Proceedings of The 35th International Conference on Ocean, Offshore and Arctic Engineering
OMAE2016
June 19-24, 2016, Busan, South Korea

OMAE2016-54355 (DRAFT)

NON-STATIONARY ESTIMATION OF JOINT DESIGN CRITERIA WITH A MULTIVARIATE CONDITIONAL EXTREMES APPROACH

Laks Raghupathi*
Shell India Markets Pvt. Ltd.,
Bangalore, 560048, India.
l.raghupathi@shell.com

David Randell
Shell Global Solutions (UK),
Manchester, M22 0RR, UK.
david.randell@shell.com

Kevin Ewans
Sarawak Shell Bhd.,
50450 Kuala Lumpur, Malaysia.
kevin.ewans@shell.com

Philip Jonathan
Shell Global Solutions (UK),
Manchester, M22 0RR, UK.
philip.jonathan@shell.com

ABSTRACT

Understanding the interaction of ocean environments with fixed and floating structures is critical to the design of offshore and coastal facilities. Structural response to environmental loading is typically the combined effect of *multiple* environmental parameters over a period of time. Knowledge of the tails of marginal and *joint* distributions of these parameters (e.g. storm peak significant wave height and associated current) as a function of covariates (e.g. dominant wave and current directions) is central to the estimation of extreme structural response, and hence of structural reliability and safety. In this paper, we present a framework for the joint estimation of multivariate extremal dependencies with multi-dimensional covariates. We demonstrate proof of principle with a synthetic bi-variate example with two covariates quantified by rigorous uncertainty analysis. We further substantiate it using two practical applications (associated current given significant wave height for northern North Sea and joint current profile for offshore Brazil locations). Further applications include the estimation of associated criteria for response-based design (e.g., T_P given H_S), extreme current profiles with depth for mooring and riser loading, weathervaning systems with non-stationary effects for the design of FLNG/FPSO installations, etc.

1 Introduction

Metocean engineers are often tasked with estimating joint extremes. This task in general is framed as the estimation of associated values of a set of environmental variables given the occurrence of large values of a dominant environmental or structural response variable. Typical examples might be the specification of associated peak period corresponding to a significant wave height with given return period, or the specification of significant wave height and spectral peak period corresponding to a large value of a particular structural response. This is central to the estimation of the statistics of extreme structural response, and hence of structural reliability and safety. Thus characterising the joint structure of extremes of environmental variables is 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 to avoiding these particular restrictions. Ewans and Jonathan [2014] discusses the Heffernan and Tawn model for the specification of joint extremes in the offshore industry in the context of other approaches. Broadly this framework applies to estimating multivariate extremal dependence in the *stationary*

*Address all correspondence to this author.

case, i.e. in the absence of covariates. Extremal marginal and dependence characteristics of environmental variables typically vary with covariates (such as wave direction). Therefore reliable descriptions of extreme environments should also therefore characterise any non-stationarity.

The directional variability of extreme environments is important to estimate, particularly for complex responses of floating structures, but also for the design of safety-critical top-side facilities on fixed structures (see, for example, Tromans and Vanderschuren [1995], Winterstein et al. [1993]). Estimation of joint extremal behaviour is of considerable interest in ocean engineering in particular (see, for example Jonathan and Ewans [2013]). Jonathan et al. [2010] applies the conditional extremes model to storm peak significant wave height and associated spectral peak period for hindcast samples from different ocean basins. Jonathan et al. [2012] considers the joint modelling of directional ocean currents. Jonathan et al. [2013] extends the conditional extremes model of Heffernan and Tawn to the *non-stationary* case, i.e. to include covariate effects, using Fourier representations of model parameters for single periodic covariates, extended by [Jonathan et al., 2014] to include general-purpose spline representations for model parameters as functions of multidimensional covariates, common to all inference steps. The current work further extends the bivariate non-stationary conditional extremes work with one-dimensional covariate to a framework allowing two-dimensional covariates (extendible to *multi-dimensional* covariates and to *multivariate* responses) quantified by rigorous uncertainty analysis. To handle the increased computational burden in the multi-dimensional covariate space, we exploit fast techniques already presented in [Raghupathi et al., 2016] which estimated whole-basin independent criteria for a large spatial neighbourhood in the Gulf of Mexico (GoM) accounting for spatial and storm directional variability of peaks over threshold.

The approach to modelling is as follows. We use a non-crossing quantile regression to estimate appropriate non-stationary marginal quantiles simultaneously 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. Then we perform marginal extreme value and conditional extremes modelling within a roughness-penalised likelihood framework, with cross-validation to estimate suitable model parameter roughness. Finally, we use a bootstrap re-sampling procedure, encompassing all inference steps, to quantify uncertainties in, and dependence structure of, parameter estimates and estimates of conditional extremes of one variate given large values of another. We validate the approach using simulations from known joint distributions, the extremal dependence structures of which change with covariate.

The rest of the paper is organised as follows. §2 presents the modelling steps in detail explaining the 2D case formulation as well as the extensions to higher-dimensions. §3 describes the basis of the validation strategy through rigorous uncertainty quantification with the help of a synthetic example with known dependence characteristics. In §4 and 5, we present the results of applying the method to

1. storm peak significant wave height and associated currents for extra-tropical storms at a northern North Sea (NNS) location, with storm direction and current direction as covariates
2. current profile with current direction at different depths given an extreme surface current at an offshore Brazil (OB) location.

Finally, in §6 we summarise possible extensions and recommendations for future work.

2 Non-Stationary Conditional Extremes (NSCE) Method

2.1 Outline of bivariate NSCE with $2D$ covariate

Consider two random variables $\dot{X}_1$ and $\dot{X}_2$ (representing oceanographic parameters such as significant wave height and current speed), respectively functions of the $1D$ covariates θ_1 and θ_2. We want to characterise the joint extremal behaviour of $\dot{X}_1$ and $\dot{X}_2$ as a function of θ_1 and θ_2. Briefly, we want to estimate $\dot{X}_1, \dot{X}_2 | \theta_1, \theta_2$ for large values of at least one of $\dot{X}_1$ and $\dot{X}_2$. We achieve this by estimating of conditional extremes model for $\dot{X}_2 | \dot{X}_1, \theta_1, \theta_2$. The steps needed to estimate these conditional extremes models and simulate under them follow the Heffernan amd Tawn approach for stationary joint extremes can be summarised as follows.

1. Estimate marginal extreme value model $\dot{X}_1 | \theta_1$,
2. Estimate marginal extreme value model $\dot{X}_2 | \theta_2$,
3. Transform $\dot{X}_1$ and $\dot{X}_2$ to random variables X_1 and X_2 respectively with stationary standard Gumbel distributions,
4. Estimate conditional extremes models $X_2 | X_1, \theta_1, \theta_2$
5. Using conditional extremes model, simulate realisations of covariates and variates corresponding to a specified return period of interest on Gumbel scale,
6. Transform the realisations from Gumbel to original (physical) scales to get the estimates of associated design criterion for a given return period

Threshold selection for marginal and dependence modelling is a key concern, and can be quite challenging. Estimating models for different threshold levels and then averaging over models for different thresholds to incorporate uncertainty in threshold selection is a wise procedure. In future, Bayesian implementation of the conditional extremes model will allow the most natural incorporation of threshold uncertainty.

2.2 Extension to p-variate NSCE with nD covariate

We now characterise parameter non-stationarity in marginal and dependence models with respect to multi-dimensional covariates in terms of tensor products of splines. Given sufficient computational resource, and slick spline algorithms such as GLAMS

(Currie et al. 2006 as applied in [Raghupathi et al., 2016] for the marginal case), the bivariate conditional extremes approach with $2D$ covariate is therefore straightforward to extend to nD covariates. Moreover, since the dependence model for p ($p > 2$) responses (i.e. $\dot{X}_1, \dot{X}_2, ..., \dot{X}_p$) can be written in terms of a set of pairwise dependence models, the approach is also extendible to multivariate responses. A p-variate conditional extremes model with nD covariate can therefore be constructed. The dependence model for p ($p > 2$) responses $X_1, X_2, ..., X_p$ on Gumbel scale can be expressed in terms of the set of pairwise dependence models $X_j|X_k$ for $j = 1, 2, ...p$ and $k = 1, 2, ...p$. For positively dependent random variables X_j, X_k with standard Gumbel marginal distributions for any fixed value t_j of θ_j and t_k of θ_k (or $t_{jk} = t_j \cup t_k$ of $\theta_j \cup \theta_k$), we extend the asymptotic argument of Heffernan and Tawn [2004] for the form of the conditional distribution of X_j given large values of X_k. We assume

$$(X_j|X_k = x_k, \theta_j \cup \theta_k = t_{jk}) = \alpha_{jk}(t_{jk})x_k + x_k^{\beta_{jk}(t_{jk})} W_{jk}(t_{jk}) \quad (1)$$

for $x_k > \psi_k(t_k)$, a threshold non-stationary with respect to θ_k only, with non-exceedance probability τ_k above which the conditional extremes model fits well. The parameter functions $\alpha_{jk} \in [0,1]$, $\beta_{jk} \in (-\infty, 1]$ vary smoothly with covariates $\theta_j \cup \theta_k$. W_{jk} is a random variable drawn from an unknown distribution, the characteristics of which also vary smoothly with $\theta_j \cup \theta_k$. We assume that the standardised variable $Z_{jk} = (W_{jk} - \mu_{jk})/\sigma_{jk}$ follows a common distribution G_{jk}, independent of covariates, for smooth location and scale parameter functions μ_{jk} and σ_{jk} of $\theta_j \cup \theta_k$, with $\sigma_{jk} > 0$. We rewrite (1) for any value t_{jk} of $\theta_j \cup \theta_k$:

$$\begin{aligned}(X_j|X_k = x_k, \theta_j \cup \theta_k = t_{jk}) &= \alpha_{jk}(t_{jk})x_k \\ &+ x_k^{\beta_{jk}(t_{jk})}(\mu_{jk}(t_{jk}) + \sigma_{jk}(t_{jk})Z_{jk})\end{aligned} \quad (2)$$

For potentially negatively dependent variables (corresponding to $\alpha_{jk} = 0$ and $\beta_{jk} < 0$, see Heffernan and Tawn 2004), extended forms of the equations above are available in the stationary case, and can be easily specified for the non-stationary case also. However, for most applications of practical interest, and for oceanographic applications in particular, we expect $\alpha_{jk} > 0$ so that positive dependence only need be considered. In practice, were we to estimate α_{jk} to be approximately zero, the extended form of the conditional extremes model would then need to be considered. To estimate the parameter functions α_{jk}, β_{jk}, μ_{jk} and σ_{jk}, we follow Heffernan and Tawn [2004] in assuming that G_{jk} is the standard Gaussian distribution. The corresponding negative log likelihood for a sample of pairs (x_j^i, x_k^i) from the original sample with covariate t_{jk}^i for which $x_k^i > \psi_k(t_i^k)$

$$\ell_{CE,jk} = \sum_{i, x_k^i > \psi_k^i} \log s_{jk}^i + \frac{(x_j^i - m_{jk}^i)^2}{2(s_{jk}^i)^2} \quad (3)$$

where

$$\begin{aligned} m_{jk}^i &= \alpha_{jk}(t_{jk}^i)x_k^i + \mu_{jk}(t_{jk}^i)(x_k^i)^{\beta_{jk}(t_{jk}^i)} \\ s_{jk}^i &= \sigma_k^i(t_{jk}^i)(x_k^i)^{\beta_{jk}(t_{jk}^i)} \end{aligned}$$

Adopting a penalisation procedure to regulate parameter roughness, the penalised negative log likelihood is:

$$\ell_{CE,jk}^* = \ell_{CE,jk} + \lambda_{\alpha_{jk}} R_{\alpha_{jk}} + \lambda_{\beta_{jk}} R_{\beta_{jk}} + \lambda_{\mu_{jk}} R_{\mu_{jk}} + \lambda_{\sigma_{jk}} R_{\sigma_{jk}} \quad (4)$$

where parameter roughnesses $R_{\alpha_{jk}}$, $R_{\beta_{jk}}$, $R_{\mu_{jk}}$, $R_{\sigma_{jk}}$ are easily evaluated, and roughness coefficients $\lambda_{\alpha_{jk}}$, $\lambda_{\beta_{jk}}$, $\lambda_{\mu_{jk}}$, $\lambda_{\sigma_{jk}}$ are estimated using cross-validation. To reduce computational burden, we choose to fix the relative size of the roughness coefficients so that only one roughness coefficient λ_{jk} ($= \delta_{\alpha_{jk}}\lambda_{\alpha_{jk}} + \delta_{\beta_{jk}}\lambda_{\beta_{jk}} + \delta_{\mu_{jk}}\lambda_{\mu_{jk}} + \delta_{\sigma_{jk}}\lambda_{\sigma_{jk}}$) is estimated.

The values of $\delta_{\alpha_{jk}}, \delta_{\beta_{jk}}, \delta_{\mu_{jk}}$ and $\delta_{\sigma_{jk}}$ used are set by careful experimentation. Residuals:

$$r_{jk}^i = \frac{1}{\hat{\sigma}_{jk}(t_{jk}^i)}((x_j^i - \hat{\alpha}_{jk}(t_{jk}^i)x_k^i)(x_k^i)^{-\hat{\beta}_{jk}(t^i{}_{jk})} - \hat{\mu}_{jk}(t_{jk}^i)) \quad (5)$$

evaluated for $x_k^i > \psi_k(t_k^i)$ are inspected to confirm reasonable model fit, where all estimated parameter functions are evaluated at covariate values corresponding to x_k^i. The set of residuals is also used as a random sample of values for Z_{jk} from the unknown distribution G_{jk} for simulation to estimate extremes quantiles.

2.3 Parameter functional forms

Physical considerations usually suggest that parameters $\dot{\psi}_k$, $\dot{\xi}_k$ and $\dot{\zeta}_k$ (marginal threshold and GP model estimates) would be expected to vary smoothly with respect to covariates θ_k, and that α_{jk}, β_{jk}, μ_{jk} and σ_{jk} would vary smoothly with respect to $\theta_j \cup \theta_k$. Adopting the notation η for a typical parameter function of covariate vector θ, this can be achieved by expressing $\eta(\theta)$ in terms of an appropriate basis for the covariate domain. For a one-dimensional covariate, periodic on $[0, 360)$, we might adopt a basis of periodic B-splines of appropriate order. We calculate the B-spline basis matrix B ($m_s \times p_s$) for an index set of m_s (typically less than sample size, n) covariate values, at p_s uniformly spaced knot locations on $[0, 360)$. Specifically, for example in the case of a bi-directional covariate, we would define

B-spline bases B_{θ_1} $(m_{\theta_1} \times p_{\theta_1})$, B_{θ_2} $(m_{\theta_2} \times p_{\theta_2})$ for two covariate direction respectively, from which the full bi-directional basis B $(m_s \times p_s = m_{\theta_1} m_{\theta_2} \times p_{\theta_1} p_{\theta_2})$ is evaluated as

$$B = B_{\theta_1} \otimes B_{\theta_2} \tag{6}$$

where $\otimes$ represents the Kronecker product. Values η of $\eta(\theta)$ on the index set can then be expressed as $\eta = B\beta$ for some vector β $(p_s \times 1)$ of basis coefficients to be estimated.

The roughness R of η is evaluated on the index set using the approach of 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 $\Delta^g\beta = \Delta(\Delta^{g-1}\beta)$, $g > 1$, the roughness R of β is given by

$$R = \beta' P \beta \tag{7}$$

where $P = (\Delta^g)'(\Delta^g)$ for differencing at order g. We use $g = 1$ throughout this work.

2.4 Simulation procedure for pairwise dependence

Inferences concerning the probabilities of extreme sets involving $\dot{X}_j$ and $\dot{X}_k$, for some value or interval of covariates are drawn using the non-stationary conditional extremes approach by simulating joint occurrences using the estimated marginal and conditional extremes models. To simulate realisations of exceedances of a high quantile of the conditioning variate $\dot{X}_k$ and a corresponding value of the conditioned variate $\dot{X}_j$ corresponding to a pre-specified return period, for some interval I of covariates $\theta_j \cup \theta_k$, we proceed as follows:

1. Estimate the number of values to be simulated from the covariate interval I (for example, using the estimated Poisson rate ρ corresponding to I for the return period of interest).
2. For each individual to be simulated
   (a) Draw the necessary number of value of covariates t^s_{jk} at random from the interval I.
   (b) Draw a value of residual r^s_{jk} from the set of residuals obtained during model fitting,
   (c) Draw a value x^s_k of the conditioning variate from its standard Gumbel distribution,
   (d) If the value x^s_k exceeds ψ_k continue, else return to the start of 2.
   (e) Estimate the value of the conditioned variate x^s_j using:

$$x^s_j = \hat{\alpha}_{jk}(t^s_{jk}) x^s_k + (x^s_k)^{\hat{\beta}_{jk}(t^s_{jk})} (\hat{\mu}_{jk}(t^s_{jk}) + \hat{\sigma}_{jk}(t^s_{jk}) r^s_{jk}) \tag{8}$$

   where the obvious notation is used for the estimated values of model parameters, evaluated at covariate value t^s_{jk}.

3. Transform the pair x^s_k, x^s_j in turn to the original scale (to $\dot{x}^s_k, \dot{x}^s_j$) using the probability integral transform.

Note that this procedure can be extended to include realisations for which $x^s_k \leq \psi_k(t^s_k)$, by drawing a pair of values (for the conditioning and conditioned variates) at random from the subset of the Gumbel-transformed original sample within interval I (for which the conditioning variate is $\leq \psi_k$) at step 2d.

2.5 Simulation procedure for multivariate dependence

Extending the simulation procedure from §2.4 to $p > 2$ responses is straightforward, given the following two observations.

Firstly, the tail of the distribution of $X_{-k}|X_k$, where $X_{-k} = X_1, X_2, ..., X_{k-1}, X_{k+1}, ..., X_p$ can be estimated from the $p-1$ pairwise models $X_j|X_k$ $j \neq k$ provided that the residuals r^i_{jk} for each individual i (and all j) are sampled together during the simulation procedure. This retains the dependence structure of of $X_{-k}|X_k$ in the simulated realisations. Therefore, given that we can specify a conditioning variate X_k, we are able to simulate conditional extremes of all other variates.

Secondly, we can specify the conditioning variate X_k as follows. Suppose we are required to simulate realisations from an extreme set E of the joint distribution of $\dot{X}_1, \dot{X}_2, ..., \dot{X}_p$. We can always *partition* E into a set of (non-overlapping) subsets, such that in each subset one response $\dot{X}_k$ is more extreme in its marginal distribution that all other variables. For this subset, it is sensible to use $\dot{X}_k$ (or its Gumbel analogue X_k) as conditioning variate. Simulation from E therefore can be achieved by a set of simulations, one for each of the subsets of the partition of E. In this work, we focus discussion of method development on the bivariate case $p = 2$, but apply it to $p > 2$ also below.

3 Uncertainty quantification for synthetic data

We validate the non-stationary conditional extremes methodology by first applying it on a synthetic example before venturing into practical applications. First, we specify an underlying "true" non-stationary bivariate extreme value model whose characteristics are known. Using the true model, we simulate one or more samples of data and then fit an NSCE model. We then compare estimates of marginal and conditional cumulative distribution functions (cdfs) for arbitrary combinations of covariates and specified return periods by simulation under the "true" model, and simulation under the "fitted" NSCE model.

Uncertainty present in simulations under the true model is true, natural, inherent or *aleatory*. It cannot be reduced without changing the true underlying physics. Aleatory uncertainty is therefore the lowest level of uncertainty we can realistically hope to achieve in the fitted model also, and corresponds to the width or spread of the "true" cdf. In comparing cdfs under the true and fitted models, we therefore seek to assess that (a) the location of the "fitted" cdf is approximately the same as that of the truth, and

that (b) the width of the fitted cdf is at least as large as (but not too much larger than) that of the truth.

The extra uncertainty in cdfs under the fitted model is due to our lack of understanding of the truth as represented by the data sample used for fitting. This uncertainty is sometimes known as *epistemic* uncertainty, and can in principle be reduced by fitting using a larger sample. But the overall uncertainty in fitted cdfs can never be reduced below the level of aleatory uncertainty.

We choose to compare cdfs in the following way, to explore aleatory and epistemic uncertainty of marginal and dependence characteristics.

1. Generate many samples directly under the *true* model and estimate empirical cdfs. Estimates are labelled "*Truth*" or "*Data*". This represents the "*Aleotary (A)*" uncertainty described above.
2. For a particular sample generated under the true model, fit the NSCE model. Simulate under the fitted model to estimate empirical cdfs. Estimates are labelled "*1-EVA*" or "*Model*". This represents the "*Epistemic (E)*" uncertainty described above.
3. For each of a large number of samples generated under the true model, fit the NSCE model. Simulate under the fitted models to estimate empirical cdfs using all simulations. Estimates are labelled ''*M-EVA*" or "*M-Model*". This represents the combined "*Aleotary + Epistemic (A+E)*" uncertainty described above.

The number of comparisons which could be made is huge; we judge that those made is sufficient to demonstrate the quality of the NSCE approach. Examples of model validation for applications to synthetic and metocean data are given in forthcoming sections.

3.1 2D Synthetic Case

We now present the results of applying our non-stationary conditional extremes model using a synthetic example and validating using the diagnostics procedure described earlier. First, we describe the test dataset with known parameters (truth) used for model validation throughout this section. We then present the results of model validation and return value diagnostics for the marginal models and finally the conditional validation.

Figure 1 illustrates the generation of synthetic example used for the development and validation of the proposed approach. Figure 1 shows the plot of bi-variate Gaussian distributed variates X,Y with different correlation ($\alpha = 0.1, 0.3, 0.6, 0.9$) in different directional sectors θ_X, θ_Y. Notice the sector ($\theta_X > 180, \theta_Y > 180$) with high correlation specified with $\alpha = 0.9$. We have used the same rate of occurrence $\rho = 1$ for all the covariate sectors for simplicity. For sake of illustration from now onwards, we shall use practical metocean terminology in the place of generic variates X and Y, i.e., we shall be dealing with a common metocean variates H_S which is the significant wave height (measured in meters) and T_P which is the peak spectral period (measured in seconds). We would like to first marginally model H_S and T_P and eventually estimate the conditional distribution of T_P given H_S.

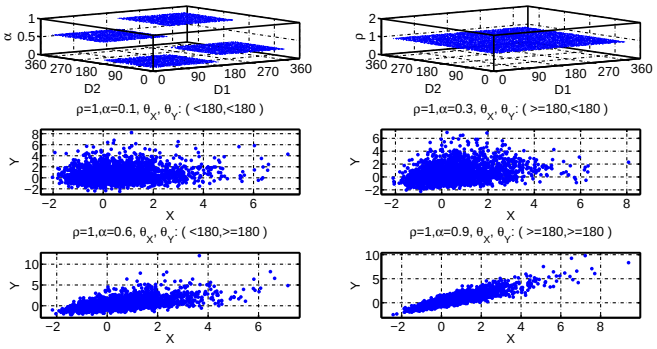


FIGURE 1. Synthetic Case: True parameters based on which the marginal and dependence data has been generated. Bi-variate Gaussian dependent data in Gumbel scale with different correlation parameters (α) for different directional covariate sectors with the same rate of occurrence (ρ).

3.2 Marginal Diagnostics

Figure 2 shows the plot of marginal parameter estimates of the threshold, rate of occurrence, GP shape and scale for the simulated case of H_S and T_P. This shows the results of marginally fitting a threshold (using non-crossing quantile regression), Poisson rate of occurrence and a GP shape and scale model. using the same synthetic example shown in Figure 1) with 1000 storm events as input for this analysis. For the threshold selection, we chose the threshold corresponding to a non-exceedance probability or NEP of 0.8 (Figure 2 (a) and (b)). The GP shape and scale model parameters are estimated by determining the optimal roughness penalty that minimizes lack of fit (Figure 2 (i) and (j)).

Synthetic Case Marginal Diagnostics: Having estimated the model parameters, we now validate its output by generating samples from the truth and model for a given return period and plotting omni-directional and sectoral distributions. Specifically, we simulate multiple realisations of sets of sea-state H_S events corresponding exactly to the period of the original sample under a model. For each set simulated, we construct a cumulative distribution function F, potentially restricted to some co-variate interval A, and estimate 2.5% and 97.5% values of the cdf for each value of H_S. To emphasise tail behaviour, we actually plot $1-\log(F)$. These curves (representing the *epistemic (E)* uncertainty), with 95% confidence intervals, are illustrated in black in

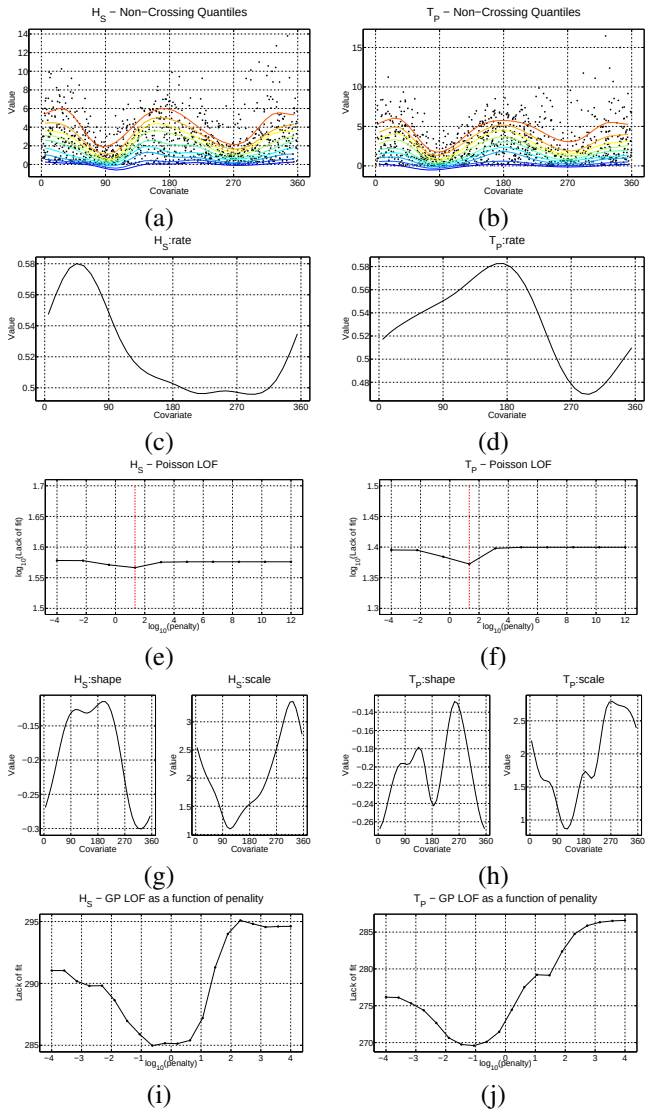


FIGURE 2. Synthetic Case: Marginal parameter estimates of H_S (left) and T_P (right). (a) & (b) Raw data with non-crossing quantile estimates. (c) & (d) Poisson rate of occurence estimates. (e) & (f) Cross-validatory selection of optimal roughness penalty of the Poisson model. (g) & (h) GP model shape and scale estimates. (i) & (j) Cross-validatory selection of optimal roughness penalty of the GP model.

Figure 3, are compared with the cumulative distribution function for the original sample, illustrated in red (representing the *aleotary (A)* uncertainty) in the Figure in terms of an empirical 95% bootstrap uncertainty band estimated from 1000 bootstrap resamples of the original sample. In addition, we simulate multiple realisations of sets of sea-state H_S events corresponding exactly to the period of the original sample under *many* models shown in green (representing the combined *aleotary and epistemic (A+E)* uncertainty). Agreement is good between all the three curves for omni-directional and sectoral estimates shown for H_S. Similar illustrations for T_P estimates per directional octant also indicate good agreement between the original sample and simulation.

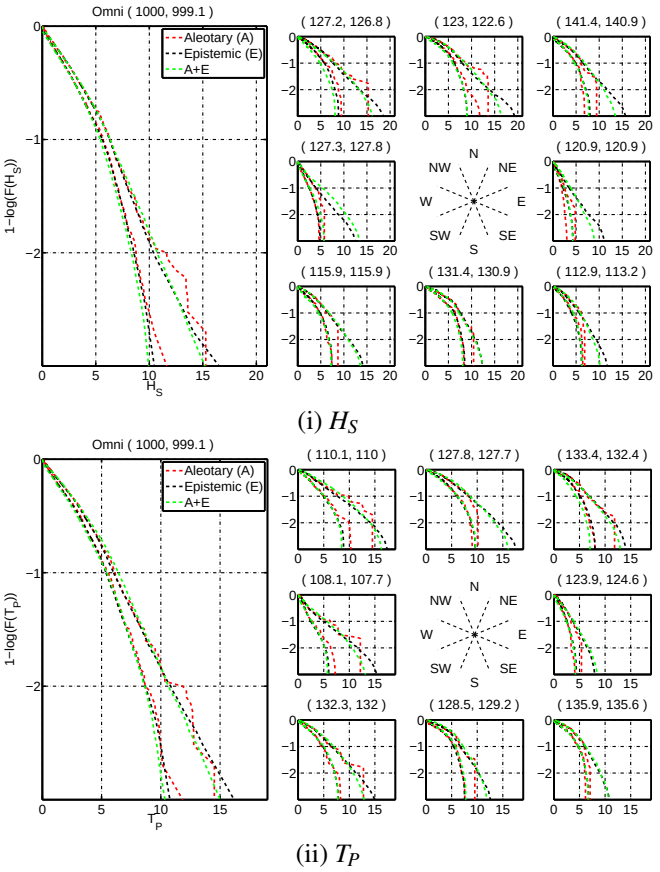


FIGURE 3. Synthetic Case: Validation of the marginal return value estimation by comparing cdfs of bootstrap resamples of the original data and realisations under the model(s) corresponding to the same time period as the original for (top) H_S and (bottom) T_P. The plots show $1-\log(F)$ to emphasise tail behaviour showing 95% quantiles for realisations of data (red), from a single model (black) and from many models (green). For each plot, the 8 right hand panels show the comparisons per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional comparison. The title for each plot, in brackets are the numbers of actual and simulated events for each directional sector.

3.3 Conditional Diagnostics

We first present the model as a result of applying our NSCE to the synthetic data with a known dependence structure. This allows us to validate the model in a straightforward manner. Figure 4 shows the plot of conditional parameter estimates from the NSCE model along with the individual diagnostic plots for the simulated case of H_S and T_P. Notice the α parameter of the HT model with different values for the four covariate sectors which reasonably matches the true input dependence model of Figure 1. This covariate dependence is also reflected in case of β which often acts as a "counter balance". Finally μ roughly equals to 0 and σ equals 1 as expected from fitting on a bi-variate Gaussian distribution. We thus can be satisfied that the model parameters are sensible and reflect the underlying true dependence structure.

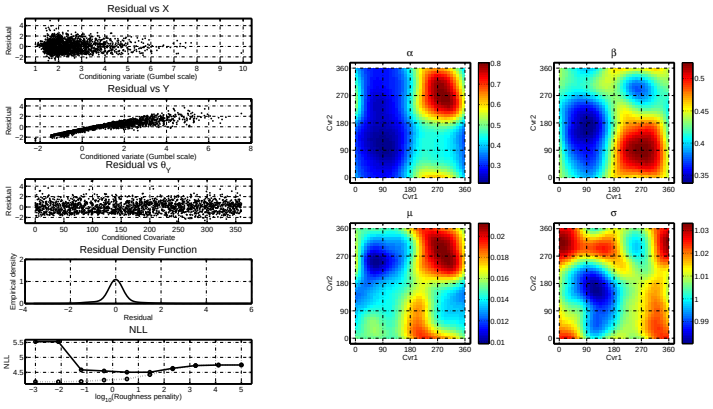


FIGURE 4. Synthetic Case: Bi-variate conditional extremes model parameter estimates and diagnostic plots. (Right) showing α, β, μ and σ parameters from the HT model as a function of bi-directional covariates. Notice the α parameter of the HT model with different values for covariate sectors which matches the true input dependence model of Figure 1. This covariate dependence is also reflected in case of β while μ roughly equals to 0 and σ equals 1 as expected. (Left) The top three show plots of residual values as a function of conditioning and conditioned variate and covariates. The fourth plot shows the empirical density fit of the residue showing a Gaussian-like distribution. The last plot shows the negative log-likelihood (NLL) as a function of the roughness penalty.

Synthetic Case Conditional Diagnostics & Return Value Estimation: We now describe the methodology to validate our conditional extremes model. First we build the distribution of T_P *conditioned* on the variate H_S satisfying certain predefined criteria. For sake of illustration let us consider the case of $T_P|H_S > h$ where we characterise the instantaneous behaviour of the peak period T_P given that the significant wave height H_S exceeding a set threshold h. Here, we simulate multiple realisations of sets of sea-state H_S events and build the *corresponding* T_P satisfying the given condition (say $H_S > h$ in this case) exactly to the period of the original sample under the model. This can be done for realisations of the original samples and also the realisations under the model.

Using this simulation procedure, we construct a cumulative distribution function F, potentially restricted to some co-variate interval A, and estimate 2.5% and 97.5% values of the cdf for each value of $T_P|H_S$. As earlier we actually plot $1-\log(F)$ to emphasise tail behaviour. These curves with 95% confidence intervals, are illustrated in black in Figure 5, is compared with the cumulative distribution function for the original sample, illustrated in red in the Figure in terms of an empirical 95% bootstrap uncertainty band estimated from 1000 bootstrap resamples of the original data. Agreement is good for omni-directional and sectoral estimates shown for $T_P|H_S > 4,6$. We notice that the uncertainty bands are getting wider as we increase the threshold as one expects due to decreasing number of samples above the threshold.

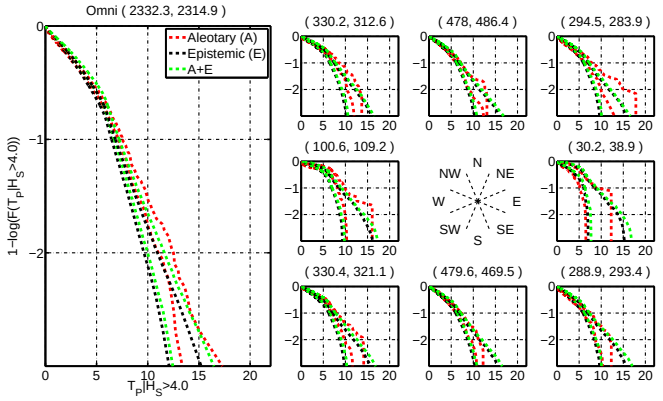
Using a synthetic example with known dependency characteristics, we have thus demonstrated the model adequacy through appropriate diagnostics and uncertainty analysis. We shall now illustrate the application of our approach to two practical examples.

4 Northern North Sea (NNS): Currents given waves

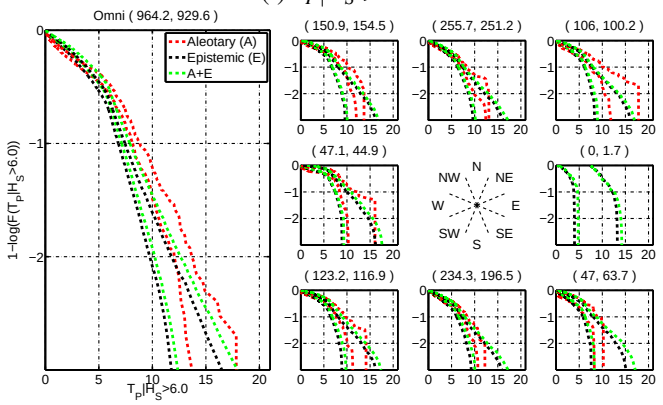
We now describe the application of the nD NSCE approach for the modelling of extreme wave height (H_S) and current speed (C_S) as a function of its two covariates i.e., wave direction θ_H and current direction θ_C for the site in NNS. In this analysis, only the winter period (Oct-Mar) is considered. For analysis of current speed, residual current speed after removal of tidal current is considered. We estimate the associated C_S directionally given a return value of H_S.

Storm characteristics are isolated from the original time-series using the procedure described in Ewans and Jonathan [2008]. Contiguous intervals of H_S above a low peak-picking threshold are identified, each interval now assumed to correspond to a storm event. The peak-picking threshold corresponds to a directional quantile of H_S with specified non-exceedance probability, estimated using quantile regression. The maximum of significant wave height during the storm interval is taken as the storm peak significant wave height or H_S^{SP}. The values of other variables (θ_H, C_S and θ_C) at the time of the storm peak significant wave height are referred to as storm peak values of those variables. The resulting storm peak sample consists of 310 observations of (H_S, θ_H, C_S, θ_C) spanning 31 years.

Figure 6 shows the plot the raw values of peaks of significant wave height H_S as a function of its covariate, wave direction θ_H ("FROM" orientation), current speeds C_S as a function of current



(i) $T_P|H_S > 4$



(ii) $T_P|H_S > 6$

FIGURE 5. Synthetic Case: Validation of the conditional return value estimation by comparing cdfs F of bootstrap resamples of the original data and realisations under the model(s) corresponding to the same time period as the original sample for $T_P|H_S > 4,6$. The plots show $1-\log(F)$ to emphasise tail behaviour showing 95% quantiles of data (red) and realisations from a single model (black) and many models (green). In each plot, the 8 right hand panels show the cdf comparisons per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_H) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional comparison. The title for each plot, in brackets are the numbers of actual and simulated events for each directional sector.

direction θ_C ("FROM" orientation). We observe significant directional variation both for H_S and C_S. For H_S, there is a shielding effect from the East (around 90^o) due to the European landmass. There are currents present for all directions with higher values from the North. Thus when we plot the sectoral scatter plot (top-left of Figure 6 (c)), we see relatively less correlation in the sector $\theta_H < 180^o, \theta_C < 180^o$. In the other sector $\theta_H > 180^o, \theta_C > 180^o$, there is a higher correlation. For each variate individually, we marginally estimate the threshold, the rate of occurrence above the threshold, and finally the GP shape and scale as a function of their respective covariates (Figure 7). The threshold and rate of occurrence estimates are consistent with the underlying data characteristics. While the GP scale for H_S correctly reflects the lack of events from the East (around 90^o again due to landmass effect), there is a corresponding counter balance effect for the GP shape. No such trend is observed in C_S. The model estimates are validated by observing the marginal return value diagnostics generated using the same procedure described in §3.2 shown in Figure 8 indicating good agreement for the period of data.

For the dependence model, we then convert the variates to Gumbel scale and fit a NSCE model as described earlier. Similar to the model description for the synthetic example of §3, we see covariate effects observed for the North Sea case clearly reflected in the model. Figure 9 shows the plot of conditional parameter estimates from the NSCE model (right) along with the individual diagnostic plots (left) for the NNS case of H_S and C_S. The diagnostic plots (left) showing the residuals, its density functions and fit negative log likelihood are all consistent and needs no further elaboration. We now focus on the right-side plots describing the model parameters as a function of bi-directional covariates.

Notice the α parameter of the HT model with different values across the four covariate sectors. As observed in the raw correlation plots, the α values are relatively small around 90^o for the H_S (shown as "Cvr1" in the plot) which is due to the landmass effect. Similarly, for the sector $\theta_H > 180^o, \theta_C > 180^o$, the α values are higher as observed earlier in the scatter plot. This effect is almost "counter-balanced" in case of β which again shows strong covariate effects. The values of μ and σ are again close to 0 and 1 respectively, consistent with the earlier case. We thus can be satisfied that the model parameters are sensible and reflects the underlying physical process. Of course, since we do not know the actual true model (if there is one), the only way we can validate it rigorously is by generating return value diagnostics as done for the synthetic case.

Similar to the marginal case, we use a return value simulation to assess the dependence model adequacy as described for the synthetic case in §3.3. Accordingly, we simulate multiple realisations of sets of sea-state H_S events and build the *corresponding* C_S satisfying a given condition (say $H_S > h$) exactly to the period of the original sample under the model and estimate 2.5% and 97.5% values of the cdf for each value of $C_S|H_S$. We plot $1-\log(F)$, with 95% confidence intervals, illustrated in black in Figure 10, and compare with the cumulative distribution function for the original sample, illustrated in red in the figure in terms of an empirical 95% bootstrap uncertainty band estimated from 1000 bootstrap resamples of the original data. Agreement is good for omni-directional and sectoral estimates shown for $C_S|H_S > 4,8$. We observe that the uncertainty bands are progressively getting wider with higher thresholds due to decreasing sample size.

Having validated both the marginal and the conditional model,

we can now estimate design criteria for any given return period using the same simulation procedure. Figure 11 compares the estimates of the independent design criterion of C_S from the marginal model and the joint design criterion of $C_S|H_S$ for a period of 100 years. It is is clear that the omni-directional joint criterion (around 0.5 m/s at the median) is much less than the independent criterion (0.8 m/s) which would have a significant impact on the eventual design of the structures. Likewise our approach allows to estimate joint design criteria for any arbitrary subset of covariates.

5 Offshore Brazil (OB): Current profile with depth

We now describe the second application of the nD NSCE approach for the modelling of extreme currents (Cs_j) at depths $j = 1, 2, ..., p$ as a function of its respective current directions θ_{Cj} for the OB site. We would like to estimate the associated Cs_j directionally given a return value of surface current Cs_1. Current characteristics are isolated from the original time-series using the procedure described in §4. The maximum surface current during a time interval is taken as the Cs_1. The values of other variables (θ_{C1}, Cs_j and θ_{Cj}) at the time of the maximum surface current are referred to as peak values of those variables. The resulting storm peak sample consists of 348 observations spanning 5.2 years.

In this analysis, for illustration we have chosen two pairs of depths to do the conditional analysis, one with a high degree of correlation owing to closer depths and second that are sufficiently far apart. Accordingly we chose Cs_1, Cs_2 and Cs_{43} for our analysis. Figure 12 plots the raw values of peaks of surface current Cs_1 as a function of its covariate, θ_{C1}, current speeds Cs_2 and Cs_{43} as a function of their respective current directions θ_{C2} and θ_{C43}. From the figure, Cs_1 and Cs_2 have a high degree of correlation across many sectors (mostly in the first sector $\theta_{C1} < 180^o, \theta_{C2} < 180^o$ and the last sector $\theta_{C1} >= 180^o, \theta_{C2} >= 180^o$) whereas Cs_1 and Cs_{43} are much less correlated across all the sectors.

As in the previous cases, here too we fit 3 marginal models each for Cs_1, Cs_2 and Cs_{43} followed by two dependence models for $Cs_2|Cs_1$ and $Cs_{43}|Cs_1$. We validated these models with the same diagnostic procedure as earlier but have chosen to only present the results for sake of brevity. Figures 13 and 14 compare the estimates of the independent design criteria of $Cs_j, j = 2, 43$ from the marginal model and the conditional design criteria of $Cs_j|Cs_1, j = 2, 43$ for a period of 100 years. It is clear that the conditional median criterion for $Cs_2|Cs_1$ (around 0.8 m/s) is higher and closer to the independent criterion (1.6 m/s) owing to higher correlation compared to $Cs_{43}|Cs_1$ (0.5 m/s compared to 1.4 m/s) which is much less correlated. The effect of conditioning in reducing conditional design values relative to unconditioned values is nevertheless clear, and careful structural design should exploit this.

6 Discussion & Conclusions

Covariate effects are important in extreme value analysis, for individual metocean parameters and for joint modelling. Incorporating covariate effects in multivariate extreme value models is challenging in general. However, as demonstrated in this paper, incorporation of covariate effects in the conditional extremes model is possible. Moreover, for both the NNS and OB applications illustrated, estimated models including covariate effects are different to those excluding covariates, reflect physical reality more adequately, and lead to different estimates for return values.

Estimation of joint occurrences of extremes of environmental variables is crucial for design of offshore facilities and achieving consistent levels of reliability. Specification of joint extremes in design criteria has often been somewhat ad hoc, being based on fairly arbitrary combination of extremes of variables estimated independently. Such approaches are even outlined in design guidelines. More rigorous methods for modelling joint occurrences of extremes of environmental variables are now available. In particular, the conditional extremes model provides a straightforward approach to joint modelling of extreme values, based on solid theory. It admits different forms of extremal dependence, ensuring that the data (rather than unwittingly made modelling assumptions) drive the estimation of design values. The model admits uni- and multi-variate covariate effects, is scalable to high dimensions and allows uncertainty analysis via simulation. For this reason, we recommend the conditional extremes model for joint extremes modelling in both response-based and response-independent metocean 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. [2010] 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.

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 of 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.

Acknowledgments

We are thankful to the support and encouragement of Vianney Koelman and Bertwim van Beest from the Computational Technology Platform, metocean engineers Graham Feld and Fan Shejun, useful discussions with team members in Bangalore and Manchester, and academic colleagues at Lancaster University.

REFERENCES

H. D. Bondell, B. J. Reich, and H. Wang. Noncrossing Quantile Regression Curve Estimation. *Biometrika*, 97:825–838, 2010.

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.

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 Mech. Arc. Engg.*, 130(10), 2008.

K. C. Ewans and P. Jonathan. Evaluating Environmental Joint Extremes for the Offshore Industry Using the Conditional Extremes Model. *J. Marine Systems*, 120:124–130, 2014.

S. Haver. Wave Climate Off Northern Norway. *Applied Ocean Research*, 7:85–92, 1985.

J. E. Heffernan and J. A. Tawn. A Conditional Approach for Multivariate Extreme Values. *J. R. Statist. Soc. B*, 66:497–546, 2004.

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, J. Flynn, and K. C. Ewans. Joint Modelling of Wave Spectral Parameters for Extreme Sea States. *Ocean Engg.*, 37: 1070–1080, 2010.

P. Jonathan, K. C. Ewans, and J. Flynn. Joint Modelling of Vertical Profiles of Large Ocean Currents. *Ocean Engg.*, 42:195–204, 2012.

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 of Northern North Sea Storm Characteristics. *Environmetrics*, 25(3):172–188, 2014.

L. Raghupathi, D. Randell, P. Jonathan, and K. Ewans. Fast Computation of Large Scale Marginal Spatio-Directional Extremes. *Comp. Stat. Dat. Anal.*, 95:243–258, 2016.

P. S. Tromans and L. Vanderschuren. Risk Based Design Conditions in the North Sea: Application of a New Method. *Offshore Technology Confernence, Houston (OTC–7683)*, 1995.

S. R. Winterstein, T. C. Ude, C. A. Cornell, P. Bjerager, and S. Haver. Environmental Parameters for Extreme Response: Inverse FORM with Omission Factors. In *Proc. 6th Int. Conf. on Structural Safety and Reliability, Innsbruck, Austria*, 1993.

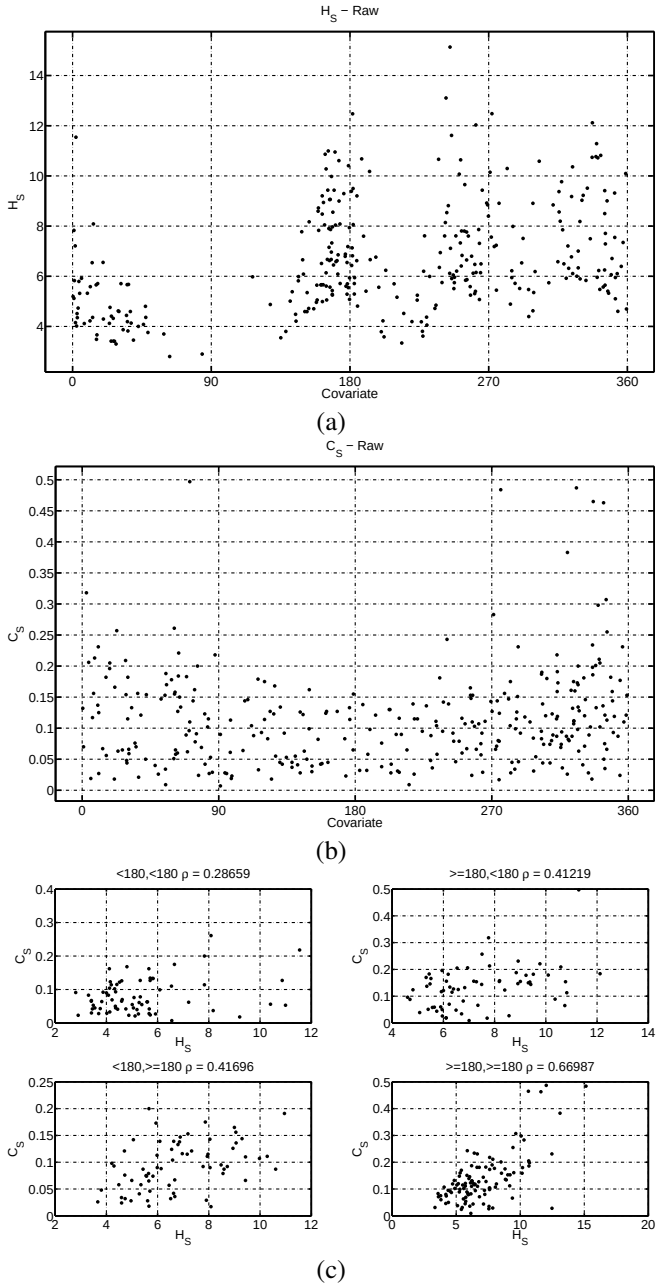


FIGURE 6. NNS: Raw input data showing (a) H_S (m) and (b) C_S (m/s) plotted as a function of their respective directional covariates i.e., θ_H and θ_C. All directions (including the currents) are direction "FROM". Notice H_S values indicating directional effects are mostly absent in the east (around 90^o) due to landmass effect. C_S also shows directional effects with higher values around North West (270-360^o) (c) Scatter plot of C_S vs H_S shown in different directional sectors indicating lower correlation in the first sector ($\theta_H > 180^o, \theta_C > 180^o$) but higher correlation in the fourth sector ($\theta_H > 180^o, \theta_C > 180^o$).

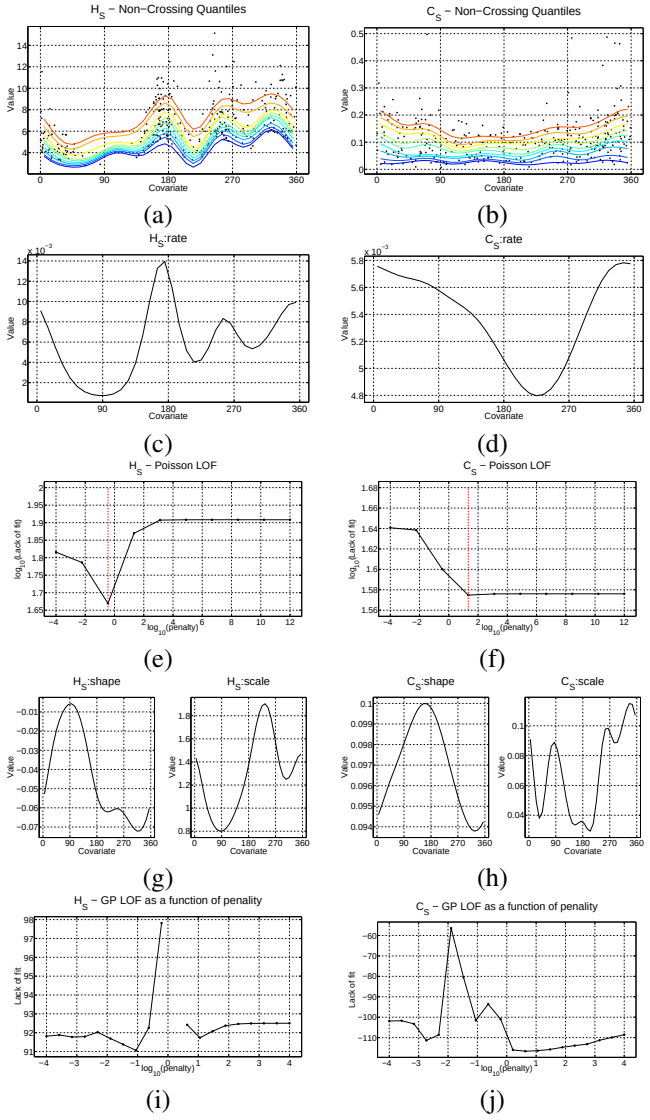
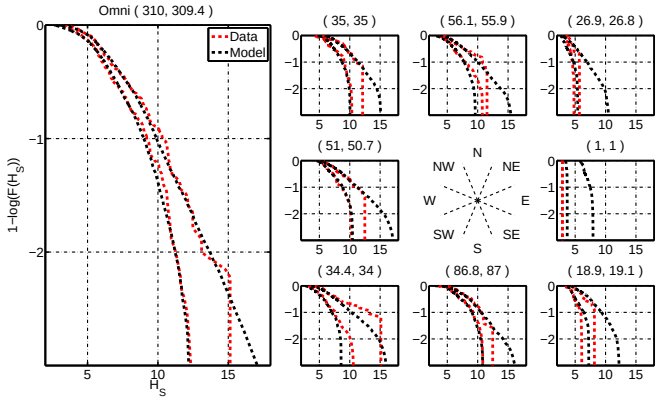
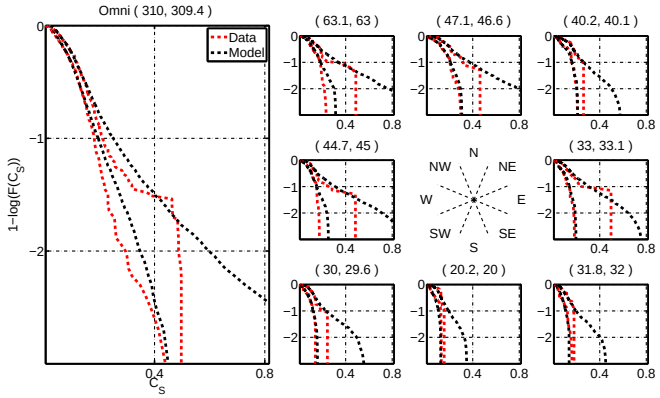


FIGURE 7. NNS: Marginal parameter estimates for H_S (left) and C_S (right). (a) & (b) Raw data with non-crossing quantile estimates. (c) & (d) Poisson rate of occurence estimates. (e) & (f) Cross-validatory selection of optimal roughness penalty of the Poisson model. (g) & (h) GP model shape and scale estimate. (i) & (j) Cross-validatory selection of optimal roughness penalty of the GP model.



(i) NNS: H_S



(ii) NNS: C_S

FIGURE 8. NNS: Validation of the marginal return value estimation by comparing cdfs F of bootstrap resamples of the original data and realisations under the model corresponding to the same time period as the original for (top) H_S and (bottom) C_S. Plot of $1-\log(F)$ showing 95% quantiles for realisations of data (red) and model (black). For each plot, the 8 right hand panels show the cdf comparisons per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_H and θ_C) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional comparison. The title for each plot, in brackets are the numbers of actual and simulated events for each directional sector.

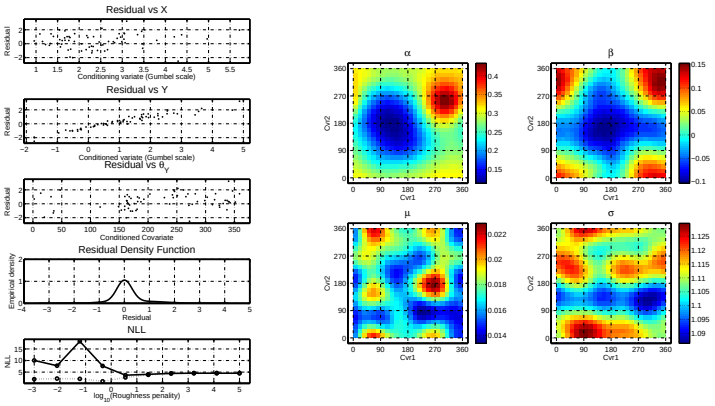
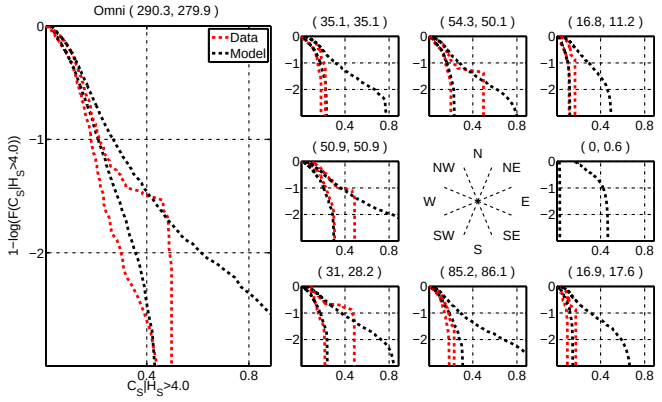
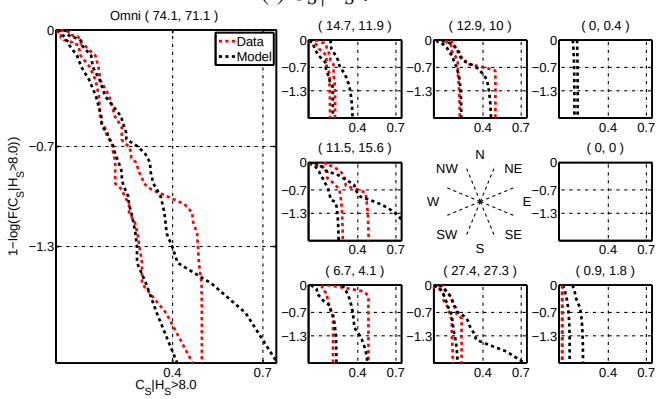


FIGURE 9. NNS: Bi-variate conditional extremes model parameter estimates and diagnostic plots.(Right) showing α, β, μ and σ parameters from the HT model as a function of bi-directional covariates. Notice the α parameter of the HT model with different values for covariate sectors agrees with the scatter plots of Figure 6. This covariate dependence is also reflected in case of β while μ roughly equals to 0 and σ equals 1 as expected. (Left) The top three show plots of residual values as a function of conditioning and conditioned variate and covariates. The fourth plot shows the empirical density fit of the residue showing a Gaussian-like distribution. The last plot shows the negative log-likelihood (NLL) as a function of the roughness penalty.

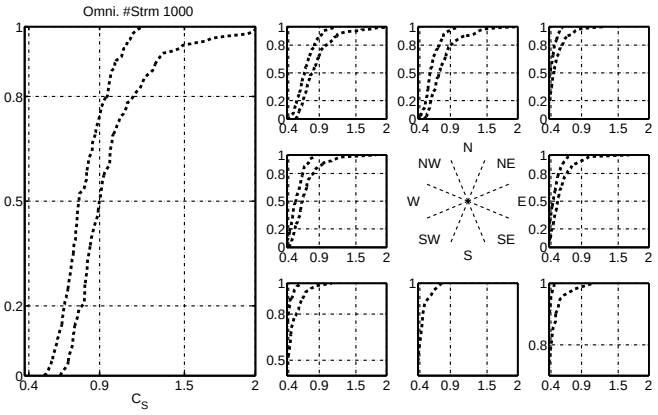


(i) $C_S|H_S > 4$

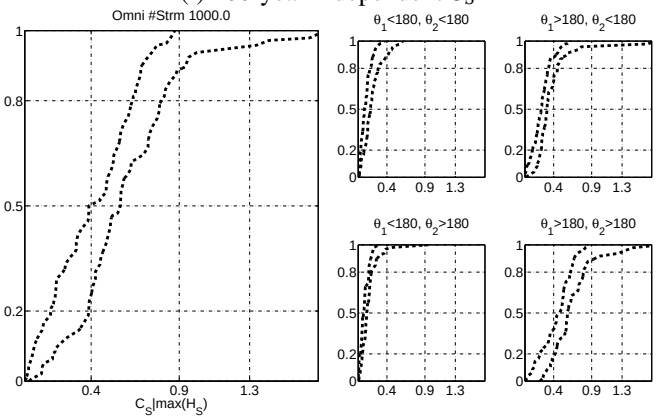


(ii) $C_S|H_S > 8$

FIGURE 10. NNS: Validation of the conditional return value estimation by comparing cdfs F of bootstrap resamples of the original data and realisations under the model corresponding to the same time period as the original for (top) $C_S|H_S > 4$ and (bottom) $C_S|H_S > 8$. The plots show $1 - \log(F)$ to emphasise tail behaviour showing 95% quantiles of data (red) and model (black). In each plot, the 8 right hand panels show the cdf comparisons per directional sector of the conditioning variate for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_H) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional comparison. The title for each plot, in brackets are the numbers of actual and simulated events for each directional sector.

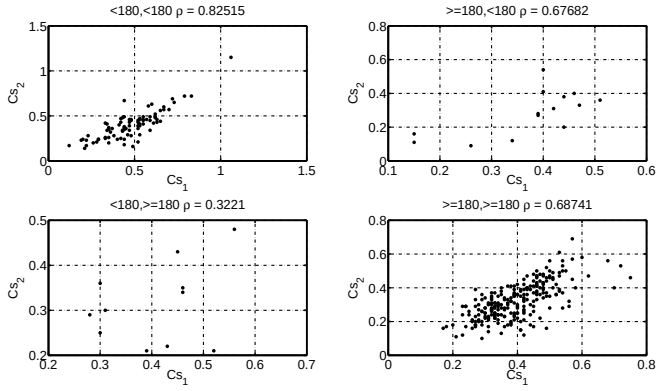


(i) 100-year independent C_S

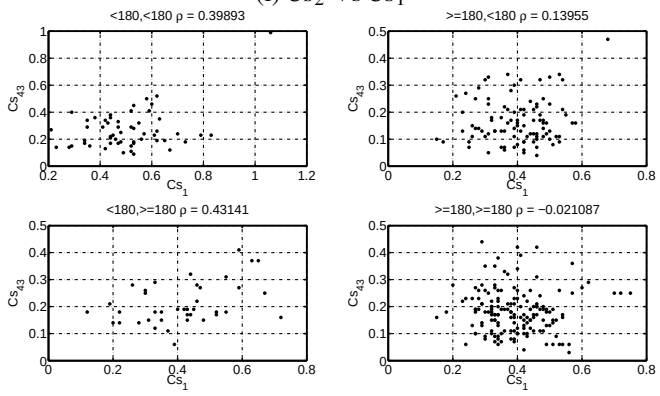


(i) 100-year joint $C_S|\max(H_S)$

FIGURE 11. NNS: Sectoral marginal and conditional return value estimates showing 95% quantiles of realisations the model showing (top) C_S and (right) $C_S|\max(H_S)$ (both in m/s) for a 100-year period. For each plot, the 8 right hand panels show the cdf estimates per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_H and θ_C) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional estimate with its title indicating the average number events for the return period of simulation.

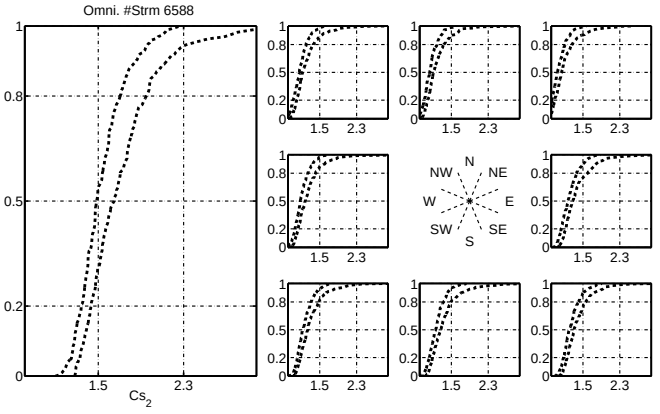


(i) Cs_2 Vs Cs_1

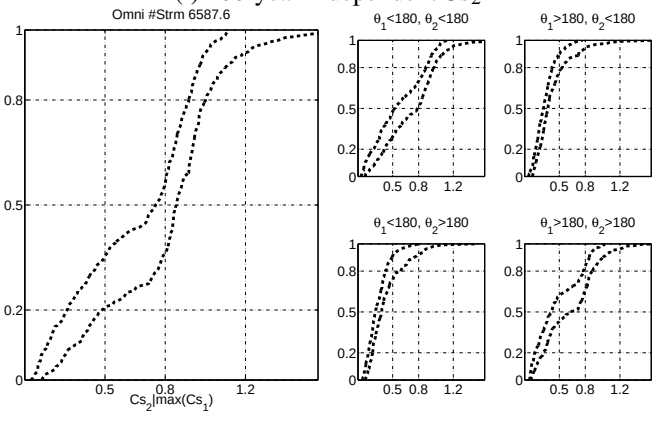


(ii) Cs_{43} Vs Cs_1

FIGURE 12. OB: Current speed scatter plots at different depths in different directional sectors (i) Cs_2 Vs Cs_1 and (ii) Cs_{43} Vs Cs_1. The title in each of the four subplot indicates the bi-directional covariate quadrant range (e.g., $\theta_{C1} < 180^o, \theta_{C2} < 180^o$ for top left and so on), followed by the correlation coefficient ρ for the observations in that quadrant.

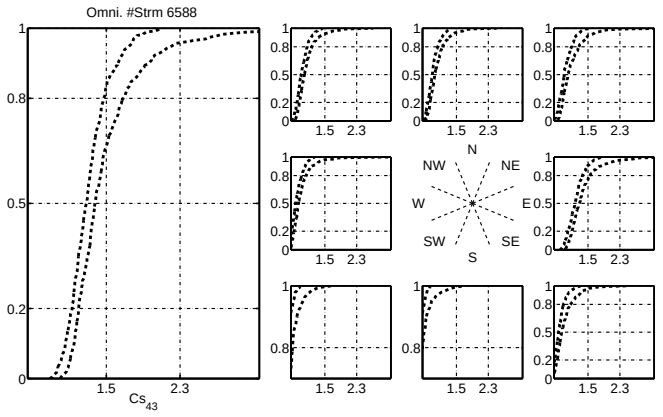


(i) 100-year independent Cs_2

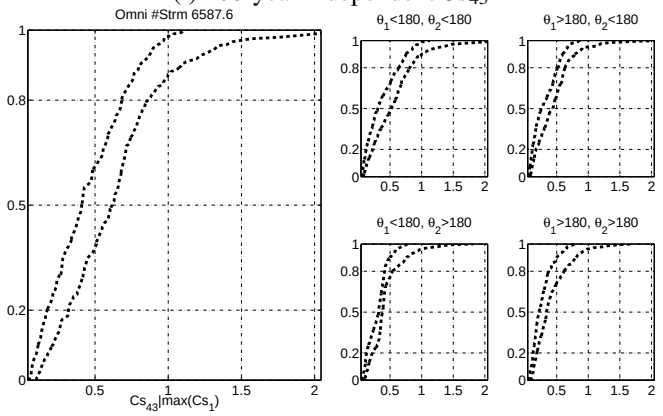


(ii) 100-year joint $Cs_2|Cs_1$

FIGURE 13. OB: Sectoral marginal and conditional return value estimates showing 95% quantiles of realisations the model showing (top) Cs_2 and (right) $Cs_2|\max(Cs_1)$ (both in m/s) for a 100-year period. For each plot, the 8 right hand panels show the cdf estimates per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_{Cs_1} and θ_{Cs_2}) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional estimate with its title indicating the average number events for the return period of simulation.



(i) 100-year independent Cs_{43}



(ii) 100-year joint $Cs_{43}|Cs_1$

FIGURE 14. OB: Sectoral marginal and conditional return value estimates showing 95% quantiles of realisations the model showing (top) Cs_{43} and (right) $Cs_{43}|\max(Cs_{43})$ (both in m/s) for a 100-year period. For each plot, the 8 right hand panels show the cdf estimates per directional sector for the 8 directional octants centred (from left to right, top to bottom) on covariate directions (i.e., θ_{Cs_1} and $\theta_{Cs_{43}}$) from NW, N, and NE; W and E; SW, S and SE respectively. The left hand panel shows the equivalent omni-directional estimate with its title indicating the average number events for the return period of simulation.