

## RESEARCH ARTICLE

# Bayesian inference for nonstationary marginal extremes

D. Randell<sup>1</sup> | K. Turnbull<sup>2</sup> | K. Ewans<sup>3</sup> | P. Jonathan<sup>1\*</sup><sup>1</sup>Shell Projects & Technology, Manchester, M22 0RR, U.K.<sup>2</sup>Department of Mathematics & Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.<sup>3</sup>Sarawak Shell Bhd., Kuala Lumpur, 50450, Malaysia**Correspondence**Philip Jonathan, Shell Projects & Technology, Brabazon House, Concord Business Park, Manchester M22 0RR, U.K.  
Email: philip.jonathan@shell.com

We propose a simple piecewise model for a sample of peaks-over-threshold, non-stationary with respect to multidimensional covariates, and estimate it using a carefully designed and computationally efficient Bayesian inference. Model parameters are themselves parameterized as functions of covariates using penalized B-spline representations. This allows detailed characterization of non-stationarity extreme environments. The approach gives similar inferences to a comparable frequentist penalized maximum likelihood method, but is computationally considerably more efficient and allows a more complete characterization of uncertainty in a single modelling step. We use the model to quantify the joint directional and seasonal variation of storm peak significant wave height at a northern North Sea location and estimate predictive directional–seasonal return value distributions necessary for the design and reliability assessment of marine and coastal structures.

**KEYWORDS**

Bayesian, covariate, extreme, generalized Pareto, non-stationarity, ocean wave, Poisson process, return value, spline, storm severity, Weibull

## 1 | INTRODUCTION

Bayesian inference provides an intuitive framework for environmental applications of extreme value analysis, allowing, for example, incorporation of prior knowledge and thorough uncertainty quantification within a single inference step, predictive inference, and when carefully designed a computationally efficient inference. Coles and Powell (1996) and Stephenson (2016) provide useful reviews. There are many applications of Bayesian inference for extreme value models in the literature. For example, Coles and Tawn (1996) reports a Bayesian analysis of extreme rainfall data. Coles and Tawn (2005) estimates extreme surges on the UK east coast for improved flood risk assessment, Cooley et al. (2006) explores Bayesian extremes in lichenometry as a means of understanding climate change, Scotto and Guedes-Soares (2007) estimates the distributions of significant wave height and Mendes et al. (2010) considers extremes of wild fires. Beirlant et al. (2004) and Northrop et al. (2016a) consider the specification of objective priors for extreme value model parameters, and Smith and Walshaw (2003) of informative priors.

Davison et al. (2012) consider Bayesian hierarchical models for spatial extremes.

Extreme value models describe the tails of distributions; using a sample of peaks over threshold, only the largest values are generally used in extreme value estimation. Choice of extreme value threshold is therefore generally critical (e.g., Scarrott & MacDonald 2012). For descriptions of whole samples, mixture models and piecewise distributions are attractive. For example, Frigessi et al. (2002) uses a dynamic mixture model for fire loss data comprising a generalized Pareto (GP) tail and Weibull body, and Behrens et al. (2004) considers a piecewise distribution comprised of a parametric form for the body of data and a GP tail. MacDonald et al. (2011) proposes a piecewise distribution in which a kernel density estimate for the body is used in conjunction with a GP tail. In oceanographic applications, many distributional forms have been proposed for wave height and significant wave height, motivated by observation and physical considerations. Forristall (1978), Haver (1985), Smith and Naylor (1987), Battjes and Groenendijk (2000), Ferreira and Guedes Soares (2000), Prevosto et al. (2000), Scotto and Guedes-Soares

(2000), and Guedes-Soares and Scotto (2001) provide useful examples. The Weibull distribution is often used to model ocean current speed and wind speed. Monahan (2006) shows the Weibull distribution to be a fair approximation for surface wind speeds, and Chu (2008) uses the Weibull to model upper tropical Pacific current speeds. Ashkenazy and Gildor (2011) finds the Weibull distribution to be a good approximation of ocean current speeds and highlights the effect of covariates. Morgan et al. (2011) explores the performance of different distributions for wind speeds.

Incorporating non-stationarity, or the effect of covariates, in extreme value analysis is also important (e.g., Davison & Smith 1990; Jonathan et al. 2008). Anderson et al. (2001) describes a seasonal model for return values of significant wave height. Covariate effects are considered in Coles and Walshaw (1994) for directional wind speeds and in Mendez et al. (2008) for seasonal significant wave height. Exploration of temporal non-stationarity in extremes can also be found in Renard et al. (2006), Cheng et al. (2014), and Ortego Martinez et al. (2014). There are many ways to parameterize nonstationary effects in statistical models, including penalized B-splines. Ugarte et al. (2012) and Oumow et al. (2012) use this formulism to estimate smooth parameter variation, the latter within a mixture model. Bayesian penalized B-splines as described by Brezger and Lang (2006) are particularly useful for the current work.

A variety of MCMC algorithms (e.g., Gamerman & Lopes 2006) is available to sample from posterior distributions. Here, we adopt a combination of different sampling schemes to achieve good computational efficiency and chain mixing, including direct Gibbs sampling of full conditional distributions and Metropolis–Hastings (MH) within Gibbs. We find it advantageous in MH sampling from some full conditionals to exploit gradient and curvature information to improve the quality of proposals. Where possible, we employ the Riemann manifold Metropolis-adjusted Langevin algorithm (mMALA) as described by Girolami and Calderhead (2011) and Xifara et al. (2014). Estimation schemes, which exploit gradient and curvature information for problematic functions such as the GP likelihood, are recommended (e.g., Jones et al., 2016).

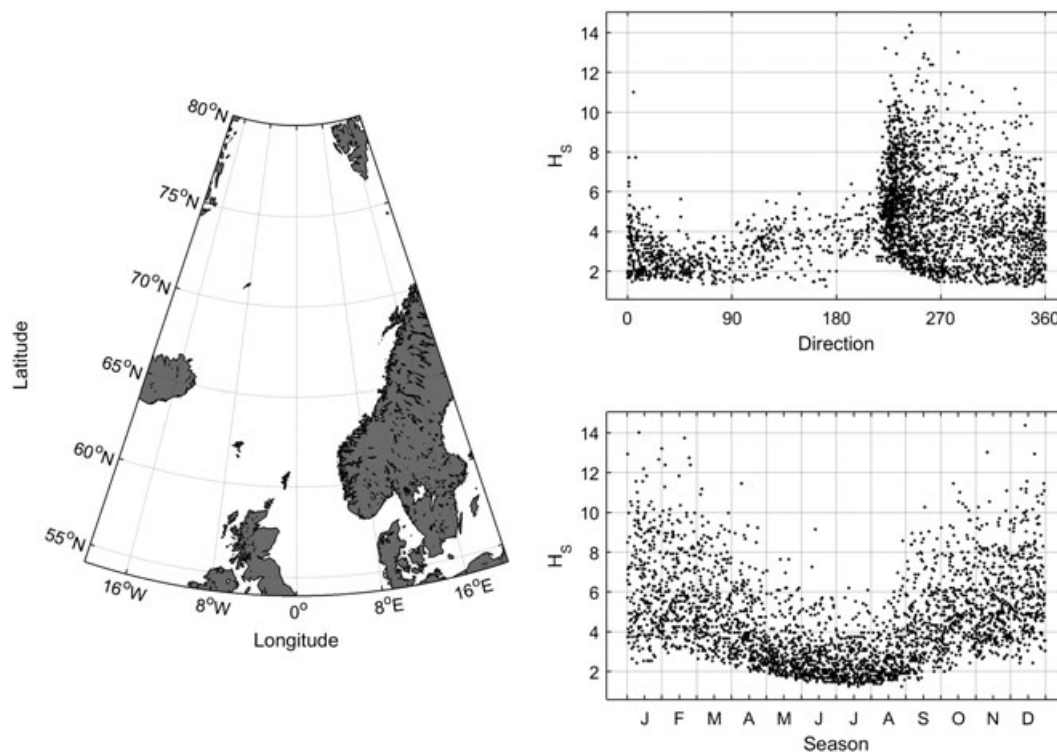
The objective of the current work is to develop a simple whole-sample model for independent observations given multidimensional covariates, incorporating the GP form for exceedances of a nonstationary threshold. The model is needed to estimate distributions of return values corresponding to long return periods, for use in the design and reliability assessment of marine and coastal structures. A major consideration in the development of the methodology is ease-of-use and computational efficiency for full-scale oceanographic applications for sample sizes from  $10^3$  to  $10^7$  with at least a two-dimensional (e.g., directional–seasonal) covariate domain.

The article is arranged as follows. Section 2 provides a motivating application to modeling of storm-peak significant

wave height events at a location in the northern North Sea, subject to directional and seasonal variability in the rate and magnitude of occurrence of events. Section 3 introduces a piecewise truncated Weibull (TW)—GP model for the size of events (Section 3.1), and a Poisson process model for the rate of occurrence of events (Section 3.2). Model parameters are nonstationary with respect to covariates, parameterized in terms of the tensor product of B-spline bases for the directional–seasonal covariate domain (Section 3.3). The specification of prior distributions for model parameters is discussed in Section 3.4 and Bayesian inference in Section 3.5. Application to the northern North Sea data is outlined in Section 4. Discussion and conclusions, including comments on a comparative study at this location and in the South China Sea, are given in Section 5. The appendix describes key elements of the inference scheme in more detail.

## 2 | MOTIVATING APPLICATION

Significant wave height measures the energy (or roughness) of the ocean surface and can be defined as four times the standard deviation of the ocean surface elevation at a spatial location for a specified period of observation. The application sample is taken from the Wave Model (WAM) hindcast of Reistad et al. (2011), which provides time series of significant wave height, (dominant) wave direction, and season (defined as day of the year, for a standardized year consisting of 360 days) for 3-hr sea states for the period September 1957 to December 2012 at a northern North Sea location. (A hindcast is a physical model of the ocean environment, incorporating pressure field, wind field, and wind-wave generation models in particular; the hindcast model is calibrated to observations of the environment from instrumented offshore facilities, moored buoys, and satellite altimeters in the neighborhood of the location for a period of time, typically decades.) Aarnes et al. (2012) and Breivik et al. (2013) study extreme value characteristics of storm severities from the hindcast. Extreme sea states in the North Sea are dominated by winter storms originating in the Atlantic Ocean and propagating eastwards across the northern part of the North Sea. Due to their proximity to the storms, sea states at northern North Sea locations are usually more intense than in the southern North Sea. Occasionally, the storms travel southeastward and intrude into the southern North Sea producing large sea states. Directions of propagation of extreme seas vary considerably with location, depending on land shadows of the British Isles, Scandinavia, and the coast of mainland Europe, and fetches associated with the Atlantic Ocean, Norwegian Sea, and the North Sea itself. In the northern North Sea, the main fetches are the Norwegian Sea to the North, the Atlantic Ocean to the west, and the North Sea to the south. Extreme sea states from the directions of Scandinavia to the east and the British Isles to the southwest are not possible. The shielding by these land masses is more effective for southern North



**FIGURE 1** Left: Sample location, offshore Norway at latitude exceeding 62°. Right: Storm peak significant wave height  $H_S$  (in meters) on direction (upper panel) and season (lower panel)

Sea locations, resulting in a similar directional distribution but reduced wave heights by comparison with northern North Sea locations.

Storm peak significant wave height characteristics are isolated from the hindcast time series using the procedure described in Ewans and Jonathan (2008) for a location offshore Norway at a latitude exceeding 62°. Contiguous intervals of significant wave height 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–seasonal quantile of significant wave height 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 ( $H_S$  for brevity in this work). The values of directional and seasonal covariates at the time of storm peak significant wave height are referred to as storm peak values of those variables. The resulting storm peak sample consists of 2,941 values of  $H_S$ . With direction from which a storm travels expressed in degrees clockwise with respect to north, Figure 1 shows a map of the region and plots of  $H_S$  (in meters) versus direction and season.

It can be seen that the land shadow of Norway (approximately the directional interval (45°, 210°)) has a considerable effect on the rate and size of occurrences with direction. In particular, there is a dramatic increase in both rate and size of occurrences with increasing direction at around 210°, corresponding to Atlantic storm events from the southwest able to pass the Norwegian headland. We therefore should expect

considerable directional variability in model parameter estimates for the sample. In contrast, the magnitude of the rate of change of both rate and size of occurrences with respect to season is lower; this should also be reflected in model parameters. Winter storms (approximately from October to March) are more intense and numerous.

### 3 | MODEL

#### 3.1 | Size of exceedance

The magnitude  $y$  of  $H_S$  peaks-over-threshold events is assumed to follow a nonstationary piecewise (or two-part) distribution, the form of which is motivated by both physical and statistical considerations. Wave models derived from theory or estimated historically from measurements (Section 1) suggest that a suitably parameterized Weibull distribution provides an adequate representation for the body of the distribution of individual wave heights, crest elevations, and storm severities (such as  $H_S$ ). Moreover, asymptotic theory suggest that an extreme value model is required to describe the largest threshold exceedances. We therefore specify the piecewise model as follows.

For response  $y$ , non-exceedances of some threshold  $\psi$  are assumed to be distributed according to a nonstationary three-parameter TW distribution with scale  $\alpha$  ( $>0$ ), shape  $\gamma$  ( $>0$ ), pre-specified location  $\zeta$  ( $<\psi$ ), and support  $[\zeta, \psi]$ . Threshold exceedances of  $\psi$  are assumed to be distributed according to a nonstationary GP distribution with scale

$\sigma (>0)$ , shape  $\xi (\in \mathbb{R})$ , and support  $(\psi, y^+)$  (where  $y^+ = \psi - \frac{\sigma}{\xi}$  for  $\xi < 0$  and  $=\infty$  otherwise). Note that all of  $\alpha$ ,  $\gamma$ ,  $\sigma$ ,  $\xi$ , and  $\zeta$  will be assumed to be smoothly varying functions of direction and season in due course.

The corresponding probability density function

$$f(y|\tau, \alpha, \gamma, \sigma, \xi) = \begin{cases} \tau \times f_{TW}(y|\tau, \alpha, \gamma) & \text{for } y \in [\zeta, \psi] \\ (1 - \tau) \times f_{GP}(y|\tau, \alpha, \gamma, \sigma, \xi) & \text{for } y \in (\psi, y^+) \end{cases}$$

is similar to the (stationary) model of Behrens et al. (2004). The pre-specified location parameter  $\zeta$  is omitted from this description since it is not inferred; in the current work,  $\zeta$  corresponds to a nonstationary “peak-picking” threshold used for storm peak identification prior to model estimation (Section 2). The factors  $\tau$  and  $(1 - \tau)$  in the piecewise density serve to weight the relative contributions of the TW and GP parts; parameter  $\tau$  corresponds to the extreme value threshold non-exceedance probability assumed stationary with respect to covariates, providing a scalar transition to describe the transition from TW to GP.  $\tau$  is estimated as part of the inference.

Below extreme value threshold  $\psi$ ,  $y$  follows a TW distribution with density

$$f_{TW}(y|\tau, \alpha, \gamma) = \frac{f_W(y|\alpha, \gamma)}{F_W(\psi|\alpha, \gamma, \tau)} \text{ for } y \in [\zeta, \psi]$$

where

$$f_W(y|\alpha, \gamma) = \frac{\gamma}{\alpha} \left( \frac{y - \zeta}{\alpha} \right)^{\gamma-1} \exp \left( - \left( \frac{y - \zeta}{\alpha} \right)^{\gamma} \right).$$

Note the threshold  $\psi$  is specified as  $\zeta + \alpha(-\log(1 - \tau))^{-1/\gamma}$  such that

$$F_W(\psi|\alpha, \gamma, \tau) = 1 - \exp \left( - \left( \frac{\psi - \zeta}{\alpha} \right)^{\gamma} \right) = \tau \text{ always.}$$

Above extreme value threshold  $\psi$ ,  $y$  follows a GP distribution with density

$$f_{GP}(y|\tau, \alpha, \gamma, \sigma, \xi) = \frac{1}{\sigma} \left( 1 + \frac{\xi}{\sigma} (y - \psi) \right)^{-1/\xi-1} \text{ for } y \in (\psi, y^+).$$

In practice, we choose to carry out computations with  $\nu = \sigma(1 + \xi)$  in place of  $\sigma$ . This choice is motivated by the fact that, with  $\xi > -0.5$ , maximum likelihood estimates for  $\nu$  and  $\xi$  are asymptotically independent, simplifying their joint estimation compared with joint estimation of  $\sigma$  and  $\xi$ . Nevertheless, (derived) estimates for  $\sigma$  will in general be reported and visualized below where possible for ease of interpretation.

### 3.2 | Rate of occurrence

The rate of occurrence of all events (both exceedances and non-exceedances of threshold  $\psi$ ) are assumed to follow a nonstationary Poisson process with rate  $\rho$  varying with direction and season. For events on covariate domain  $\mathcal{D}$  (Section 3.3) at locations  $\mathbf{r} = \{\mathbf{r}_i\}_{i=1}^n$ , the Poisson process likelihood is  $f(\mathbf{r}|\rho) = \exp(-\int_{\mathcal{D}} \rho(\mathbf{r}) d\mathbf{r}) \prod_{i=1}^n \rho(\mathbf{r}_i)$ . Following Chavez-Demoulin and Davison (2005), we approximate this integral by evaluating it on the index set of  $m$  directional-seasonal subintervals of area  $d$ , assuming  $d$  small

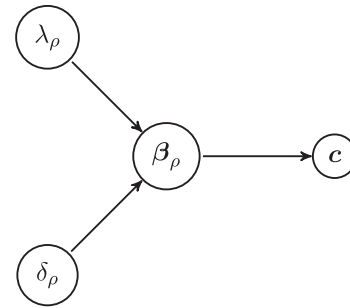


FIGURE 3 Directed acyclic graph for Poisson inference of event rate of occurrence

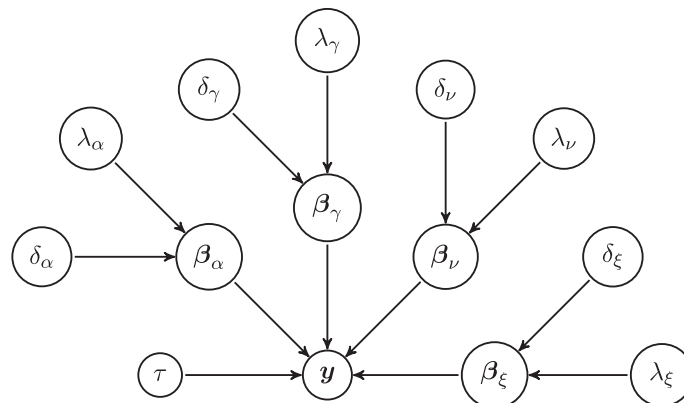


FIGURE 2 Directed acyclic graph for truncated Weibull-generalized Pareto inference of event magnitude



enough so that  $\rho$  is constant on each subinterval. Then, for a vector of counts  $\mathbf{c} = \{c_k\}_{k=1}^m$  of occurrences on the index set

$$f(\mathbf{c}|\rho) = \exp\left(-d \sum_{k=1}^m \rho_k\right) \prod_{k=1}^m \rho_k^{c_k},$$

where  $\rho = \{\rho_k\}_{k=1}^m$  is the corresponding Poisson count rate.

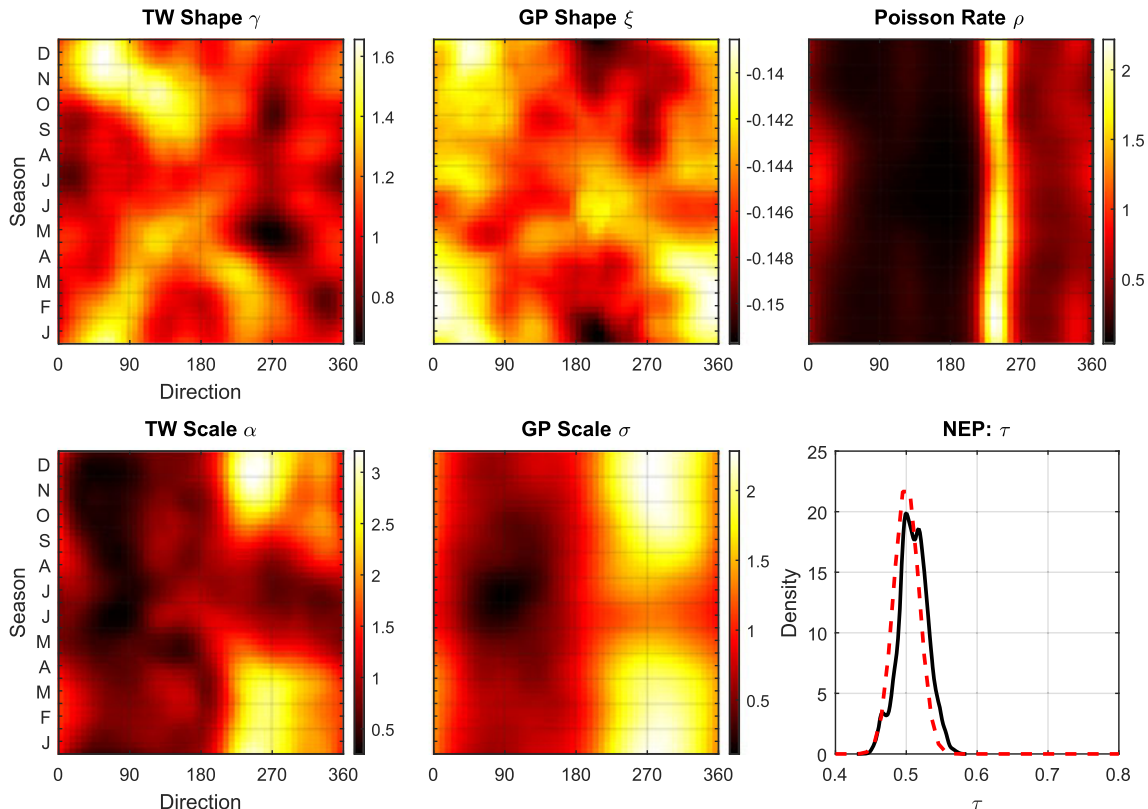
### 3.3 | Spline parameterization

Physical considerations suggest we should expect model parameters  $\rho$ ,  $\alpha$ ,  $\gamma$ ,  $\nu$ , and  $\xi$  to vary smoothly with respect to directional and seasonal covariates (whereas threshold non-exceedance probability  $\tau$  is assumed stationary). For estimation, this can be achieved by expressing each parameter in terms of an appropriate basis for the domain  $\mathcal{D}$  of covariates, where  $\mathcal{D} = \mathcal{D}_2 \times \mathcal{D}_1$ .  $\mathcal{D}_1 = \mathcal{D}_2 = [0, 360]$  are, respectively, the (marginal) domains of storm direction and season under consideration. We calculate a  $m_1 \times p_1$  periodic marginal B-spline basis matrix  $\mathbf{B}_1$  for an index set of  $m_1=50$  directional locations and a  $m_2 \times p_2$  periodic marginal B-spline basis matrix  $\mathbf{B}_2$  for an index set of  $m_2=50$  seasonal locations, yielding a total of  $m(=m_1 m_2)$  combinations of covariate values in the index set on  $\mathcal{D}$ . Following the work of Eilers and Marx and coauthors (e.g., Marx & Eilers 1998; Eilers et al. 2006

and citations thereof), we define the  $m \times p$  basis matrix  $\mathbf{B}$  (where  $p = p_1 p_2$ ) for  $\mathcal{D}$  using the tensor product  $\mathbf{B} = \mathbf{B}_2 \otimes \mathbf{B}_1$ . The value of any parameter vector  $\boldsymbol{\eta}$  on the 2D index set can therefore be expressed as  $\mathbf{B}\boldsymbol{\beta}_\eta$  for some  $p \times 1$  vector  $\boldsymbol{\beta}_\eta$  of spline parameters. Model estimation reduces to estimating appropriate sets of spline parameters for each  $\eta \in \{\rho, \alpha, \gamma, \sigma, \xi\}$ . Here, marginally for periodic directional and seasonal bases, we set  $p_1=p_2=10$ , although we also considered  $p_1=p_2=20$  (and  $m_1=m_2=100$ ) to confirm lack of sensitivity of inferences to this choice.

In any single covariate dimension  $j(=1,2)$ , the roughness  $R_{\eta_j}^\circ$  of any  $p_j \times 1$  spline parameter vector  $\boldsymbol{\beta}_{\eta_j}^\circ$  can be evaluated on the index set of  $m_j$  locations as  $R_{\eta_j}^\circ = \boldsymbol{\beta}_{\eta_j}^{\circ\prime} \mathbf{P}_{\eta_j}^\circ \boldsymbol{\beta}_{\eta_j}^\circ$  where  $\mathbf{P}_{\eta_j}^\circ$  is a  $p_j \times p_j$  penalty matrix of the form  $\mathbf{P}_{\eta_j}^\circ = \mathbf{D}_{\eta_j}^{\circ\prime} \Delta_{\eta_j}^\circ \mathbf{D}_{\eta_j}^\circ$ .  $\mathbf{D}_{\eta_j}^\circ$  is a  $r_j \times p_j$  difference matrix, taken in this work to be a first-order difference, such that  $\mathbf{D}_{\eta_j}^\circ \boldsymbol{\beta}_{\eta_j}^\circ$  yields a  $r_j \times 1$  vector of differences between values of consecutive elements of  $\boldsymbol{\beta}_{\eta_j}^\circ$ .  $\Delta_{\eta_j}^\circ$  is a  $r_j \times r_j$  diagonal matrix with the  $r_j \times 1$  vector of roughness penalties  $\boldsymbol{\delta}_{\eta_j}^\circ$  on its leading diagonal,  $\Delta_{\eta_j}^\circ = \text{diag}(\boldsymbol{\delta}_{\eta_j}^\circ)$ . In this work, since both directional and seasonal domains are periodic, we have  $r_j=p_j, j=1,2$ .

Spline roughness  $R_\eta$  on  $\mathcal{D}$  can be expressed in terms of marginal characteristics using tensor products. We extend the marginal difference matrices using  $\mathbf{D}_{\eta_1} = \mathbf{I}_2 \otimes \mathbf{D}_{\eta_1}^\circ$  and  $\mathbf{D}_{\eta_2} = \mathbf{D}_{\eta_2}^\circ \otimes \mathbf{I}_1$ , where  $\mathbf{I}_1$  and  $\mathbf{I}_2$  are respectively  $p_1 \times p_1$  and  $p_2 \times p_2$  identity matrices. The extended  $p \times p$  penalty matrices become  $\mathbf{P}_{\eta_j} = \mathbf{D}_{\eta_j}^{\prime} \Delta_{\eta_j} \mathbf{D}_{\eta_j}$ ,  $j=1,2$ .  $\Delta_{\eta_1}$  is a  $p_2 r_1 \times p_2 r_1$  diagonal



**FIGURE 4** Directional-seasonal plots for posterior median TW, GP, and Poisson rate ( $\times 10^{-3}$ ) parameter estimates on the covariate index set, and prior (dashed red) and posterior densities for extreme value threshold non-exceedance probability ( $\tau$ )

matrix, the leading diagonal of which is the vector of spline penalties  $\delta_{\eta 1}$ .  $\Delta_{\eta 2}$  is defined analogously. The overall spline roughness penalty matrix on  $\mathcal{D}$  is  $\mathbf{P}_{\eta} = \lambda_{\eta 1} \mathbf{P}_{\eta 1} + \lambda_{\eta 2} \mathbf{P}_{\eta 2}$  for spline roughness penalty coefficient vector  $\lambda_{\eta} = (\lambda_{\eta 1}, \lambda_{\eta 2})$ .  $\mathbf{P}_{\eta}$  is used as the prior precision matrix for spline coefficients  $\beta_{\eta}$ . Random roughness penalties, as opposed to fixed penalties of unit size, promote better mixing of the MCMC chain (see Brezger & Lang 2006).

The computational burden of the matrix manipulations above can become restrictive as  $m$  and  $p$  increase. Fortunately, using generalized linear array methods (GLAM, Currie et al. 2012) we avoid the need ever to calculate the full  $m \times p$  basis matrix  $\mathbf{B}$ , since  $\mathbf{B}\beta_{\eta}$  is equivalent (after rearrangement) to  $\mathbf{B}_{\eta 1} \mathcal{M}(\beta_{\eta}) \mathbf{B}'_{\eta 2}$ , where  $\mathcal{M}(\beta_{\eta})$  is a  $p_1 \times p_2$  rearrangement of the  $p \times 1$  vector  $\beta_{\eta}$ .

### 3.4 | Prior specification

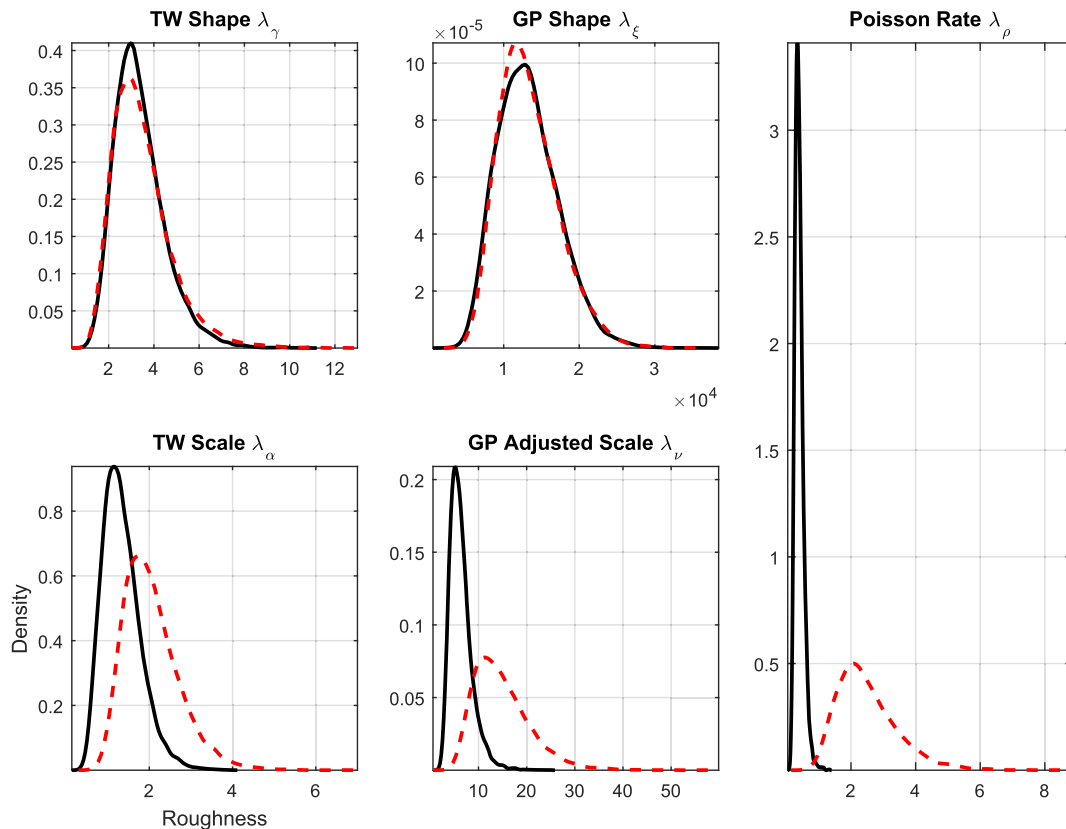
The following prior specification is made for  $\tau$ , the components of vectors  $\delta_{\eta 1}$ ,  $\delta_{\eta 2}$  and the scalars  $\lambda_{\eta 1}$ ,  $\lambda_{\eta 2}$  for  $\eta \in \{\rho, \alpha, \gamma, \nu, \xi\}$ . The extreme threshold non-exceedance probability  $\tau$  follows beta distribution  $B(a_{\tau}, b_{\tau})$ , with fixed parameters. This distributional form is suitable for the description of a probability since it has range  $[0, 1]$ . The choice of parameters  $a_{\tau}, b_{\tau}$  was motivated by the requirement to have sufficient data to estimate the GP parameters well (i.e.,  $\tau$  relatively

small) but also ensure reasonable extreme value tail fit (i.e.,  $\tau$  relatively large). The Supporting Information accompanying this article provides further discussion of this topic. The two components of positive-valued spline roughness penalty coefficient vector  $\lambda_{\eta}$  are assumed to be gamma-distributed  $\Gamma(a_{\lambda}, b_{\lambda})$  independently and identically. This choice of prior ensures that the corresponding full conditional distributions are available in closed form, enabling Gibbs sampling. The distribution of the spline roughness penalty vector  $\delta_{\eta}$  is not inferred; its positive-valued components are taken to follow gamma distribution  $\Gamma(a_{\delta}, b_{\delta})$  identically and independently. This distributional form and the parameter values for  $a_{\delta}$ ,  $b_{\delta}$  were recommended by Brezger and Lang (2006) and adopted here following some confirmatory exploratory analysis. Further description of prior specification, including values for  $a_{\tau}$ ,  $b_{\tau}$ ,  $a_{\delta}$ ,  $b_{\delta}$ ,  $a_{\lambda}$ , and  $b_{\lambda}$  is given in the Appendix.

In addition, the vector of spline parameters  $\beta_{\eta}$  corresponding to  $\eta \in \{\rho, \alpha, \gamma, \nu, \xi\}$  is distributed with density proportional to  $\exp\left(-\frac{1}{2} \beta'_{\eta} \mathbf{P}_{\eta} \beta_{\eta}\right)$  (Brezger & Lang 2006), where penalty matrix  $\mathbf{P}_{\eta}$  is a function of  $\delta_{\eta}$  and  $\lambda_{\eta}$  (Section 3.3).

### 3.5 | Inference

For a sample  $\mathbf{y}$  of event sizes, we estimate the model described in Section 3 with spline representations of the form  $\eta = \mathbf{B}\beta_{\eta}$  on the index set of covariate locations for



**FIGURE 5** Estimated posterior densities for directional and seasonal (dashed red) spline roughness penalty coefficients for TW, GP, and Poisson rate parameters

parameters  $\eta \in \{\rho, \alpha, \gamma, \nu, \xi\}$ . The roughness of each  $\eta$  with respect to directional and seasonal covariates respectively is regulated using the  $2 \times 1$  roughness coefficient vector  $\lambda_\eta$  (itself inferred) and spline penalties  $\delta_\eta$  (drawn at random from an assumed known distribution). In addition, the non-exceedance probability  $\tau$  corresponding to the extreme value threshold is estimated. The full parameter set is therefore  $\Omega = \{\tau, \beta_\alpha, \delta_\alpha, \lambda_\alpha, \beta_\gamma, \delta_\gamma, \lambda_\gamma, \beta_\nu, \delta_\nu, \lambda_\nu, \beta_\xi, \delta_\xi, \lambda_\xi\}$ . Inference can be expressed conveniently in terms of a directed acyclic graph (Figure 2).

The analogous representation for a sample  $c$  of event rates is given in Figure 3.

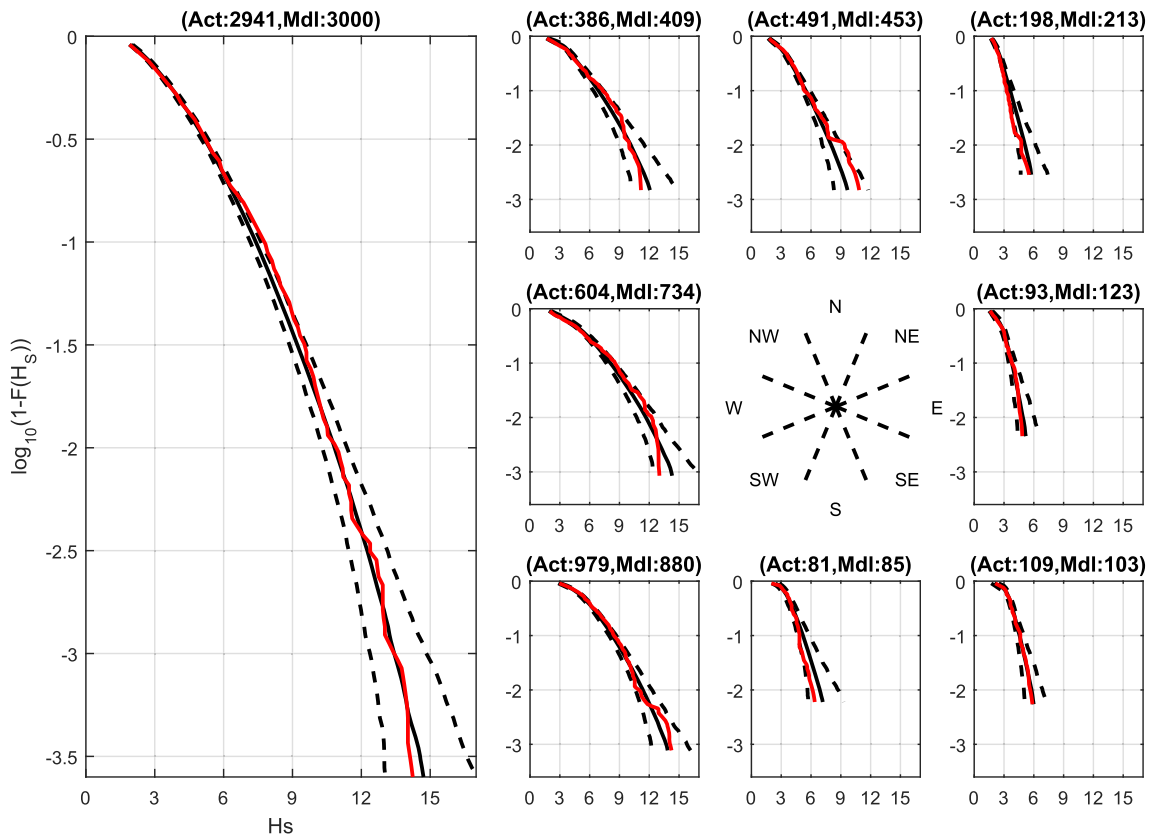
The posterior distribution of model parameters is not available in closed form for either event size or rate of occurrence. Posterior inference is therefore made using Monte Carlo Markov chain (MCMC) by sampling from full conditionals as detailed in the Appendix and outlined below. When full conditionals are available in closed form, we use Gibbs sampling; otherwise, we use MH in Gibbs. When the function to be sampled is particularly problematic (e.g., posteriors for GP shape and adjusted scale), we exploit function gradient and curvature information if possible to generate MH proposals efficiently. In particular, full conditional distributions for the spline roughness penalty coefficients  $\lambda_\eta = (\lambda_{\eta 1}, \lambda_{\eta 2})$  are avail-

able in closed form, permitting Gibbs sampling. With gamma prior specification (Section 3.4), the posterior distribution is  $\Gamma(a_{\eta j} + \frac{1}{2} \text{rank}(\mathbf{P}_\eta), b_{\eta j} + \frac{1}{2} \beta'_\eta \mathbf{P}_\eta \beta_\eta)$ .

At each iteration of the MCMC chain, we sample in turn from the full conditional distribution of extreme threshold non-exceedance probability  $\tau$ , the spline parameters  $\beta_\rho$ , then spline roughness penalties  $\delta_\rho$ , then spline roughness penalty coefficients  $\lambda_\rho$  for Weibull scale  $\alpha$ . Then, we sample from the corresponding triplets of parameters for TW shapes  $\alpha$  and  $\gamma$ , and GP shape  $\xi$  and adjusted scale  $\nu$ . We experimented with different orderings of parameters and found there to be little difference in performance. Further algorithmic details are given in the Appendix.

#### 4 | APPLICATION

The model of Section 3 was applied to the northern North Sea application sample. The posterior distribution of parameters was estimated using MCMC, incorporating 20,000 burn-in iterations, with the same number of subsequent iterations to estimate the posterior. Visual inspection of individual MCMC chains suggested good mixing; informal comparisons of multiple chains confirmed satisfactory convergence. Figure 4



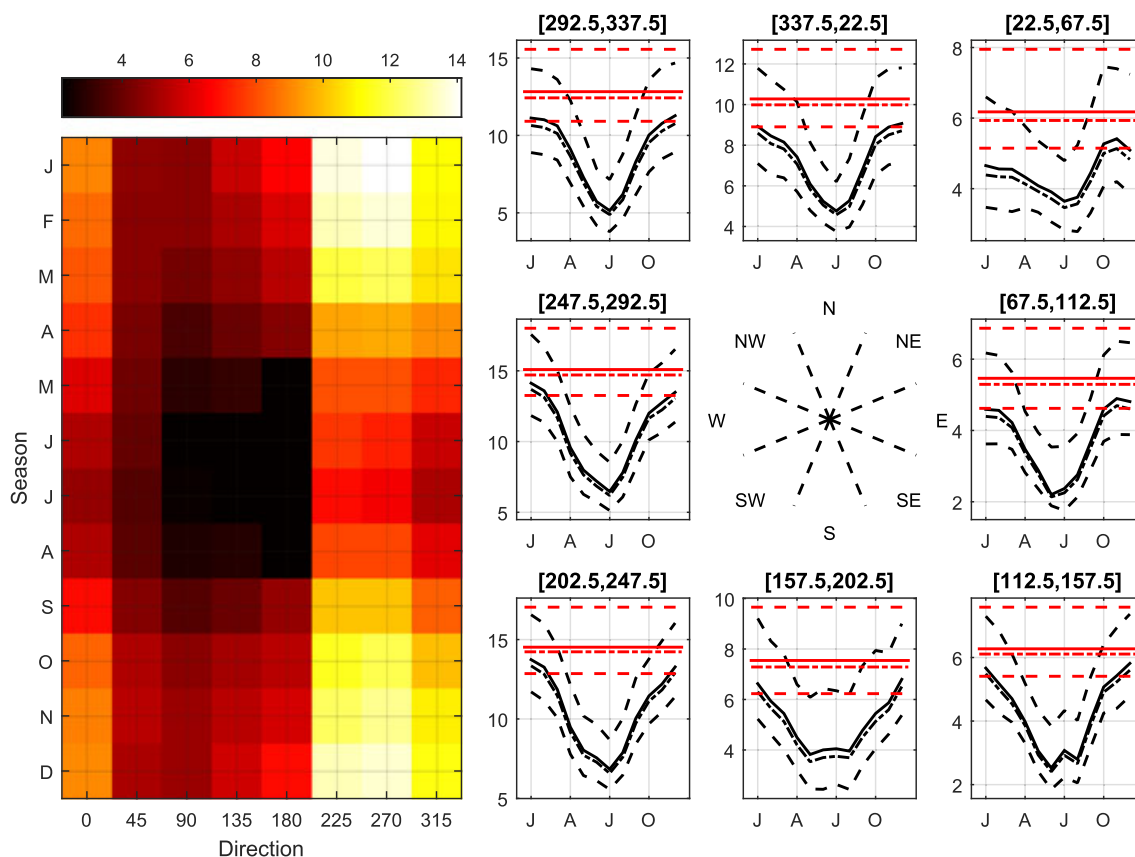
**FIGURE 6** Illustration of model validation by comparison of estimates for the distribution of  $H_S$  (in meters) corresponding to the period of the original sample, plotted as  $\log_{10}(1 - F(H_S))$  to accentuate tail behavior. The solid red curve is an empirical estimate. The black curves summarize the predictive distribution of the quantile estimate (for given tail probability  $1 - F(H_S)$ ) from multiple realizations of time series under the directional–seasonal model, as the median (solid) and 2.5% and 97.5% values (dashed). The left hand panel corresponds to the omni-directional omni-seasonal case, and the right hand panels to omni-seasonal estimates for each of eight directional octants. Titles for plots, in brackets, are the numbers of actual events and the median number of events simulated

shows directional–seasonal plots for posterior median TW, GP, and Poisson rate parameter estimates, and prior (dashed red) and posterior densities for extreme value threshold non-exceedance probability ( $\tau$ ). The strong directional variability of rate of occurrence is most apparent, compared to more modest seasonal variation. TW and GP scale parameter estimates both indicate large event magnitudes at around  $270^\circ$ , corresponding to storms from the Atlantic. The GP shape parameter estimate is approximately  $-0.14$ , typical for northern North Sea locations. The posterior distribution of non-exceedance probability is relatively similar to its prior, implying that the sample provides relatively little evidence for the value of this parameter.

Figure 5 shows estimated posterior densities for directional and seasonal (dashed red) spline roughness penalty coefficients for TW, GP, and Poisson rate parameters. Directional and seasonal densities for TW and GP shape parameters are very similar, suggesting that the same degree of spline roughness with respect to either covariate is appropriate. However, seasonal roughness penalties for TW and GP scale parameters are higher than directional roughness penalties, indicating that scale parameters show more directional variability.

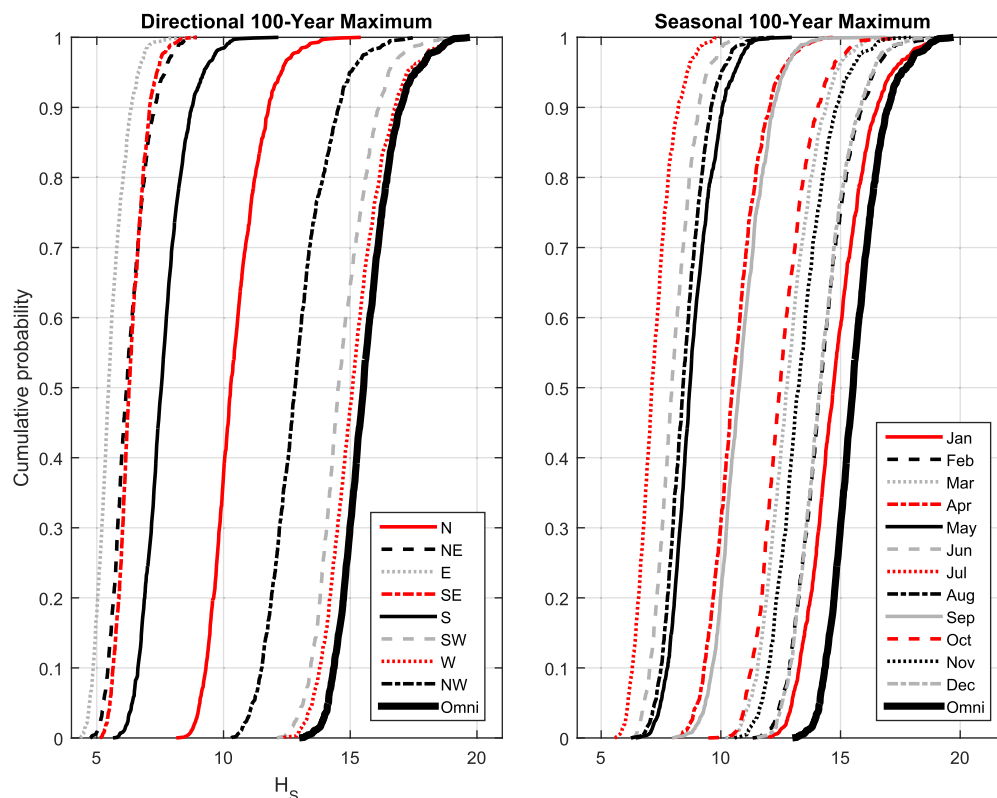
Using the estimated posterior distribution  $f(\Omega|y,c)$  of parameters  $\Omega$  given sample  $\{y,c\}$ , we can estimate predictive distributions for  $H_S$  and the maximum of  $H_S$  corresponding to any return period of interest by simulation under the estimated model. For response  $y^*$  of interest, with known density  $f(y^*|\Omega)$  potentially restricted to a specified (joint) covariate interval, the posterior predictive density  $f(y^*|y,c)$  of  $y^*$  is  $\int f(y^*|\Omega)f(\Omega|y,c)d\Omega$ . Figure 6 illustrates model validation by comparison of estimates for the distribution of  $H_S$  corresponding to the period of the original sample. The empirical estimate, found simply from sorting the original sample and shown as a solid red curve, is in good agreement with predictive simulations under the model, both omni-directionally and for 8 directional octants. Agreement is also good between the actual numbers of events observed in each directional sector examined and the corresponding median number of simulated events.

Figure 7 gives a directional–seasonal plot for the predictive distribution of the 100-year maximum. The left-hand panel shows directional and seasonal variability of the median estimate from multiple realizations of time-series of period 100 years under the directional–seasonal model. The right hand panel summarizes seasonal variation of the predictive distribution



**FIGURE 7** Directional–seasonal plot for the predictive distribution of the 100-year maximum (in meters). The left-hand panel shows directional and seasonal variability of the median estimate from multiple realizations of time series of period 100 years under the directional–seasonal model. The right hand panel summarizes seasonal variation of the predictive distribution for each of eight directional octants in terms of the 2.5%, 37%, median, and 97.5% values of the predictive distribution (black). Also shown are the corresponding omni-seasonal estimates (in red). Titles for plots in right hand panel give the angular intervals corresponding to the directional octants





**FIGURE 8** Empirical predictive cumulative distribution functions for the 100-year maximum  $H_s$  (in meters) from simulation under the directional–seasonal model. Left-hand and right-hand panels show directional (octant) and seasonal (monthly) predictive distributions, respectively

bution for each of 8 directional octants. Again, there is systematic directional and seasonal variation present. For design purposes, extreme value inferences are usually reported in terms of directional and seasonal return value distributions, for return periods such as 100 years (Figure 8). Return value estimates such as these are critical for design and reliability assessments of marine and coastal structures. They indicate the need to consider directional and seasonal variability of extreme ocean environments to maximize safety and operability. They can also be used as a basis for decision making concerning the relative risk of different short-term marine operations.

## 5 | DISCUSSION AND CONCLUSIONS

We report the development of a simple piecewise model for a sample of peaks-over-threshold, nonstationary with respect to multidimensional covariates, estimated using Bayesian inference. We demonstrated the utility of the model in application to a directional–seasonal analysis of storm peak significant wave height at a northern North Sea location and estimation of predictive directional–seasonal return value distributions necessary for the design and reliability assessment of marine and coastal structures.

These authors have performed a number of recent analyses similar to that reported here using maximum likelihood inference, incorporating cross-validation to esti-

mate optimal spline roughness penalty coefficients and an all-encompassing bootstrap scheme to estimate uncertainties in model parameters and return values (e.g., Jonathan et al., 2014; Feld et al., 2015). In particular, we have performed a direct comparison of maximum likelihood and Bayesian inference for the South China Sea application (reported in Randell et al., 2015), obtaining good agreement for model parameter and return value estimates. Moreover, maximum likelihood inferences for the northern North Sea location considered here are provided in the Supporting Information accompanying this work, again showing good agreement between maximum likelihood and Bayesian inferences. In comparison with these recent studies, the current Bayesian approach appears to offer a number of advantages. Firstly, inference is computationally considerably less demanding than maximum likelihood inference incorporating cross-validation and bootstrapping, particularly for estimation of spline roughness penalty coefficients. Model and return value estimation for a single MCMC chain can be completed in 30 min using modest computational resources (e.g., a single core PC with 2.7 GHz processor and 4 GB RAM, running MATLAB software in Microsoft Windows 7). Secondly, it appears that inference schemes, which sample a function in a random but directed fashion, rather than seeking function minimum deterministically, provide more stable estimation for the models and samples considered. Finally, for the Bayesian scheme, estimation of different spline roughness penalty coefficients for each model parameter in each (marginal) covariate is straight-

forward; achieving the same within the maximum likelihood setting would be computationally more demanding. Numerous improvements and enhancements of the approach are currently under consideration. The choice of distributional form for the “body” part of the piecewise density somewhat arbitrary, although there is physical reasoning supporting the choice of Weibull distribution. In the current formulation, the piecewise density is discontinuous at the extreme value threshold  $\psi$ , although the cumulative distribution function is continuous on  $(\zeta, y^+)$  for given values of covariates. Estimation of spline roughness penalties ( $\delta_\eta$  for  $\eta \in \{\rho, \alpha, \gamma, \nu, \xi\}$ ) as in Brezger and Lang (2006) might be considered, although sampling from the prior performs well. Estimating  $\delta_\eta$  would come at some computational cost however because of the large number of penalties involved. From an applications perspective, current work is centered on evaluating the approach for higher-dimensional covariates (e.g., longitude, latitude, direction, and season), enabling coherent (marginal) spatial analysis over large spatial domains (e.g., Raghupathi et al., 2015).

## ACKNOWLEDGMENTS

We acknowledge discussions with Graham Feld, Matthew Jones, and Rasmus Kulseng at Shell and Jonathan Tawn at Lancaster University. We are also grateful to two anonymous reviewers for thorough and thoughtful review of the work. The WAM hindcast is courtesy of Reistad et al. (2011). MATLAB algorithms for estimation of nonstationary extreme value models are available via Northrop et al. (2016b). Further assistance is available from the authors on request.

## REFERENCES

- Aarnes, O. J., Breivik, O., & Reistad, M. (1996). Wave extremes in the Northeast Atlantic. *Journal of Climate*, 25, 1529–1543.
- Anderson, C., Carter, D., & Cotton, P. (2016). Wave climate variability and impact on offshore design extremes. Report commissioned from the University of Sheffield and Satellite Observing Systems for Shell International.
- Ashkenazy, Y., & Gildor, H. (1996). On the probability and spatial distribution of ocean surface currents. *Journal of Physical Oceanography*, 41(12), 2295–2306.
- Battjes, J. A., & Groenendijk, H. W. (2005). Wave height distributions on shallow foreshores. *Coastal Engineering*, 40, 161–182.
- Behrens, C. N., Lopes, H. F., & Gamerman, D. (2006). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4, 227–244.
- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2007). *Statistics of extremes: Theory and applications*. Chichester, UK: Wiley.
- Breivik, O., Aarnes, O. J., Bidlot, J.-R., Carrasco, A., & Saetra, O. (2010). Wave extremes in the North East Atlantic from ensemble forecasts. *Journal of Climate*, 26, 7525–7540.
- Brezger, A., & Lang, S. (2004). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis*, 50, 967–991.
- Chavez-Demoulin, V., & Davison, A. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54, 207–222.
- Cheng, L., Aghakouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, 127, 353–369.
- Chu, P. C. (2012). Probability distribution function of the upper equatorial pacific current speeds. *Geophysical Research Letters*, 35, L12606.
- Coles, S., & Walshaw, D. (2012). Directional modelling of extreme wind speeds. *Applied Statistics*, 43, 139–157.
- Coles, S. G., & Powell, E. A. (2002). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, 64, 119–136.
- Coles, S. G., & Tawn, J. A. (2004). A Bayesian analysis of extreme rainfall data. *Applied Statistics*, 45, 463–478.
- Coles, S. G., & Tawn, J. A. (2011). Bayesian modelling of extreme sea surges on the UK east coast. *Philosophical Transactions of the Royal Society A*, 363, 1387–1406.
- Cooley, D., Naveau, P., Jomelli, V., Rabatel, A., & Grancher, D. (1978). A Bayesian hierarchical extreme value model for lichenometry. *Environmetrics*, 17, 555–574.
- Currie, I. D., Durban, M., & Eilers, P. H. C. (1985). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society Series B*, 68, 259–280.
- Davison, A., & Smith, R. L. (1987). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B*, 52, 393.
- Davison, A. C. (2000). *Statistical models*. Cambridge University Press, Cambridge, UK.
- Davison, A. C., Padoan, S. A., & Ribatet, M. (2000). Statistical modelling of spatial extremes. *Statistical Science*, 27, 161–186.
- Eilers, P. H. C., Currie, I. D., & Durban, M. (2000). Fast and compact smoothing on multi-dimensional grids. *Computational Statistics & Data Analysis*, 50, 61–76.
- Ewans, K. C., & Jonathan, P. (2000). The effect of directionality on northern North Sea extreme wave design criteria. *Journal of Offshore Mechanics and Arctic Engineering*, 130, 041604:1–041604:8.
- Feld, G., Randell, D., Wu, Y., Ewans, K., & Jonathan, P. (2001). Estimation of storm peak and intra-storm directional-seasonal design conditions in the North Sea. *Journal of Offshore Mechanics and Arctic Engineering*, 137, 021102:1–021102:15.
- Ferreira, J. A., & Guedes Soares, C. (2006). Modelling distributions of significant wave height. *Coastal Engineering*, 40(4), 361–374.
- Forristall, G. Z. (2008). On the statistical distribution of wave heights in a storm. *Journal of Geophysical Research*, 83, 2353–2358.
- Frigessi, A., Haug, O., & Rue, H. (2011). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5, 219–235.
- Gamerman, D., & Lopes, H. F. (2011). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Boca Raton, USA: Chapman & Hall / CRC.
- Girolami, M., & Calderhead, B. (1990). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, 73, 123–214.
- Green, P. J., & Silverman, B. (2008). *Nonparametric regression and generalised linear models: A roughness penalty approach*. London, UK: Chapman and Hall.
- Guedes-Soares, C., & Scotto, M. (2001). Modelling uncertainty in long-term predictions of significant wave height. *Ocean Engineering*, 28, 329–342.
- Haver, S. (1994). Wave climate off northern Norway. *Applied Ocean Research*, 7, 85–92.
- Jonathan, P., Ewans, K. C., & Forristall, G. Z. (2008). Statistical estimation of extreme ocean environments: the requirement for modelling directionality and other covariate effects. *Ocean Engineering*, 35, 1211–1225.
- Jonathan, P., Randell, D., Wu, Y., & Ewans, K. (2006). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88, 520–532.
- Jones, M., Randell, D., Ewans, K., & Jonathan, P. (2016). Statistics of extreme ocean environments: Non-stationary inference for directionality and other covariate effects. *Ocean Engineering*, 119, 30–46.
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., & Russell, G. (2014). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55, 2137–2157.

- Marx, B. D., & Eilers, P. H. C. (2012). Direct generalised additive modelling with penalised likelihood. *Computational Statistics & Data Analysis*, 28, 193–209.
- Mendes, J. M., de Zea Bermudez, P. C., Pereira, J., Turkman, K. F., & Vasconcelos, M. J. P. (2012). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. *Environmental and Ecological Statistics*, 17, 1–28.
- Mendez, F. J., Menendez, M., Luceno, A., Medina, R., & Graham, N. E. (2006). Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering*, 35, 131–138.
- Monahan, A. H. (2006). The probability distribution of sea surface wind speeds. Part I: theory and seawinds observations. *Journal of Climate*, 19, 497–520.
- Morgan, E. C., Lackner, M., Vogel, R. M., & Baise, L. G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1), 15–26.
- Northrop, P., Attalides, N., & Jonathan, P. (2016a). Cross-validators extreme value threshold selection and uncertainty with application to storm severity. *J. R. Stat. Soc. C*.doi: 10.1111/rssc.12159.
- Northrop, P., Jonathan, P., & Randell, D. (2016b). Threshold modeling of nonstationary extremes. In Dey, D., & Yan, J. (Eds.), *Extreme value modeling and risk analysis: Methods and applications*. Chapman and Hall / CRC, Boca Raton, USA, 87–108.
- Ortego Martinez, M. I., Egozcue Rubi, J. J., & Tolosana Delgado, R. (2011). Bayesian trend analysis of extreme wind using observed and hindcast series off Catalan coast, NW Mediterranean Sea. *Natural Hazards and Earth System Sciences - Discussion*, 2, 799–824.
- Oumow, B., de Carvalho, M., & Davison, A. C. (2012). Bayesian P-spline mixture modeling of extreme forest temperatures. Retrieved from www.mat.puc.cl/~mdcarvalho.
- Prevosto, M., Krogstad, H. E., & Robin, A. (2013). Probability distributions for maximum wave and crest heights. *Coastal Engineering*, 40, 329–360.
- Raghupathi, L., Randell, D., Ewans, K., & Jonathan, P. (2008). Fast computation of large scale marginal extremes with multi-dimensional covariates. (Accepted for publication in Comp. Stat. Data Anal., Retrieved from www.lancs.ac.uk/~jonathan).
- Randell, D., Feld, G., Ewans, K., & Jonathan, P. (2015). Distributions of return values for ocean wave characteristics in the South China Sea using directional-seasonal extreme value analysis. *Environmetrics*, 26, 442–450.
- Reistad, M., Breivik, O., Haakenstad, H., Aarnes, O. J., Furevik, B. R., & Bidlot, J.-R. (2006). A high-resolution hindcast of wind and waves for the North sea, the Norwegian sea, and the Barents sea. *Journal of Geophysical Research*, 116, 1–18.
- Renard, B., Lang, M., & Bois, P. (1998). Statistical analysis of extreme events in a nonstationary context via a Bayesian framework. Case study with peak-over-threshold data. *Stochastic Environmental Research and Risk Assessment*, 21, 97–112.
- Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat*, 10, 33–60.
- Scotto, M., & Guedes-Soares, C. (2015). Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Engineering*, 40, 313–327.
- Scotto, M., & Guedes-Soares, C. (2014). Bayesian inference for long-term prediction of significant wave height. *Coastal Engineering*, 54, 393–400.
- Smith, E. L., & Walshaw, D. (2003). Modelling bivariate extremes in a region. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting, Dedicated to Dennis V. Lindley*.
- Smith, R. L., & Naylor, J. C. (2015). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society Series C*, 36, 358–369.
- Stephenson, A. (2016). Bayesian inference for extreme value modelling. In Dey, D., & Yan, J. (Eds.), *Extreme value modeling and risk analysis: Methods and applications*. Boca Raton, USA: Chapman and Hall / CRC, 257–280.
- Ugarte, M. D., Goicoa, T., Etxeberria, J., & Militino, A. F. (2012). Projections of cancer mortality risks using spatio-temporal P-spline models. *Statistical Methods in Medical Research*, 21, 545–560.

- Xifara, T., Sherlock, C., Livingstone, S., Byrne, S., & Girolami, M. (2014). Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Statistics & Probability Letters*, 91(2002), 14–19.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Randell, D., Turnbull, K., Ewans, K., and Jonathan, P. (2016), Bayesian inference for nonstationary marginal extremes, *Environmetrics*. doi: 10.1002/env.2403

## APPENDIX

For the models described in Section 3 for event size and rate of occurrence, with parameter set  $\Omega$ , full conditional distributions to be sampled are

$$\begin{aligned} f(\tau|\mathbf{y}, \Omega \setminus \tau) &\propto f(\mathbf{y}|\tau, \Omega \setminus \tau) \times f(\tau) \\ f(\boldsymbol{\beta}_\eta|\mathbf{y}, \Omega \setminus \boldsymbol{\beta}_\eta) &\propto f(\mathbf{y}|\boldsymbol{\beta}_\eta, \Omega \setminus \boldsymbol{\beta}_\eta) \times f(\boldsymbol{\beta}_\eta|\boldsymbol{\delta}_\eta, \lambda_\eta) \\ f(\lambda_\eta|\mathbf{y}, \Omega \setminus \lambda_\eta) &\propto f(\boldsymbol{\beta}_\eta|\boldsymbol{\delta}_\eta, \lambda_\eta) \times f(\lambda_\eta) \end{aligned}$$

where  $\eta$  is each of  $\rho$ ,  $\alpha$ ,  $\gamma$ ,  $\nu$ , and  $\xi$  in turn, and the distribution of  $\boldsymbol{\delta}_\eta$  is assumed known (Section 3.4). We generate a sample from the posterior distribution of  $\Omega$  by sampling sequentially from the full conditional distributions for  $\tau$ , then the triplets of vectors  $\{\boldsymbol{\beta}_\eta, \lambda_\eta, \boldsymbol{\delta}_\eta\}$  for  $\eta \in \{\rho, \alpha, \gamma, \xi, \nu\}$  in order, a large number of times. By construction, the full conditional for  $\lambda_\eta$  is available in closed form for all choices of  $\eta$  (Brezger and Lang 2004), and hence, Gibbs sampling is used. Sampling  $\boldsymbol{\delta}_\eta$  from its (prior) distribution is also straightforward. Sampling from full conditionals for  $\tau$  and  $\boldsymbol{\beta}_\eta$  for  $\eta \in \{\rho, \alpha, \gamma, \xi, \nu\}$  is achieved using MH in Gibbs. For any parameter (or vector of parameters)  $\omega$ , we accept transition from state  $\omega$  to  $\omega^*$  with probability

$$\min \left( 1, \frac{f(\omega^*|\Omega \setminus \omega^*)g(\omega^* \rightarrow \omega)}{f(\omega|\Omega \setminus \omega)g(\omega \rightarrow \omega^*)} \right)$$

for proposal density  $g$ . For sampling of spline parameter vectors, correlated proposals are essential to reasonable chain mixing especially for larger problems. Proposals are generated using in one of two ways. For  $\eta \in \{\alpha, \gamma\}$ , random walk MH proposals  $\boldsymbol{\beta}_\eta^*$  given current state  $\boldsymbol{\beta}_\eta$  take the form  $\boldsymbol{\beta}_\eta^* = \boldsymbol{\beta}_\eta + (\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\kappa_\eta\epsilon$ , where  $\epsilon \sim N(0,1)$  and  $\kappa_\eta$  is adjusted to regulate proposal acceptance rate. This form of proposal is motivated by consideration of smoothing splines or the ridge regression  $\mathbf{y} = \mathbf{B}\boldsymbol{\beta} + \epsilon$  where  $\epsilon \sim N(0, \mathbf{I})$ . If  $\boldsymbol{\beta}$  has normal prior  $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{P}^{-1})$ , its posterior variance is  $(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}$ . Proposals for  $\tau$  are generated using a simple Gaussian random walk, again with variance adjustment on the fly to achieve suitable acceptance rate. Note

that proposal densities for  $\tau$  and  $\beta_\eta$  ( $\eta \in \{\alpha, \gamma\}$ ) are symmetric such that  $g(\omega^* \rightarrow \omega) = g(\omega \rightarrow \omega^*)$ , with resulting simplification of the expression for acceptance probability above. For  $\eta \in \{\rho, \nu, \xi\}$ , we propose  $\beta_\eta^*$  using derivatives of the target distribution at the current sample. This promotes proposals in regions of higher probability, at the additional computational cost of computing necessary derivatives and matrix inverses. Asymmetric random walk proposals are made as  $\beta_\eta^* = \beta_\eta + \frac{1}{2}\kappa_\eta^2 H^{-1} G + \kappa_\eta H^{-1/2} \epsilon$ , where  $H$  is the second derivative matrix of negative log full conditional for  $\beta_\eta$  with respect to  $\beta_\eta$  evaluated at the current state, and  $G$  is the corresponding gradient. The value of  $\kappa_\eta$  is set to achieve suitable acceptance rate. We note the obvious similarity between mMALA and the back-fitting algorithm (e.g. Green and Silverman 2008, Davison 2000) for maximum likelihood estimation. As in the back-fitting algorithm, the constraint  $\xi > -0.5$  is necessary to apply the mMALA algorithm in the GP case. Further computational details and expressions for derivatives are given in Jones et al. (2016).

The table below gives prior distributional forms and parameter values used for  $\tau$  and elements of  $\delta$  and  $\lambda$ .

Parameter	Distribution	$a$	$b$
$\tau$	$B(a, b)$	100	100
$\delta$	$\Gamma(a, b)$	$\frac{1}{2}$	$\frac{1}{2}$
$\lambda$	$\Gamma(a, b)$	$10^{-3}$	$10^{-3}$

Prior distributions for  $\tau$  and components of  $\delta$  and  $\lambda$ .

We note that the prior for  $\tau$  is relatively informative compared with those of other parameters, primarily to retain a sufficient proportion of the sample below extreme value threshold  $\psi$  for reasonable confidence in predictions of return values, as discussed in the Supporting Information.