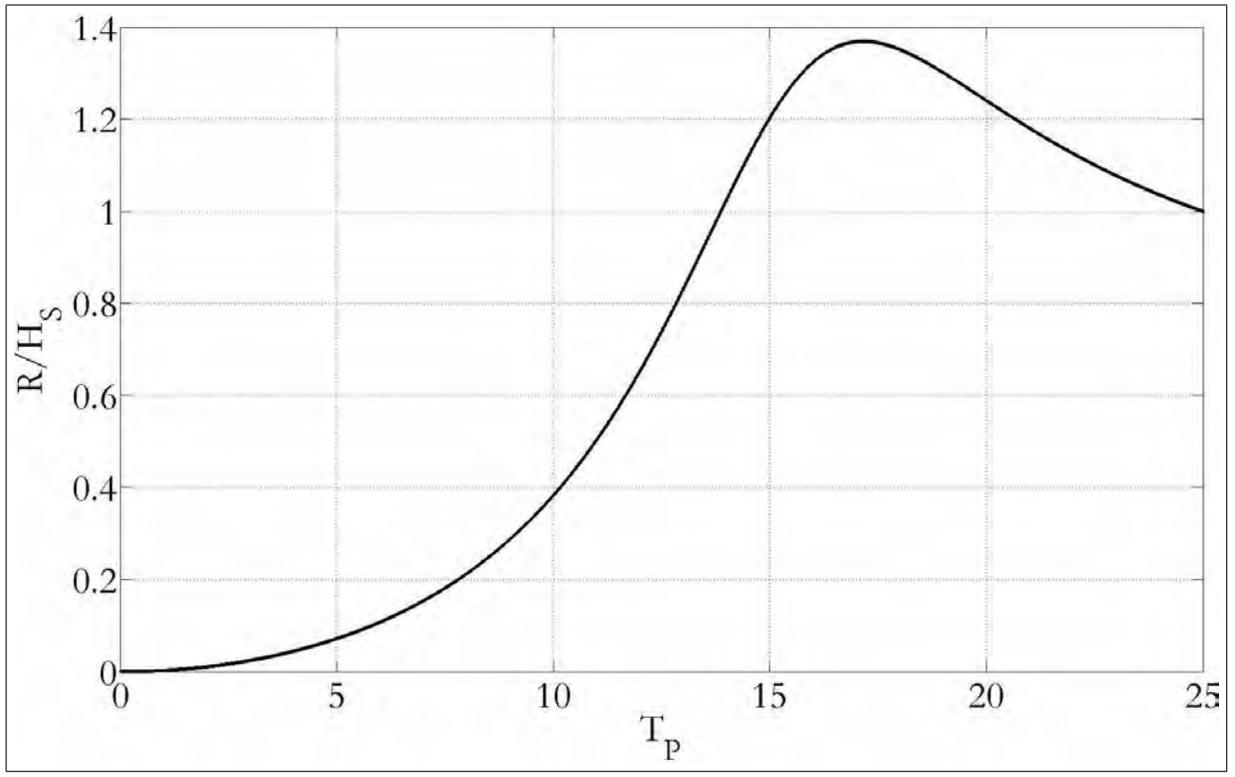


Figure 7: Simple ratio of response R to storm peak H_S as a function of storm peak T_P (seconds) for any direction.

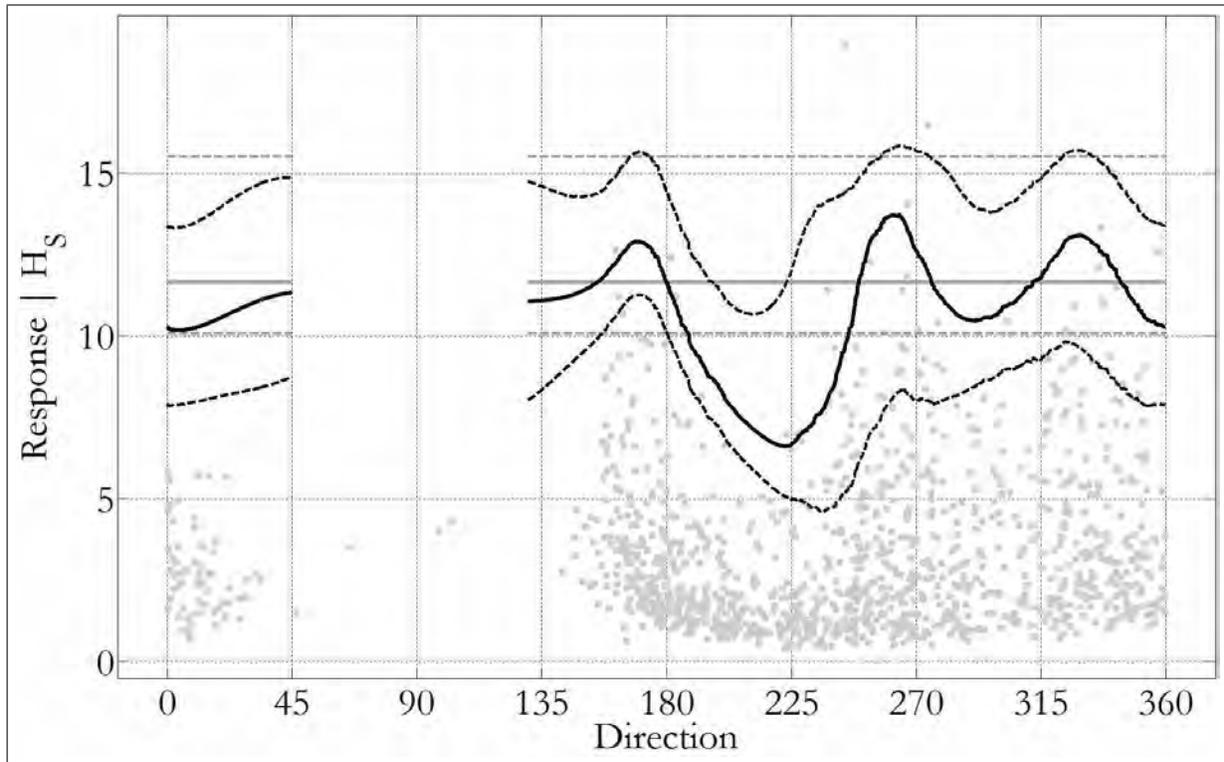


Figure 8: Median conditional structural response (in metres) for storm peak H_S (metres) exceeding its directional quantile non-exceedance probability of 0.99 incorporating (solid black) and ignoring (solid grey) directional covariate effect, with corresponding 2.5%- and 97.5% percentiles (dashed).

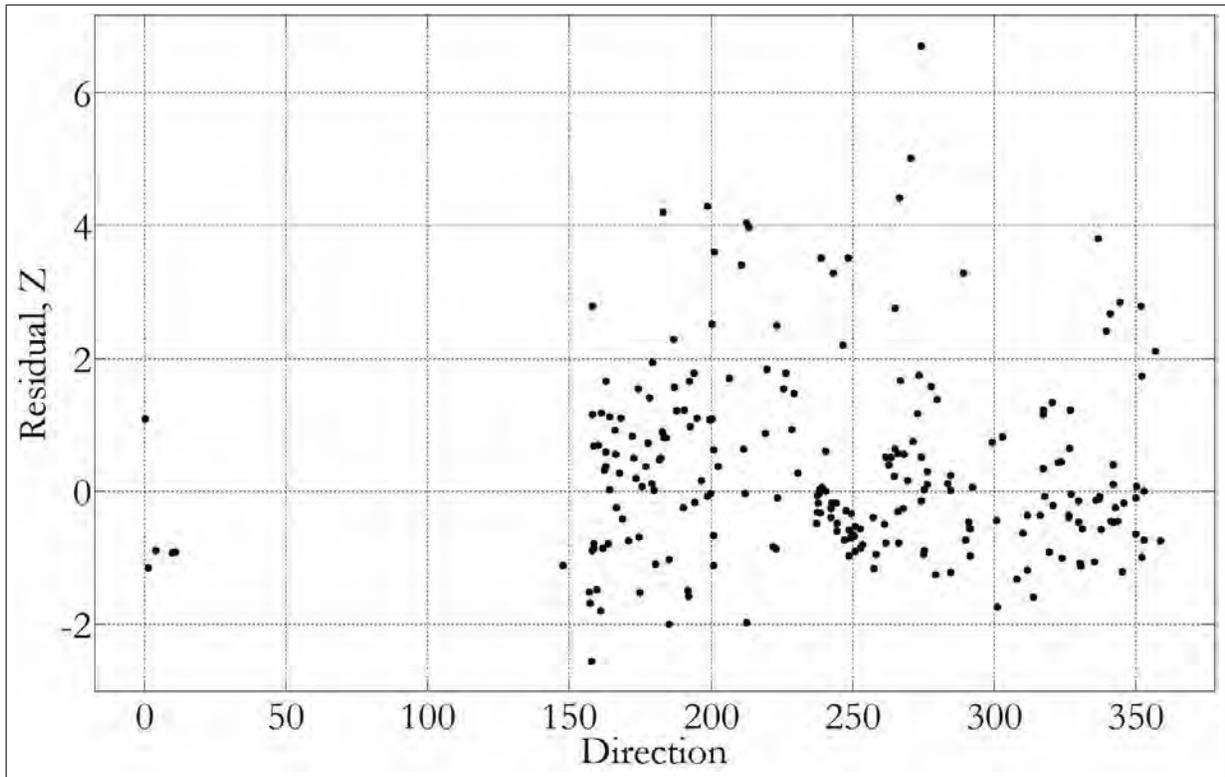


Figure 9: Values for residuals Z from conditional extremes model against direction θ . There is no obvious directional dependence.

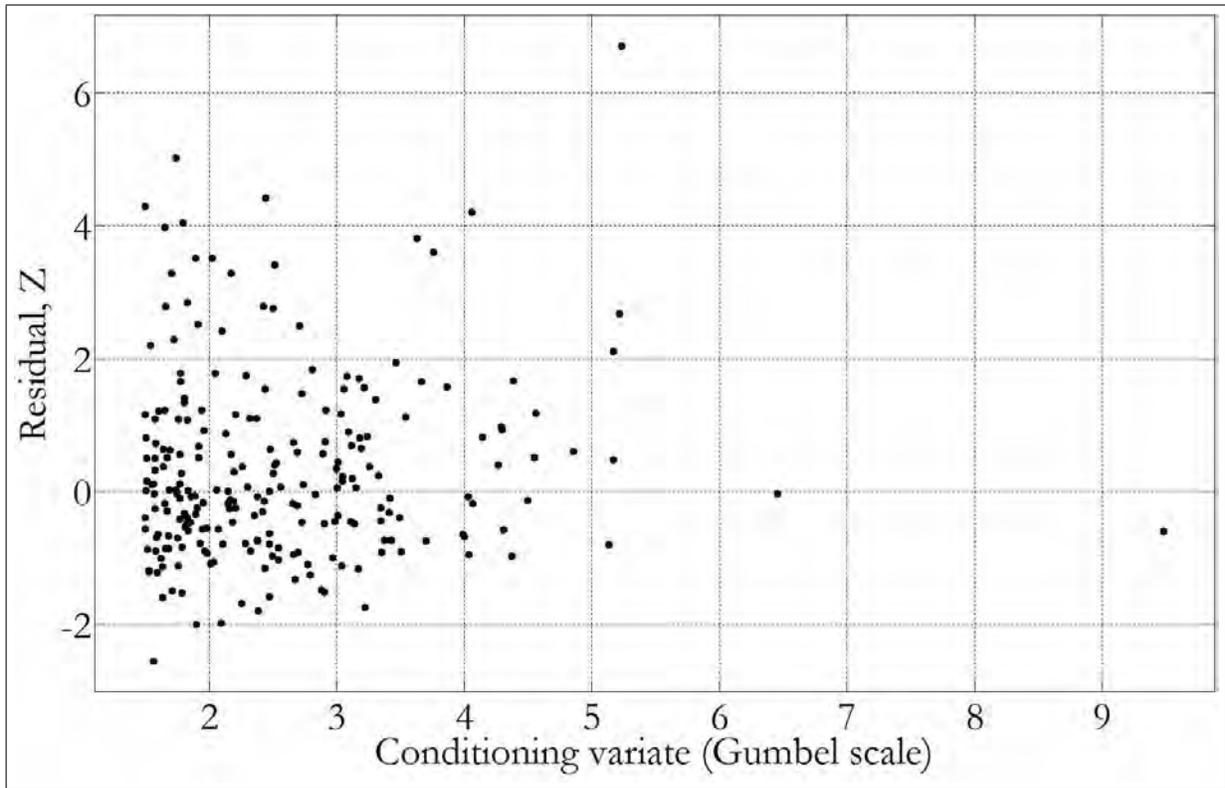


Figure 10: Values for residuals Z from conditional extremes model against conditioning variate H_S (on Gumbel scale). There is no obvious dependence.

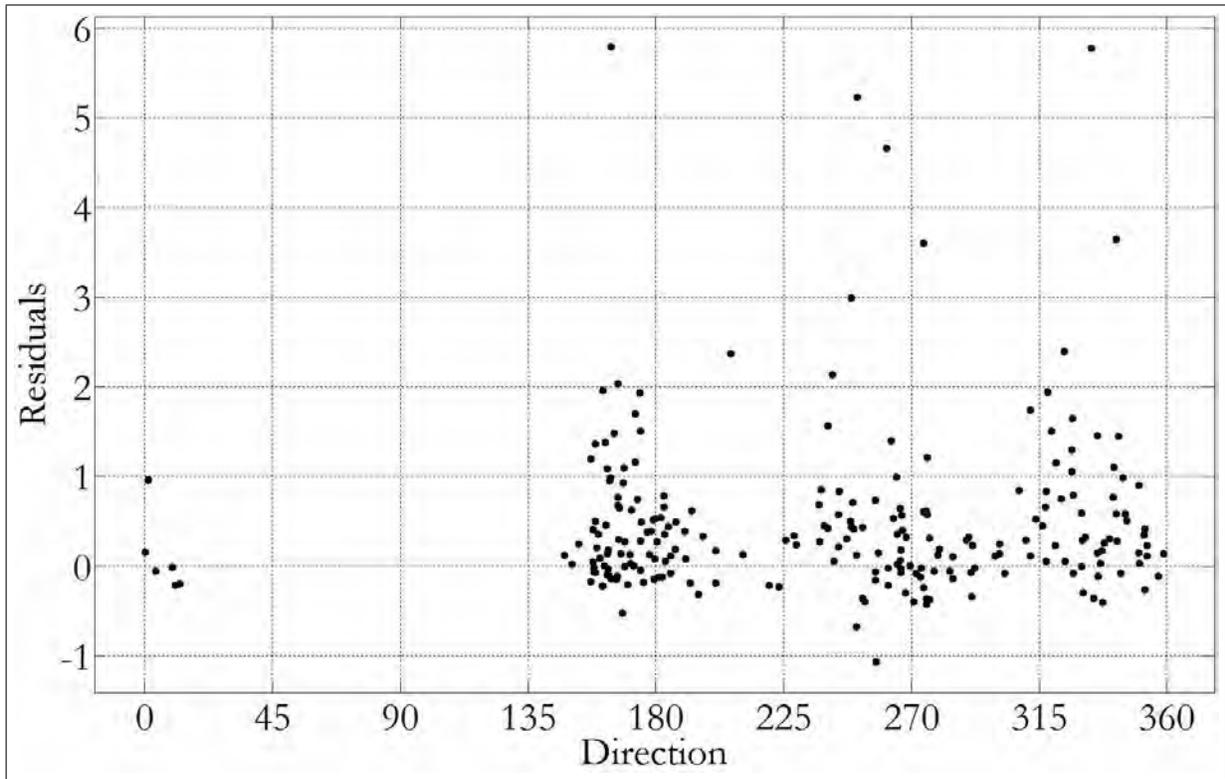


Figure 11: Values for residuals Z from a covariate-free conditional extremes model against direction θ . Residuals appear to be dependent on the value of covariate.