Flexible covariate representations for extremes

E. Zanini[a,b], E. Eastoe[b], M.J. Jones[c], D. Randell[c] and P. Jonathan[a,b]*

Summary: Environmental extremes often show systematic variation with covariates. Three different non-parametric descriptions (penalised B-splines, Bayesian adaptive regression splines, and Voronoi partition) for the dependence of extreme value model parameters on covariates are considered. These descriptions take the generic form of a linear combination of basis functions on the covariate domain, but differ (a) in the way that basis functions are constructed and possibly modified, and potentially (b) by additional penalisation of the variability (e.g. variance or roughness) of basis coefficients, for a given sample, to improve inference. The three representations are used to characterise variation of parameters in a non-stationary generalised Pareto model for the magnitude of threshold exceedances with respect to covariates. Computationally-efficient schemes for Bayesian inference are used, including mMALA (Riemann manifold Metropolis-adjusted Langevin algorithm) and reversible jump. A simulation study assesses relative performance of the three descriptions in estimating the distribution of the T-year maximum event (for arbitrary T greater than the period of the sample) from a peaks over threshold extreme value analysis with respect to a single periodic covariate. The three descriptions are also used to estimate a directional tail model for peaks over threshold of storm peak significant wave height at a location in the northern North Sea.

Keywords: extreme; non-stationarity; covariate; Bayesian; return value; ocean wave; storm severity; Poisson; generalised Pareto; spline; Voronoi; MCMC; reversible jump; mMALA

Third submission to Environmetrics, February 2020.

1. INTRODUCTION

Different aspects of covariate effects in marginal extreme value modelling have been researched for many years; Carter and Challenor (1981), Davison and Smith (1990), Coles and Walshaw (1994), Robinson and Tawn (1997), Scotto and Guedes-Soares (2000), Anderson et al. (2001), Chavez-Demoulin and Davison (2005), Fawcett and Walshaw (2007), Mendez et al. (2008), Northrop and Jonathan (2011) and Randell et al. (2016) are useful references. The ultimate goal of much of this

[a]*Shell Projects & Technology, London SE1 7NA, United Kingdom.*
[b]*Department of Mathematics & Statistics, Lancaster University, Lancaster LA1 4YF, United Kingdom.*
[c]*Shell Projects & Technology, 1031 HW Amsterdam, The Netherlands.*
*Correspondence to: Philip Jonathan, Shell Projects & Technology, London SE1 7NA, United Kingdom. E-mail: philip.jonathan@shell.com

This paper has been submitted for consideration for publication in *Environmetrics*

research is estimation of extreme quantiles (or return values or equivalent) for use in quantification of environmental risk. For instance, Jonathan et al. (2008) and Jones et al. (2016) demonstrate the impact of covariate effects on quality of estimation for extreme quantiles of ocean storm severity.

We consider non-stationary marginal extreme value models for ocean storm severity. We characterise statistical properties of peaks over threshold of storm severity, specifically of storm peak significant wave height, which varies with a single covariate, storm direction. The extreme value threshold, itself non-stationary with respect to storm direction, is assumed pre-specified. We then follow Chavez-Demoulin and Davison (2005) in adopting a non-stationary Poisson model for rate of occurrence of threshold exceedances and a generalised Pareto model for size of threshold exceedances conditional on the number of exceedances. One of three forms of covariate representations for the parameters of the generalised Pareto and Poisson models is then assumed; for clarity, we focus discussion on the generalised Pareto model.

The three covariate representations introduced in Section 2 take a generic form: they can be expressed in terms of a linear combination of basis functions on the covariate domain. The covariate representations differ in the way that basis functions are constructed and potentially modified, and also by possible penalisation of the variability (i.e. roughness) of basis coefficients during inference. A key feature of these covariate representations is that they require only the specification of a non-parametric regression relationship between model parameter and covariate. Further, basis functions have local support on the covariate domain, providing numerical stability and efficient computation of basis coefficients (e.g. Currie et al. 2006, Bodin and Sambridge 2009). Moreover, extension to the kinds of multidimensional covariates typically encountered in oceanography and ocean engineering is straightforward in terms of tensor products of (marginal) basis representations (e.g. Raghupathi et al. 2016). Further, the covariate representation can be tailored for a specific application. In some cases, the functional form for covariate dependence may be suggested by physical consideration (e.g. Davison and Smith 1990). When sample size is small, evidence for complex covariate effects will be limited; a simple penalised piecewise-constant covariate representation (e.g. Ross et al. 2018) may be preferable, providing a flexible semi-parametric representation considerably more easily implemented than competitors. For periodic covariates, such as those encountered in this study, some authors (e.g. Robinson and Tawn 1997, Ewans and Jonathan 2008) adopt a Fourier basis for covariates; computationally, a Fourier basis for covariates is problematic because it does not have compact support on the covariate domain, increasing the complexity of inference. In this article, we focus

on a single covariate for ease of description. In a companion paper (Zanini et al. 2020) we extend discussion to two-, three- and higher-dimensional covariates.

Bayesian inference (see Section 3) provides an intuitive framework for environmental applications of extreme value analysis. Bayesian inference allows coherent incorporation of prior knowledge, thorough uncertainty quantification, predictive inference, and when carefully designed, computationally-efficient inference. Stephenson (2016) provides a useful review. Different Markov chain Monte Carlo (MCMC) algorithms (e.g. Gamerman and Lopes 2006) can be used to sample from posterior distributions. Here, we use both 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). Since some of the covariate representations require that the number of parameters present changes, we also utilise reversible-jump inference (Green 1995, Richardson and Green 1997); algorithms implemented draw on the work of Biller (2000), DiMatteo et al. (2001), Brezger and Lang (2006), Bodin et al. (2009), Bodin and Sambridge (2009) and Randell et al. (2015).

Reversible-jump inference has already been used in extreme value modelling, but not for the estimation of flexible covariate representations considered here. Bottolo et al. (2003) adopts a mixture model for non-homogeneous Poisson process parameters. Boldi and Davison (2007) proposes a mixture of Dirichlet distributions to represent spectral functions for multivariate extremes. El Adlouni and Ouarda (2009) and Ouarda and El Adlouni (2011) consider linear additive covariate representations. The current methodology has many shared features with vector generalised additive models (VGAM, Yee and Wild 1996), and developments of low-rank penalised regression smoothers by Wood and co-authors (Wood 2003, 2004, 2011, Wood et al. 2016) exploiting Laplace approximations for efficient frequentist inference. We elaborate on these similarities in the discussion (Section 6).

Objectives

The objective of this work is to assess the relative merits of three covariate representations, primarily in terms of quality of estimation of the distribution of the size of the largest event observed (the T-year maximum) in a long return period of T years. We compare estimates for T-year maxima with

those obtained using simpler covariate representations. We consider relative ease of implementation of the MCMC algorithms required, complexity of the resulting models, and relative computational efficiency.

Layout of article

The article is structured as follows. Section 2 introduces the three covariate representations, namely penalised B-splines (or P-splines), Bayesian adaptive regression splines (BARS) and Voronoi partition. Section 3 describes Bayesian inference for the shape and scale parameters of a generalised Pareto model, including specification of sample likelihood and prior structures. It also describes the MCMC procedures used to sample from the joint posterior distribution of parameters. Section 4 illustrates application of the covariate representations to extreme value modelling of storm peak significant wave height with respect to storm direction, at a location in the northern North Sea, including estimation of Poisson rate of extreme value threshold exceedance, estimation of generalised Pareto parameters, and estimation of the distribution of the size of the T-year maximum. Section 5 outlines a simulation study conducted to assess predictive performance of the covariate representations in estimating the distributions of T-year maxima over both the complete covariate domain, and on subintervals of that domain. Section 6 provides discussion. Technical details are relegated to the Supplementary Material where possible.

2. COVARIATE REPRESENTATIONS

We are interested in modelling a response Y, the characteristics of which vary systematically with covariate θ. We assume that the conditional density $f(Y|\theta)$ has a known, fixed functional form parameterised in terms of p variables $\{\eta_j(\theta)\}_{j=1}^p$, such that each of the functions $\eta_j(\theta)$ $(j = 1, 2, ..., p)$ varies systematically with θ on covariate domain $\mathcal{D}_\theta$. We define each η_j on an arbitrary index set $\mathcal{I}_{\theta j}$ of m_j unique locations on $\mathcal{D}_\theta$; $\mathcal{I}_{\theta j}$ might correspond to a regular lattice on $\mathcal{D}_\theta$, or to the subset of unique covariate values in a particular sample for analysis. We write the set $\{\eta_{j;s}\}_{s=1}^{m_j}$ in vector form as $\boldsymbol{\eta}_j$. For simplicity, we take the covariate domain to be an interval $[L, U)$ of the real line, and assume that the covariate is periodic on this interval. This description is therefore appropriate for a single directional or seasonal covariate. The use of higher-dimensional covariates has been considered

by the current authors elsewhere (e.g. Randell et al. 2016, Feld et al. 2019); extensions of the current work to higher-dimensional covariates will be reported in a companion article (Zanini et al. 2020).

Since the description of covariate representations applies to all variables η_j, for the remainder of this section we suppress subscript j for clarity. On the index set $\mathcal{I}_\theta = \{\theta_s\}_{s=1}^m$, we then assume that

$$
\begin{aligned}
\eta_s &= \sum_{k=1}^n B_{sk}\beta_k, s = 1, 2, ..., m, \text{ or} \\
\boldsymbol{\eta} &= \boldsymbol{B}\boldsymbol{\beta}
\end{aligned}
\tag{1}
$$

where $\boldsymbol{B} = \{B_{sk}\}_{s=1;k=1}^{m;n}$ forms a suitable basis for the covariate domain with real-valued elements, and allows evaluation of η at each point on $\mathcal{I}_\theta$ in terms of $n > 0$ real-valued basis coefficients $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^n$. The focus of statistical inference for given $\boldsymbol{B}$ becomes estimation of coefficient vector $\boldsymbol{\beta}$. A given choice of $\boldsymbol{B}$ leads to a covariate representation with particular statistical and computational characteristics. It is likely that, for a given inference task, some covariate representations will prove more attractive than others, in both computational and physical terms. We consider three choices of basis, leading to covariate representations in terms of P-splines (e.g. Marx and Eilers 1998, Eilers et al. 2006, Brezger and Lang 2006, Eilers and Marx 2010), Bayesian adaptive regression splines (e.g. Denison et al. 1998, Biller 2000, DiMatteo et al. 2001, Wallstrom et al. 2008), and Voronoi partition (e.g. Denison et al. 2002, Costain 2008, Bodin and Sambridge 2009).

2.1. B-splines

A spline is a piecewise polynomial function. Its polynomial pieces are connected at knots. Splines and their derivatives are typically constructed to be continuous functions. B-splines (or basis splines) provide basis functions for splines of a particular order; all splines can therefore be constructed as a unique linear combination of B-splines. B-spline bases of different orders d in one-dimension can be constructed for arbitrary knot locations $r_1 \leq r_2 \leq r_m \in (L, U)$, using the Cox - de Boor recursion formula. A B-spline of order d is defined using exactly $d + 1$ knot locations. Construction of B-spline bases for periodic covariate θ is relatively straightforward; a brief overview is given in the Supplementary Material. For a periodic covariate, the number n of B-spline coefficients β is equal to the number of knot locations on the covariate domain, independently of spline order, and the covariate representation for any model parameter η is defined by the set $\{r_k, \beta_k\}_{k=1}^n$. For instance, we might use a fixed set of n regularly-spaced knots on (L, U); this is the approach used in the penalised

B-spline (or P-spline) covariate representation. More generally, we might place knots arbitrarily, and allow the number of knots n to vary to improve inference; this approach is the motivation for Bayesian adaptive regression splines (BARS, see Section 3.3).

For a P-spline model, the number n of spline knots is required to be large, so that the resulting set $\{\eta_s\}_{s=1}^n$ easily captures the variability of $\eta_j(\theta)$. To prevent over-fitting, the values of spline coefficients β are penalised to produce sets $\{\eta_s\}$ with an appropriate level of smooth variation (or roughness) as a function of covariate. In a Bayesian inference, the smoothness requirement is specified in terms of a prior distribution for $\boldsymbol{\beta}$, discussed in Section 3.2. Penalisation can also be used in a BARS model, although it is not strictly necessary.

2.2. Voronoi partition

A Voronoi partition (or Dirichlet tessellation) partitions the covariate domain into convex polygons (or cells) such that each polygon contains exactly one generating point (or centroid) and every point in a given cell is closer (by some definition) to its centroid than to any other centroid. For a set of n centroid locations $\{r_k\}_{k=1}^n \in (L, U)$, a periodic one-dimensional Voronoi partition partitions $[L, U)$ into n cells. The kth cell consists of every point on the covariate domain whose (shortest wrapped) Euclidean distance from r_k is less than or equal to that from any other $r_{k'}$ $(k' \neq k)$. On the index set $\mathcal{I}_\theta$, the covariate representation in Equation 1 applies, with $B_{sk} = 1$ if θ_s belongs to the kth cell, and $B_{sk} = 0$ otherwise. The covariate representation is defined by set $\{r_k, \beta_k\}_{k=1}^n$. Note the equivalence, in one dimension, of a Voronoi partition and an order-one (i.e. $d = 1$) BARS characterisation.

3. INFERENCE

Assume now that observations of $Y|\theta$ follow a generalised Pareto distribution with parameters shape $\xi(\theta)$, (modified) scale $\nu(\theta)$ and pre-specified threshold $\psi(\theta)$. We seek to estimate functional forms of $\xi(\theta)$ and $\nu(\theta)$ using Bayesian inference, and thereby estimate return values for Y. We consider estimation using the three approaches outlined in Section 2, namely P-splines, BARS and Voronoi partition.

To perform Bayesian inference, we need to specify the sample likelihood (Section 3.1), a prior model structure (Section 3.2), and a MCMC algorithm to sample from the appropriate target posterior

distribution (Section 3.3). The latter may include exploring models of different sizes i.e. varying $\{n_j\}$.

For clarity and brevity, we describe inference only for the non-stationary generalised Pareto distribution of threshold exceedance in some detail in the paragraphs following. In practical applications, models for non-stationary threshold itself, and non-stationary rate of threshold exceedance would also need to be inferred (Section 4). A typical full inference procedure is outlined in Section A4 of the Supplementary Material, following the description of Ross et al. (2017).

3.1. Sample likelihood

For convenience, we denote an observed sample of independent threshold exceedances Y by $\boldsymbol{y} = \{y_i\}_{i=1}^{n_Y}$ and corresponding covariate values. The sample likelihood can be written

$$L(\boldsymbol{\xi}, \boldsymbol{\nu}|\boldsymbol{y}) = f_{\mathrm{GP}}(\boldsymbol{y}|\boldsymbol{\xi}, \boldsymbol{\nu}) = \prod_{i=1}^{n_Y} f_{\mathrm{GP}}\left(y_i|\xi_{A(i)}, \nu_{A(i)}\right)$$

where mapping $s = A(i)$ allocates the ith observation from the sample to exactly one member $\theta_s \in \mathcal{I}_\theta$ of the index set of covariates, and each member of the index set can represent the covariate allocation of none, one or multiple observations. Note that $\mathcal{I}_\theta$ is assumed to be common for all parameters estimated, containing m values, so that $\boldsymbol{\xi} = \{\xi_s\}_{s=1}^m$ with $\xi_s = \xi(\theta_s)$, and $\boldsymbol{\nu} = \{\nu_s\}_{s=1}^m$ with $\nu_s = \nu(\theta_s)$. The density f_{GP} of the generalised Pareto distribution is

$$f_{\mathrm{GP}}(y|\xi, \nu) = \frac{1}{\sigma}\left[\left(1 + \frac{\xi}{\sigma}(y - \psi)\right)^{-1/\xi - 1}\right]_+ \quad \text{with } \nu = \sigma(1 + \xi)$$

where $\xi \in (-\infty, \infty)$ and $\sigma > 0$. For any real x, $[x]_+ = x$ when $x > 0$ and 0 otherwise. When $|\xi| \le 10^{-6}$, we assume f_{GP} has the exponential form $(1/\sigma)\exp(-(y - \psi)/\sigma)$ appropriate for $\xi \to 0$. Equation 1 relates values of ξ and ν on $\mathcal{I}_\theta$ to the corresponding basis coefficients β.

Next we provide a general description of the inference appropriate for all covariate representations, noting variations as necessary. The description is also relevant for estimation of extreme value threshold and rate of threshold exceedance, with appropriate change of likelihood (and number of model parameters p). With reference to Section 2 and the generalised Pareto case, the full parameter set to be estimated is $\Omega = \{\Omega_j\}_{j=1}^p$ where $\Omega_j = (n_j, \boldsymbol{r}_j, \lambda_j, \boldsymbol{\beta}_j)$ and $p = 2$; for definiteness we assume $j = 1$ refers to ξ and $j = 2$ to ν. Further, for the jth variable, n_j is the number of spline knots (or

Voronoi centroids), $\boldsymbol{r}_j = \{r_{jk}\}_{k=1}^{n_j}$ are knot (or centroid) locations, λ_j is the roughness parameter and $\boldsymbol{\beta}_j = \{\beta_{jk}\}_{k=1}^{n_j}$ are spline (or Voronoi) coefficients. Throughout, we use the convention that a set defined using curly braces (e.g. $\{a_\ell, b_\ell\}_{\ell=1}^{n_\ell}$ or $\{a_\ell, b_\ell\}_\ell$ or $\{a_\ell, b_\ell\}$ in brief) includes all possible pairs a_ℓ, b_ℓ (i.e. for all of $\ell = 1, 2, ..., n_\ell$, such that $\{a_\ell, b_\ell\} = \cup_{\ell=1}^{n_\ell}(a_\ell, b_\ell)$) whereas a set (or vector) defined using round brackets refers to a specific value of ℓ only; thus (a_ℓ, b_ℓ) includes two elements, one each of a_ℓ and b_ℓ for a single specific value of ℓ. Inference is performed using modified scale ν, since maximum likelihood estimators for ξ and ν are asymptotically independent (see e.g. Cox and Reid 1987, Hosking and Wallis 1987). Results below are discussed in terms of ν.

3.2. Prior structure

Model prior probability density $f(\Omega)$ is

$$f(\Omega) = \prod_{j=1}^{p} f(\Omega_j)$$

where $f(\Omega_j)$ is the prior probability density for the jth variable. $f(\Omega_j)$ can be factorised using the chain rule as

$$
\begin{aligned}
f(\Omega_j) &= f(n_j, \boldsymbol{r}_j, \lambda_j, \boldsymbol{\beta}_j) && (2)\\
&= f(n_j)f(\boldsymbol{r}_j|n_j)f(\lambda_j|\boldsymbol{r}_j, n_j)f(\boldsymbol{\beta}_j|\lambda_j, \boldsymbol{r}_j, n_j)\\
&= f(n_j)f(\boldsymbol{r}_j|n_j)f(\lambda_j)f(\boldsymbol{\beta}_j|\lambda_j, \boldsymbol{r}_j, n_j).
\end{aligned}
$$

where prior densities $f(n_j)$, $f(\boldsymbol{r}_j|n_j)$, $f(\lambda_j)$ and $f(\boldsymbol{\beta}_j|\lambda_j, \boldsymbol{r}_j, n_j)$ take different forms for the three covariate representations.

P-splines

For penalised B-splines, the number of spline knots n_j $(= 1, 2, ...)$ is fixed, and locations $\boldsymbol{r}_j$ of spline knots $(r_{jk} \in [a_r, b_r) = [L, U) = [0, 360)^\circ, \ k = 1, 2, ..., n_j)$ are pre-specified. Prior specification thus involves only $f(\lambda_j)$ and $f(\boldsymbol{\beta}_j|\lambda_j, \boldsymbol{r}_j, n_j)$. The first of these is

$$f(\lambda_j) = f_{\text{Gmm}}(\lambda_j|a_{\lambda j}, b_{\lambda j}), \quad \lambda_j > 0,$$

a gamma distribution with pre-specified hyper-parameters $a_{\lambda j} > 0$ and $b_{\lambda j} > 0$, typical values for which are given in the Supplementary Material. Ideally, the vector of knot coefficients $\boldsymbol{\beta}_j \in \mathbb{R}^{n_j}$ would take a Gaussian conditional density, so that $f(\boldsymbol{\beta}_j | \lambda_j, \boldsymbol{r}_j, n_j) = ((2\pi)^{n_j/2} |\boldsymbol{P}_j|^{1/2})^{-1} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_j' \boldsymbol{P}_j \boldsymbol{\beta}_j\right)$. However for P-splines

$$\boldsymbol{P}_j = \boldsymbol{P}_j(\lambda_j) = \lambda_j \boldsymbol{D}_j' \boldsymbol{D}_j$$

where $\boldsymbol{D}_j$ is a $n_j \times n_j$ differencing matrix (wrapped for periodic covariate domain) with integer elements, so that $\boldsymbol{D}_j \boldsymbol{\beta}_j$ yields a vector of differences between consecutive elements of $\boldsymbol{\beta}_j$. Since for P-splines $\boldsymbol{P}_j$ is rank-deficient, an improper conditional prior for $\boldsymbol{\beta}_j \in \mathbb{R}^{n_j}$ is used, with density

$$f(\boldsymbol{\beta}_j | \lambda_j, \boldsymbol{r}_j, n_j) \propto |\lambda_j|^{\mathrm{rank}(\boldsymbol{P}_j)/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_j' \boldsymbol{P}_j \boldsymbol{\beta}_j\right).$$

This choice of prior for $\boldsymbol{\beta}_j$ is motivated by the work of Lang and Brezger 2004 and Brezger and Lang (2006), who also provide guidance (based on the work of Hobert and Casella 1996) for the specification of the gamma prior for λ_j to ensure a proper posterior for $\boldsymbol{\beta}_j$. Larger values of λ_j favour choices of $\boldsymbol{\beta}$ for which $|\boldsymbol{D}_j \boldsymbol{\beta}_j|$ is nearer zero, corresponding to smoother variation of $\eta_j(\theta)$. Different orders of differencing result in different roughness penalties; a choice of zero-order difference corresponds to setting $\boldsymbol{D}_j$ to an identity matrix, amounting to a ridge-type penalty. A first-order difference is typically adequate.

BARS

For Bayesian adaptive regression splines, we must specify all four terms in Equation 2. The number n_j of knots per variable is given a Poisson prior with parameter $a_n > 0$

$$f(n_j) = f_{\mathrm{Poisson}}(n_j | a_n)$$

and the prior location of each knot is taken to be uniformly distributed on the covariate domain $[a_r, b_r)$, so that

$$f(\boldsymbol{r}_j | n_j) = \frac{n_j!}{V_r^{n_j}}$$

where $V_r = b_r - a_r$, and the $n_j!$ term is included since permutations of the set of knot locations are indistinguishable. The prior density for the vector of knot coefficients $\boldsymbol{\beta}_j$ and the roughness parameter

λ_j are the same as those for P-splines; a zero-order difference penalty is used in the prior for $\boldsymbol{\beta}_j$. DiMatteo et al. (2001) suggest setting $\boldsymbol{P}_j \propto \boldsymbol{B}'_j\boldsymbol{B}_j$, allowing the prior for $\boldsymbol{\beta}_j$ to reflect the correlation structure of the spline basis, but this was not found to be particularly useful here.

Voronoi partition

For Voronoi partition, the number n_j of cells per variable is typically given a Poisson prior, as for BARS; however, in our experience, a uniform prior also performs well in practice. Similarly to BARS, the prior location of each cell centroid is taken to be uniformly distributed on the covariate domain. We adopt a Gaussian conditional prior for cell coefficients $\boldsymbol{\beta}_j$ with zero-order differences; as for BARS, a gamma prior for λ_j is used, although in general there is no need to include a roughness penalty for the Voronoi covariate representation.

Note that Bodin and Sambridge (2009) assumes that $\boldsymbol{\beta}_j$ is uniformly-distributed on support $(a_{\beta j}, b_{\beta j}) \subset \mathbb{R}$, so that $f(\boldsymbol{\beta}_j|\lambda_j, \boldsymbol{r}_j, n_j) = f(\boldsymbol{\beta}_j|n_j) = V_{\beta j}^{-n_j}$ where $V_{\beta j} = b_{\beta j} - a_{\beta j}$.

3.3. MCMC scheme

The posterior distribution of model parameters Ω is not available in closed form. Posterior inference is therefore made using reversible-jump Markov chain Monte Carlo (RJ-MCMC, Green 1995) by sampling from full conditionals as outlined below. Two types of proposal schemes are used, the first to sample from the complete set of full conditional distributions when model size does not change (applicable to P-splines, BARS and Voronoi), and the second to explore models of different size using dimension-jumping proposals (for BARS and Voronoi only).

Sampling from full conditionals for given model size

For a model of given size, we follow the inference scheme described in Randell et al. (2016). When full conditionals are available in closed form, we use Gibbs sampling. Otherwise, we use MH within Gibbs. When the distribution to be sampled is particularly problematic (e.g. posteriors for generalised Pareto shape and adjusted scale), we exploit gradient and curvature information if possible to generate MH proposals efficiently. In particular, full conditional distributions for roughness coefficients $\{\lambda_j\}$ are available in closed form, permitting Gibbs sampling (see Randell et al. 2016). For fixed $\{n_j\}$, at each iteration of the MCMC chain, we sample in turn from the full conditional distributions of coefficients

$\{\boldsymbol{\beta}_j\}$, roughness penalties $\{\lambda_j\}$, and knot locations $\{\boldsymbol{r}_j\}$. We experimented with different orderings of parameters and found there to be little difference in performance. Further details are given in the Supplementary Material.

Reversible jump

RJ-MCMC allows transitions between a current state with a given number of parameters, and a proposed state with greater (or fewer) parameters, by augmenting the set of parameters for the current (or proposed) state with extra auxiliary random variables. As a result, the number of parameters in the (possibly augmented) current and proposed states is equal, and a deterministic bijective mapping can be defined between them. An appropriate expression for transition acceptance probability is required to ensure that a detailed balance condition is satisfied, so that the MCMC eventually samples from the desired "trans-dimensional" target distribution.

In more detail, for current state $\Omega = (n_j, \boldsymbol{r}_j, \lambda_j, \boldsymbol{\beta}_j)$, and for each variable j in turn, we propose reversible dimension-jumping transitions to state $\Omega^* = (\Omega \setminus \omega, \omega^*)$ where triplet $\omega = (n_j, \boldsymbol{\beta}_j, \boldsymbol{r}_j)$ in the current state changes to triplet $\omega^* = (n_j^*, \boldsymbol{\beta}_j^*, \boldsymbol{r}_j^*)$ in the proposed state. The transition is accepted with probability

$$\min\left(1, \frac{f(\boldsymbol{y}|\Omega^*)}{f(\boldsymbol{y}|\Omega)} \frac{f(\omega^*)}{f(\omega)} \frac{q(\omega|\omega^*)}{q(\omega^*|\omega)} \left|\frac{\partial(\omega^{a*})}{\partial(\omega^a)}\right|\right) \tag{3}$$

where $f(\boldsymbol{y}|\Omega)$ and $f(\boldsymbol{y}|\Omega^*)$ represent the sample likelihood in the respective states, $f(\omega)$ and $f(\omega^*)$ represent priors for respective triplets; $q(\omega^*|\omega)$ is the proposal density for transition from ω to ω^*, and $q(\omega|\omega^*)$ is the proposal density for reverse transition. The final term in the expression is the Jacobian for the transformation of augmented triplets (with superscript "a"). The manner in which augmented triplets are constructed is dictated by the type of dimension-jumping transition considered: here we consider two transition types, "birth" and "death". Birth involves increasing n_j to $n_j^* = n_j + 1$, by creating a new knot (centroid) with some location and coefficient. Each of $\boldsymbol{\beta}_j^*$ and $\boldsymbol{r}_j^*$ has $n_j + 1$ elements, compared with n_j elements for $\boldsymbol{\beta}_j$ and $\boldsymbol{r}_j$. To enable proposal of reversible transitions, we augment ω with two random variables u_r and u_β so that the resulting augmented triplet $\omega^a = (n_j, \boldsymbol{\beta}_j, u_\beta, \boldsymbol{r}_j, u_r)$ has the same number of elements as ω^*; no augmentation of ω^* is necessary, and we set $\omega^{a*} = \omega^*$. Conversely, death involves reducing n_j to $n_j^* = n_j - 1$. Following a similar logic, we now augment ω^{a*}, leaving $\omega^a = \omega$. Thus we create a deterministic bijection between augmented

triplets ω^a and ω^{a*} for any transition type. The expression for the acceptance probability above then ensures that the detailed balance condition for the transition is satisfied. The distributions of u_r and u_β are discussed below.

The proposal ratio can be written more usefully as

$$
\begin{aligned}
\frac{q(\omega|\omega^*)}{q(\omega^*|\omega)} &= \frac{q(n_j, \boldsymbol{r}_j, \boldsymbol{\beta}_j | n_j^*, \boldsymbol{r}_j^*, \boldsymbol{\beta}_j^*)}{q(n_j^*, \boldsymbol{r}_j^*, \boldsymbol{\beta}_j^* | n_j, \boldsymbol{r}_j, \boldsymbol{\beta}_j)} = \frac{q(n_j|\omega^*)}{q(n_j^*|\omega)} \frac{q(\boldsymbol{r}_j|\omega^*, n_j)}{q(\boldsymbol{r}_j^*|\omega, n_j^*)} \frac{q(\boldsymbol{\beta}_j|\omega^*, n_j, \boldsymbol{r}_j)}{q(\boldsymbol{\beta}_j^*|\omega, n_j^*, \boldsymbol{r}_j^*)} \\
&= \frac{q(n_j|n_j^*)}{q(n_j^*|n_j)} \frac{q(\boldsymbol{r}_j|\boldsymbol{r}_j^*, n_j^*, n_j)}{q(\boldsymbol{r}_j^*|\boldsymbol{r}_j, n_j, n_j^*)} \frac{q(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^*, \boldsymbol{r}_j^*, \boldsymbol{r}_j)}{q(\boldsymbol{\beta}_j^*|\boldsymbol{\beta}_j, \boldsymbol{r}_j, \boldsymbol{r}_j^*)}.
\end{aligned}
\tag{4}
$$

Sample likelihood (Section 3.1) and model prior (Section 3.2) have already been discussed, and can be easily evaluated. Specification of proposal distributions and Jacobian depend on both transition type and model type; we start with Voronoi partition, since the description is somewhat simpler.

Voronoi partition

For a Voronoi birth transition, we assume that location r^+ of the new centroid is sampled uniformly on the covariate domain so that

$$
r^+ = u_r, \text{ where } f(u_r) = \frac{1}{V_r} \text{ and } u_r \in [a_r, b_r).
$$

The corresponding coefficient $\beta^+ \in \mathbb{R}$ is sampled from a Gaussian density with standard deviation $\tau_{\beta j} > 0$, centred at the value of coefficient β_{j+} at the location of r^+ in the current state, so that

$$
\beta^+ = \beta_{j+} + u_\beta, \text{ where } u_\beta \sim N(0, \tau_{\beta j}^2).
$$

Using these relationships, as shown in the Supplementary Material, the proposal ratio becomes

$$
\left[\frac{q(\omega|\omega^*)}{q(\omega^*|\omega)} \right]_{\text{VrnBrt}} = \frac{p_D}{p_B} \frac{V_r}{n_j^*} \sqrt{2\pi} \tau_{\beta j} \exp\left(\frac{(\beta^+ - \beta_{j+})^2}{2\tau_{\beta j}^2} \right)
$$

where $n_j^* = n_j + 1$ and VrnBrt refers to Voronoi birth. Constants $p_B \in [0,1]$ and $p_D \in [0,1]$ are probabilities of attempting a birth and death, which we are free to choose in most cases; sometimes, it is useful to make these functions of the model size n_j. The Jacobian in this case is unity.

For a Voronoi death transition, one of the $n_j + 1$ centroids is eliminated at random. The proposal ratio is the inverse of that given above, with β^+ replaced by the coefficient β^- of the eliminated centroid, and β_{j+} by the coefficient β_{j-}^* for the Voronoi cell within which the location of the eliminated

centroid falls in the proposed state. Specifically

$$\left[\frac{q(\omega|\omega^*)}{q(\omega^*|\omega)}\right]_{\text{VrnDth}} = \frac{p_B}{p_D}\frac{n_j}{V_r}\frac{1}{\sqrt{2\pi}\tau_{\beta j}}\exp\left(-\frac{(\beta^- - \beta^*_{j-})^2}{2\tau_{\beta j}^2}\right)$$

where VrnDth refers to Voronoi death. The Jacobian for the Voronoi death transition is again unity.

BARS

For a BARS birth transition, we assume that location r^+ of the new knot is sampled uniformly from the covariate domain as for Voronoi partition above. Given knowledge of knot locations $\boldsymbol{r}_j$ and $\boldsymbol{r}_j^*$ in the current and proposed states, we use a regression to establish the bijection between an augmented coefficient vector $\boldsymbol{\beta}_j^a = (\boldsymbol{\beta}_j, u_\beta)$ for the current state, and the vector $\boldsymbol{\beta}_j^*$ for the proposed state. Suppose that $\boldsymbol{B}_j$ is the spline basis matrix for the jth variable in the current state (with knot locations $\boldsymbol{r}_j$), such that the values of the variable on the index set of covariate values is $\boldsymbol{B}_j\boldsymbol{\beta}_j$. The corresponding spline basis in the proposed state (with knot locations $\boldsymbol{r}_j^* = (\boldsymbol{r}_j, r^+)$) is $\boldsymbol{B}_j^*$. We choose to append the new knot location to the existing location vector here, for ease of description; in reality we reorder knot locations so that the elements $\boldsymbol{r}_j^*$ create an increasing sequence. Solving the regression $\boldsymbol{B}_j^*\boldsymbol{\beta}_j^* = \boldsymbol{B}_j\boldsymbol{\beta}_j$ provides an estimate $\widehat{\boldsymbol{\beta}}_j^*$ for $\boldsymbol{\beta}_j^*$ in the proposed state, with $\widehat{\boldsymbol{\beta}}_j^* = \left[(\boldsymbol{B}_j^{*\prime}\boldsymbol{B}_j^*)^{-1}\boldsymbol{B}_j^{*\prime}\boldsymbol{B}_j\right]\boldsymbol{\beta}_j = \boldsymbol{G}_j\boldsymbol{\beta}_j$. Motivated by this result, we set

$$\boldsymbol{\beta}_j^* = \left[\begin{array}{c|c} & 0 \\ & \vdots \\ \boldsymbol{G}_j & 0 \\ & 0 \\ & 1 \end{array}\right] \times \left[\begin{array}{c} \boldsymbol{\beta}_j \\ u_\beta \end{array}\right] = \boldsymbol{F}_j\boldsymbol{\beta}_j^a$$

where $u_\beta \sim N(0, \tau_{\beta j}^2)$. This expression provides a bijection between the sets $\boldsymbol{\beta}_j^a = (\boldsymbol{\beta}_j, u_\beta)$ and $\boldsymbol{\beta}_j^*$. The proposal ratio for BARS birth becomes

$$\left[\frac{q(\omega|\omega^*)}{q(\omega^*|\omega)}\right]_{\text{BrsBrt}} = \frac{p_D}{p_B}\frac{V_r}{n_j^*}\sqrt{2\pi}\tau_{\beta j}\exp\left(\frac{(\beta^+ - \boldsymbol{g}_j\boldsymbol{\beta}_j)^2}{2\tau_{\beta j}^2}\right)$$

$$= \frac{p_D}{p_B}\frac{V_r}{n_j^*}\sqrt{2\pi}\tau_{\beta j}\exp\left(\frac{u_\beta^2}{2\tau_{\beta j}^2}\right)$$

where β^+ is now the last element of $\boldsymbol{\beta}_j^*$, and $\boldsymbol{g}_j$ is the last row of $\boldsymbol{G}_j$, and BrsBrt refers to BARS birth.

For a death transition, we select a knot at random to be possibly eliminated; each knot is considered for elimination with probability $1/n_j$. Once the location vectors $\boldsymbol{r}_j$ and $\boldsymbol{r}_j^*$ are established, coefficient vector $\boldsymbol{\beta}_j^*$ is calculated using $\boldsymbol{\beta}_j^{a*} = (\boldsymbol{\beta}_j^*, u_\beta) = \boldsymbol{F}_j^{-1}\boldsymbol{\beta}_j$, where $\boldsymbol{F}_j^{-1}$ is the inverse of $\boldsymbol{F}_j$ above (for a birth). The proposal ratio for a death is essentially the inverse of that shown above for birth, as discussed in the Supplementary Material. There also we show that the Jacobian term for a birth is $|\boldsymbol{F}_j|$, and for a death is $|\boldsymbol{F}_j|^{-1}$.

Note that the Supplementary Material also outlines optional "local" knot birth and death steps (not used in the current work) which have been considered by others (e.g. DiMatteo et al. 2001).

4. APPLICATION

To illustrate the use of the three different covariate representations in extreme value analysis, we consider the problem of estimating characteristics of a severe ocean wave environment. The severity of an ocean storm, in terms of the magnitude of surface gravity waves it generates, is quantified using significant wave height (H_S). H_S is a measure of the energy content (or roughness) of the ocean surface, and can be defined as four times the standard deviation of ocean surface elevation at a spatial location for a specified period of observation. The overall severity of a storm can be usefully summarised in terms of the highest value of H_S observed, over a set of contiguous consecutive time intervals (referred to as sea states) corresponding to the time period of the storm; this value is referred to as storm peak H_S.

The application sample corresponds to storm peak H_S and associated (dominant) wave direction generated from a physical "hindcast" model of the ocean environment for a northern North Sea location, for the period September 1957 to December 2012. These data were previously considered by Randell et al. (2016). The hindcast model utilises pressure field, wind field and wind-wave generation models in particular to simulate the ocean environment, and is calibrated to observational data from instrumented offshore facilities, moored buoys and satellite altimeters in the neighbourhood of the location for a period of time, typically decades. Further details of the specific hindcast used (based on the WAM wave model) are available from Reistad et al. (2011).

Extreme sea states in the North Sea tend to be associated with winter storms originating in the Atlantic Ocean and propagating eastwards across the northern part of the North Sea. Directions of propagation of extreme seas vary considerably, due to long fetches associated with the Atlantic Ocean, Norwegian Sea and the North Sea itself, and land shadows of the British Isles and Scandinavia. Extreme sea states from the directions of Scandinavia to the east and the British Isles to the south-west are unlikely. Storm peak H_S and direction are isolated from sea-state time-series using the procedure described in Ewans and Jonathan (2008). The resulting storm peak H_S sample consists of 1076 values.

Figure 1 shows a plot of storm peak H_S (in metres) versus direction, with direction from which a storm travels expressed in degrees clockwise with respect to north. The land shadow of Norway reduces the rate and size of occurrences of events when $\theta \in (45°, 210°)$. The rate and size of occurrences increases with direction at around 210°, corresponding to Atlantic storms from the south-west passing the Norwegian headland. We therefore expect considerable directional variability in model parameter estimates. Figure 1 also shows directional extreme value threshold ψ used, with non-exceedance probability for given θ of $\tau = 0.8$.

The overall modelling strategy is outlined algorithmically in Supplementary Material A4. The threshold ψ is estimated by first characterising the distribution of $\tilde{Y}|\theta$ (for $\tilde{Y}$ = storm peak H_S) for the full sample $\tilde{\boldsymbol{y}} = \{\tilde{y}_i\}_{i=1}^{n_{\tilde{Y}}}$ of $n_{\tilde{Y}}$ storm peak H_S (as opposed to just threshold exceedances) using a non-stationary gamma distribution with likelihood

$$L_{\text{Gmm}}(\boldsymbol{\alpha}, \boldsymbol{\kappa}|\tilde{\boldsymbol{y}}) = f_{\text{Gmm}}\left(\tilde{\boldsymbol{y}}|\boldsymbol{\alpha}, \boldsymbol{\kappa}\right) = \prod_{i=1}^{n_{\tilde{Y}}} f_{\text{Gmm}}\left(\tilde{y}_i|\alpha_{A(i)}, \kappa_{A(i)}\right).$$

Here, the gamma density is $f_{\text{Gmm}}(y|\alpha, \kappa) = \kappa^{\alpha} y^{\alpha-1} \exp(-\kappa y)/\Gamma(\alpha)$ for shape $\alpha > 0$ and rate (or scale) $\kappa > 0$, and mapping A is defined as for generalised Pareto estimation. We allow α and κ parameters to vary smoothly with θ using a P-spline parameterisation for definiteness. For each θ, ψ is then defined as the quantile of the fitted distribution with pre-specified non-exceedance probability τ.

[Figure 1 about here.]

The rate of occurrence of threshold exceedances Y of ψ is assumed to follow a non-stationary Poisson process with rate ρ varying with direction. Following Chavez-Demoulin and Davison (2005) and Randell et al. (2016), we approximate the corresponding intensity on index set $\mathcal{I}_\theta$ of m directional sub-intervals of length d, taking d small enough so that ρ is assumed constant on each sub-interval.

Then, for a vector of counts $\boldsymbol{c} = \{c_s\}_{s=1}^m$ of occurrences of threshold exceedances on the index set

$$f(\boldsymbol{c}|\rho) = \exp\left(-d\sum_{s=1}^m \rho_s\right)\prod_{s=1}^m \rho_s^{c_s},$$

where $\boldsymbol{\rho} = \{\rho_s\}_{s=1}^m$ is the corresponding Poisson count rate, assumed to vary smoothly with θ using some covariate representation.

The objective of the current work is to compare different covariate representations. However the P-spline gamma model used to estimate extreme value threshold ψ is used for all extreme value fits. Hence the sample of threshold exceedances for subsequent Poisson and generalised Pareto modelling is fixed, enabling fair comparison of different covariate representations for the Poisson and generalised Pareto models. We arbitrarily choose to use the same representation for both rate and size of threshold exceedance, although this choice is not necessary; that is, when a Voronoi partition is used for the shape ξ and scale ν of the generalised Pareto distribution, it is also used for rate ρ. For conciseness of presentation, we focus on estimation of generalised Pareto parameters and resulting distributions of T-year maxima for long return periods T, incorporating models for all of threshold, rate and size.

Figure 2 illustrates the posterior distribution of ξ, ν and ρ for each of P-spline, BARS and Voronoi representations. All estimates of ρ reflect directional characteristics of H_S from Figure 1; the same is true of estimates for ν. The estimated ξ is essentially constant with direction, and relatively uncertain; these features are particularly clear for the Voronoi representation. The P-splines parameter estimates are somewhat more uncertain that those for other covariate representations. The piecewise-constant nature of the Voronoi parameter estimates is particularly clear for ξ. Overall, there is good agreement between the sets of parameter estimates for the three covariate representations.

It is important to confirm that inferences are not overly sensitive to threshold choice: in general we seek the lowest value of threshold non-exceedance probability τ such that estimates for generalised Pareto shape parameter ξ are relatively stable for all higher thresholds whilst admitting a sufficiently large sample. The choice of $\tau = 0.8$ was informed by inspection of plots of estimated ξ and median T-year maxima on θ for different τ.

[Figure 2 about here.]

For each covariate representation, Figure 3 summarises the complexity of covariate representation for each of ξ, ν and ρ, using densities of both knot (or centroid) locations r, and the number n of knots. For BARS and Voronoi representations, knot (centroid) locations are approximately evenly

spread on the covariate domain for ξ, with some suggestion that more knots are located around $230°$ corresponding to the end of the land-shadow of Norway; this direction also sees a higher density of knots for ρ, and also to some extent for ν. BARS locates more knots around $0°$ compared to Voronoi. For P-splines, the 20 knots are evenly spaced on the covariate domain. For the Voronoi case, approximately 7 centroids on average are used to describe ρ and ν, but the spread of the distribution for ν is somewhat narrower; the mode of the posterior density is at unity for ξ. The fact that a greater number of centroids is selected to describe the variation of ρ and ν with covariate, compared with the number of centroids used to describe the variation of ξ with covariate, is consistent with expectation that in general the sample is more informative for ρ and ν than for ξ.

Figure 4 shows the posterior density of roughness coefficients λ for each of ξ, ν and ρ in each covariate representation. The P-splines model employs a different prior for spline coefficients β compared with BARS and Voronoi: as described in Section 3.2, a difference penalty is used for P-splines, whereas BARS and Voronoi use a (zero-order) ridge-type penalty. For ν and ρ, roughness coefficients for BARS and Voronoi are very similar. The Voronoi representation requires less penalisation for ξ, associated with a lower number of centroids in Figure 2, compared to BARS.

The prior density for n was set to a Poisson distribution with expectation 5 (see Supplementary Material) for both BARS and Voronoi representations. Since the estimated posterior distribution for n also lies around 5, to confirm that prior specification is not too influential, the analysis was repeated for Poisson priors on n with expectation 3 and 15, and a uniform prior on $[1, 30]$. Posterior estimates for these cases were very similar to those shown in Figure 3.

[Figure 3 about here.]

[Figure 4 about here.]

Using estimated Poisson and generalised Pareto models, the distribution of M_T, the T-year maximum storm peak H_S event in any covariate bin centred on the index set $\mathcal{I}_\theta$ follows a generalised extreme value distribution (Ross et al. 2017)

$$F_{M_T}(x|\xi_s, \nu_s, \rho_s) = \exp\left(-T\rho_s(1 - F(x|\xi_s, \nu_s))\right) \text{ for location } s \in \mathcal{I}_\theta$$

where $F(y|\xi_s, \nu_s)$ is the generalised Pareto cumulative distribution function. Since occurrences of storm peak H_S in different covariate bins are independent, the corresponding distributions of the T-year maximum storm peak H_S event over all covariate bins is simply $F_{M_T}(x|\boldsymbol{\xi}, \boldsymbol{\nu}, \boldsymbol{\rho}) =$

$\prod_{s=1}^{m} F_{M_T}(x|\xi_s, \nu_s, \rho_s)$. The posterior predictive cumulative distribution function for the T-year maximum storm peak H_S is

$$F_{M_T}(x) = \int_{\boldsymbol{\xi},\boldsymbol{\nu},\boldsymbol{\rho}} F_{M_T}(x|\boldsymbol{\xi},\boldsymbol{\nu},\boldsymbol{\rho})f(\boldsymbol{\xi},\boldsymbol{\nu},\boldsymbol{\rho})d\boldsymbol{\xi}d\boldsymbol{\nu}d\boldsymbol{\rho} \tag{5}$$

where $f(\boldsymbol{\xi},\boldsymbol{\nu},\boldsymbol{\rho})$ is the estimated joint posterior density of $\boldsymbol{\xi}$, $\boldsymbol{\nu}$ and $\boldsymbol{\rho}$. We evaluate this integral using Monte Carlo integration (Ross et al. 2017).

Figure 5 illustrates estimates for the posterior predictive distribution of M_{1000}, in terms of 2.5%, 25%, 50%, 75% and 97.5% percentiles for each covariate representation. Corresponding estimates (not shown) were generated for other return periods. There is good agreement between estimates for M_{1000}, both per directional sector and omni-directionally, and also for different return periods, suggesting consistency between all of P-spline, BARS and Voronoi representations as would be hoped. In the NE, E, SE and S sectors, where storms are less severe, the estimate for M_{1000} and its uncertainty is reduced, particularly for the Voronoi representation. Overall, P-splines provide somewhat more uncertain estimates than BARS and Voronoi: because of its reliance on equally-spaced knots and global roughness penalty, this representation is not sufficiently flexible to accommodate local directional characteristics.

[Figure 5 about here.]

It is illustrative to compare the empirical tail of threshold exceedances against posterior predictive tail both for different directional sectors, and for all directions. Figure 6 suggests that an extreme value model with any of the P-spline, BARS or Voronoi covariate parameterisations of ξ, ν and ρ provides a reasonable description of the sample. Omni-directionally, there is greater similarity between tail functions from P-splines and BARS. Voronoi gives more certain estimates for NE, E, SE and S sectors in particular. In summary, based on visual comparison of quality of fit, there is little to choose between the competing covariate representations.

[Figure 6 about here.]

5. SIMULATION STUDY

A simulation study provides more structured comparison of P-spline, BARS and Voronoi partition covariate representations. We generate $n_S = 100$ samples, each containing exactly $n_O = 1000$

observations of threshold exceedances with a generalised Pareto distribution. The true Poisson rate ρ of occurrence of threshold exceedance, and known shape ξ and scale ν of size of threshold exceedances, are assumed to vary systematically with covariate ("direction") θ. Functional forms of $\xi(\theta)$, $\nu(\theta)$ and $\rho(\theta)$ are generated using the sum of 10 weighted (wrapped) Gaussian kernels of standard deviation 30°, randomly located on the periodic covariate domain; weights are drawn at random from suitable distributions, so that variation of ξ, ν and ρ with θ approximately reflects that of the sample considered in Section 4. In particular, rate of occurrence of events on $[180°, 270°)$ was constrained to be low compared to other regions of the covariate domain. Corresponding distributions for T-year maxima are easily calculated. We choose a specific value of T corresponding to a period of $10\times$ the period of the sample, so that if we assume the sample of n_O events represents T_O years, we are estimating the distribution of the $10T_O$-year maximum, M_{10T_O}. For illustration, Figure 7 gives the "true known" directional variation of each of ξ, ν and ρ with θ for a typical realisation, along with the corresponding distribution of M_{10T_O} for 8 octants, and omni-directionally.

[Figure 7 about here.]

Figure 8 illustrates model fit for each of $\xi(\theta)$, $\nu(\theta)$ and $\rho(\theta)$ for the sample realisation illustrated in Figure 7. The figure suggests excellent estimation of ρ, and reasonable estimation of ν and ξ. As might be expected, credible intervals for ξ in particular are relatively large. However, variation of ρ, ν and ξ with covariate is identified by each of P-spline, BARS and Voronoi representations. Visual inspection of corresponding figures (not shown) for other sample realisations suggests consistent fitting performance. We infer that all covariate representations give reasonable models for the n simulated cases considered.

[Figure 8 about here.]

In a typical application, primary interest lies in estimation of extreme values associated with long return periods. It is natural to assess the three covariate representations in this respect. For each sample realised, we estimate the distribution of M_{10T_O} omni-directionally, and also for each of 8 directional octants centred on cardinal and semi-cardinal directions, using P-spline, BARS and Voronoi partition representations. We assess the performance of a covariate representation by comparing the distribution of M_{10T_O} from the estimated extreme value model with the truth. This comparison is summarised in terms of two quantities: an offset (the median of the estimated posterior

predictive distribution minus the true median), and a Kullback-Leibler (KL) divergence statistic (used e.g. for similar comparisons in Jones et al. 2016).

Distributions of estimated median offset for P-spline, BARS and Voronoi representations over the n_S samples are illustrated in Figure 9, for each directional octant and omni-directionally, together with the corresponding distributions of KL divergence statistics. The distribution of median offset is centred close to zero for all directional sectors. Omni-directionally, the estimated median value of median offset is approximately -0.3, 0.3 and 0.4 for P-spline, BARS and Voronoi representations. Variability of median offset is also somewhat lower for BARS. KL divergence for BARS is generally somewhat lower than for P-splines, which is lower than for Voronoi. Omni-directionally for KL divergence, there is little to choose between the covariate representations.

Figure 9 also shows median offsets and KL divergence for estimates based on independent, stationary Poisson and generalised Pareto fits to each directional octant (in light green), alongside corresponding omni-directional median offset and KL divergence. Further, the figure shows omni-directional median offset and KL divergence for a stationary (covariate-free) model (in dark green). Figure 9 suggests that median offsets for M_{10T_O} from independent stationary octant fits are generally reasonable when the sample size is sufficient; for SW, W and W octants, for which ρ is small (Figure 7), variability in median offset increases. Resulting omni-directional median offset also shows relatively large bias of approximately -2.1 (independent) and 1.0 (covariate-free). KL divergence for M_{10T_O} from independent stationary octant fits is generally larger than for P-splines, BARS and Voronoi; but there are some octants (e.g. N) for which Voronoi intervals are wider. We attribute this to the piecewise constant nature of the Voronoi representation. Omni-directionally, independent stationary octant (light green) and stationary (covariate-free, dark green) models perform more poorly than P-spline, BARS and Voronoi covariate representations.

[Figure 9 about here.]

Figure 10 shows densities for the number n of BARS knots or Voronoi centroids, estimated by averaging the corresponding posterior densities from models for each of the n_S sample realisations, for each of ξ, ν and ρ. Perhaps not surprisingly, general features of the figure are somewhat similar to those of Figure 3 (bottom row). The posterior probability mass of n for ρ using BARS is similar to the prior (a Poisson distribution with expectation 5). The mode of the corresponding posterior probability mass using Voronoi is approximately 9, reflecting the greater number of Voronoi cells

required to capture variability in ρ with θ (c.f. Figure 8). The posterior masses of n for ν and ρ using Voronoi are somewhat wider than those for BARS knots. For ξ however, the mode of the posterior mass of n is displaced from the prior to near unity, corresponding e.g. to a scalar (covariate-free) constant estimate for ξ for Voronoi.

[Figure 10 about here.]

6. DISCUSSION AND CONCLUSIONS

We have considered different covariate representations for the parameters of extreme value models, and demonstrated that, for one-dimensional covariates, each of P-spline, Bayesian adaptive regression spline and Voronoi partition representations provides reasonable estimates of extreme value parameter variation with a covariate, and of extreme quantiles corresponding to specific intervals of the covariate domain. Successful application of the P-spline representation typically relies on over-specifying the number of components (i.e. knots) in the covariate representation whilst penalising the roughness of parameter variation with covariate to optimise model fit. The BARS and Voronoi partition representations allow the number of basis components (i.e. BARS knots and Voronoi cells) itself to be estimated using reversible-jump MCMC inference, and the locations of knots and centroids to be changed. This flexibility increases the complexity of inference, but eliminates the need for difference penalisation; further it potentially reduces the complexity of the resulting model, and may increase model interpretability. P-spline and BARS-based parameter estimates are by construction smoother with covariate than those achieved by Voronoi partition, but in terms of estimation of extreme quantiles, there is little difference in performance between approaches.

There is no requirement that the same covariate representation be used for estimation of all parameters. In the current work, we might have chosen different covariate representations for the extreme value threshold ψ, exceedance rate ρ and generalised Pareto shape ξ and (modified) scale ν, but chose not to for simplicity.

We assumed a common P-spline representation for parameters of the gamma model used to estimate extreme value threshold ψ, so that the sample of threshold exceedances for subsequent rate and size modelling is fixed. The fixed sample is advantageous for purposes of comparing subsequent inferences. However, estimation of (non-stationary) extreme value threshold is itself challenging. It is possible

that a particular covariate representation, coupled with an appropriate model, might prove more useful than others in this estimation. In the current application, for example, the rapid change in the characteristics of H_S at around $230°$ suggests that a higher density of knots (or centroids) is required near that direction. As noted in Jones et al. (2016), in practical application, careful estimation of non-stationary ψ is at least as important for reliable inference as estimation of ρ, ν and ξ; see also Scarrott and MacDonald (2012) and Northrop et al. (2017). Any non-stationarity of extreme value threshold should be examined and identified either before or alongside non-stationarity of Poisson and generalised Pareto parameters. Indeed, intuition suggests, for modelling of magnitudes of events, that non-stationarity should be introduced sequentially and only as required, first in (a) extreme value threshold, before (b) generalised Pareto scale and finally (c) generalised Pareto shape. In the current work, we assume the extreme value threshold has already been identified without error.

All covariate representations have been implemented in a computationally-efficient manner in MATLAB software for non-stationary extreme value analysis. Full Bayesian inference (i.e. estimation of ψ, ρ, ν and ξ, and of T-year maxima for $T \in [10^2, 10^8]$ years) for a one-dimensional covariate (e.g. for the hindcast data from Section 4 or for any of the 100 simulated samples in Section 5) can be completed in approximately five minutes on a workstation with reasonable performance specification.

For a typical implementation of VGAM for marginal extreme value analysis (e.g. Yee and Stephenson 2007), we might adopt functional forms for generalised Pareto parameters ξ and ν of the form $\eta_j = \sum_{k=1}^{q} f_{jk}(\theta_{(k)})$, $j = 1, 2$, for q covariates $\{\theta_{(k)}\}_{k=1}^{q}$ and $2q$ smooth functions $\{f_{jk}(\theta_{(k)})\}_{j=1,k=1}^{2,q}$, where each f_{jk} is a function of a single covariate $\theta_{(k)}$. The variation of f_{jk} with $\theta_{(k)}$ might be described using a P-spline, with global smoothness constraint(s) on the covariate domain. In the current study with a one-dimensional covariate, this VGAM description is closely related to the P-spline approach. However, a typical VGAM implementation would not allow optimal estimation of the number and placement of knots (or centroids), achieved using BARS and Voronoi representations. Moreover, when applied to two- and higher-dimensional covariates, a typical VGAM would only allow additive descriptions of the form $\sum_{k=1}^{q} f_{jk}(\theta_{(k)})$ for generalised Pareto parameter η_j, in contrast to the more general description $f_j(\theta_{(1)}, \theta_{(2)}, ..., \theta_{(q)})$ in the current approach.

Further, Wood (2003) presents an approach to estimation of low-rank smoothers, motivated by approximate solutions to thin plate regression spline models. Since estimation of thin plate spline models in multiple dimensions is computationally demanding, this approach is beneficial when the approximate solution can be achieved considerably more efficiently. Jonathan and Ewans (2011)

report the application of thin-plate splines to non-stationary extreme value analysis. Thin plate spline approaches are naturally multidimensional. Typically, for a q-dimensional covariate, (a) Euclidean distance is assumed as the natural metric for distance between two points in covariate space, and (b) function penalisation is applied globally on the covariate domain. Characteristic (a) is common to the Voronoi representation in the current work. Characteristic (b) is assumed by all of P-spline, BARS and Voronoi representations, but is less critical for BARS and Voronoi since spline knot and Voronoi cell numbers and locations are also estimated. Wood (2004) proposes generalised additive models (GAMs) with multiple smoothing parameters and improved numerical stability. Krivobokova et al. (2008) propose a fast approach to adaptive penalised splines, in which the spline smoothing parameter itself is also modelled as (another) penalised spline, and the Laplace approximation used for marginal likelihood estimation. Wood (2011) considers fast estimation of semi-parametric generalised linear models, amounting to optimisation of Laplace approximate restricted maximum likelihood or maximum likelihood criteria. This work is extended by Wood et al. (2016) to encompass non-exponential family models. The covariate representations here also prove useful in general regression settings beyond extreme value analysis of peaks over threshold; estimation strategy described is largely independent of likelihood form. Indeed, in other studies, we have found the representations useful for non-stationary exponential and non-exponential family models.

This study of covariate representations focusses on application to marginal extreme value modelling, but we believe that some of its findings might prove more generally useful. When extending the current work to multidimensional covariates (Zanini et al. 2020), it might prove useful to adopt (e.g.) a two-dimensional tensor product spline basis, consisting of a P-spline representation for seasonal variation and a BARS representation for direction. Further, for strongly coupled covariates (e.g. direction and season for storm peak H_S in the South China Sea, Randell et al. 2015) a two-dimensional Voronoi partition will probably prove an adequate parsimonious description. For yet higher-dimensional covariates, it is likely that a BARS or (in particular) a Voronoi representation will prove computationally considerably more efficient to estimate, notwithstanding slick computational techniques such as GLAMs (Currie et al. 2006) available for covariate descriptions in terms of penalised tensor products of B-splines.

ACKNOWLEDGEMENT

ONLINE SUPPLEMENTARY MATERIAL

There is online Supplementary Material accompanying this article.

REFERENCES

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

Biller, C., 2000. Adaptive Bayesian regression splines in semiparametric generalized linear models. J. Comput. Graph. Statist. 9, 122–140.

Bodin, T., Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm. Geophysical Journal International 178, 1411–1436.

Bodin, T., Sambridge, M., Gallagher, K., 2009. A self-parametrizing partition model approach to tomographic inverse problems. Inverse Problems 25, 055009.

Boldi, M.O., Davison, A.C., 2007. A mixture model for multivariate extremes. J. Roy. Statist. Soc. B 69, 217–229.

Bottolo, L., Consonni, G., Dellaportas, P., Lijoi, A., 2003. Bayesian analysis of extreme values by mixture modeling. Extremes 6, 27–36.

Brezger, A., Lang, S., 2006. Generalized structured additive regression based on Bayesian P-splines. Comput. Statist. Data Anal. 50, 967–991.

Carter, D.J.T., Challenor, P.G., 1981. Estimating return values of environmental parameters. Quart. J. R. Met. Soc. 107, 259.

Chavez-Demoulin, V., Davison, A., 2005. Generalized additive modelling of sample extremes. J. Roy. Statist. Soc. Series C: Applied Statistics 54, 207–222.

Coles, S., Walshaw, D., 1994. Directional modelling of extreme wind speeds. Applied Statistics 43, 139–157.

Costain, D.A., 2008. Bayesian partitioning for modeling and mapping spatial case-control data. Biometrics 65, 1123–1132.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. J. Roy. Statist. Soc. B 49, 1–39.

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

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

Denison, D.G.T., Adams, N.M., Holmes, C.C., Hand, D.J., 2002. Bayesian partition modelling. Comput. Statist. Data Anal. 38, 475–485.

Denison, D.G.T., Mallick, B.K., Smith, A.F.M., 1998. Automatic Bayesian curve-fitting. J. Roy. Statist. Soc. B 60, 330–50.

DiMatteo, I., Genovese, C.R., Kass, R.E., 2001. Bayesian curve-fitting with free-knot splines. Biometrika 88, 1055–1071.

Eilers, P.H.C., Currie, I.D., Durban, M., 2006. Fast and compact smoothing on multi-dimensional grids. Comput. Statist. Data Anal. 50, 61–76.

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

El Adlouni, S., Ouarda, T.B.M.J., 2009. Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo. Water Resour. Res. 45.

Ewans, K.C., Jonathan, P., 2008. The effect of directionality on northern North Sea extreme wave design criteria. J. Offshore. Arct. Eng. 130, 041604:1–041604:8.

Fawcett, L., Walshaw, D., 2007. Improved estimation for temporally clustered extremes. Environmetrics 18, 173–188.

Feld, G., Randell, D., Ross, E., Jonathan, P., 2019. Design conditions for waves and water levels using extreme value analysis with covariates. Ocean Eng. 173, 851–866.

Gamerman, D., Lopes, H.F., 2006. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman and Hall / CRC, Boca Raton, USA.

Girolami, M., Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. J. Roy. Statist. Soc. B 73, 123–214.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Hobert, J.P., Casella, G., 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. J. Am. Statist. Soc. 91, 1461–1473.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339–349.

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

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 Eng. 35, 1211–1225.

Jones, M., Randell, D., Ewans, K., Jonathan, P., 2016. Statistics of extreme ocean environments: non-stationary inference for directionality and other covariate effects. Ocean Eng. 119, 30–46.

Krivobokova, T., Crainiceanu, C.M., Kauermann, G., 2008. Fast adaptive penalized splines. J. Comput. Graph. Statist. 17, 1–20.

Lang, S., Brezger, A., 2004. Bayesian p-splines. J. Comput. Graph. Statist. 13, 183–212.

Marx, B.D., Eilers, P.H.C., 1998. Direct generalised additive modelling with penalised likelihood. Comput. Statist. Data Anal. 28, 193–209.

Mendez, F.J., Menendez, M., Luceno, A., Medina, R., Graham, N.E., 2008. Seasonality and duration in extreme value distributions of significant wave height. Ocean Eng. 35, 131–138.

Northrop, P., Attalides, N., Jonathan, P., 2017. Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. J. Roy. Statist. Soc. C 66, 93–120.

Northrop, P., Jonathan, P., 2011. Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights. Environmetrics 22, 799–809.

Ouarda, T.B.M.J., El Adlouni, S., 2011. Bayesian nonstationary frequency analysis of hydrological variables. J. Am. Water Resour. As. 47, 496–505.

Raghupathi, L., Randell, D., Ewans, K., Jonathan, P., 2016. Fast computation of large scale marginal extremes with multi-dimensional covariates. Comput. Statist. Data Anal. 95, 243–258.

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.

Randell, D., Turnbull, K., Ewans, K., Jonathan, P., 2016. Bayesian inference for non-stationary marginal extremes. Environmetrics 27, 439–450.

Reistad, M., Breivik, O., Haakenstad, H., Aarnes, O.J., Furevik, B.R., Bidlot, J.R., 2011. A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea. J. Geophys. Res. 116, 1–18.

Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). J. Roy. Statist. Soc. B 59, 731–792.

Robinson, M.E., Tawn, J.A., 1997. Statistics for extreme sea currents. Appl. Statist. 46, 183–205.

Ross, E., Randell, D., Ewans, K., Feld, G., Jonathan, P., 2017. Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. Ocean Eng. 142, 315–328.

Ross, E., Sam, S., Randell, D., Feld, G., Jonathan, P., 2018. Estimating surge in extreme North Sea storms. Ocean Eng. 154, 430–444.

Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Revstat 10, 33–60.

Scotto, M., Guedes-Soares, C., 2000. Modelling the long-term time series of significant wave height with non-linear threshold models. Coastal Eng. 40, 313–327.

Stephenson, A., 2016. Bayesian inference for extreme value modelling, in: Dey, D., Yan, J. (Eds.), Extreme Value Modeling and Risk Analysis: Methods and Applications. Chapman and Hall / CRC, pp. 257–280.

Wallstrom, G., Liebner, J., Kass, R.E., 2008. An implementation of Bayesian adaptive regression splines (BARS) in C with S and R wrappers. Journal of Statistical Software 26.

Wood, S.N., 2003. Thin plate regression splines. J. Roy. Statist. Soc. B 65, 95–114.

Wood, S.N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Am. Statist. Soc. 99, 673–686.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. Roy. Statist. Soc. B 73, 3–36.

Wood, S.N., Pya, N., Safken, B., 2016. Smoothing parameter and model selection for general smooth models. J. Am. Statist. Soc. 111, 1548–1563.

Xifara, T., Sherlock, C., Livingstone, S., Byrne, S., Girolami, M., 2014. Langevin diffusions and the Metropolis-adjusted Langevin algorithm. Stat. Probabil. Lett. 91, 14–19.

Yee, T.W., Stephenson, A., 2007. Vector generalized linear and additive extreme value models. Extremes 10, 1–19.

Yee, T.W., Wild, C.J., 1996. Vector generalized additive models. J. R. Stat. Soc. B 58, 481–493.

Zanini, E., Eastoe, E., Jones, M., Randell, D., Jonathan, P., 2020. Multidimensional covariate representations for non-stationary extremes. (Draft at www.lancs.ac.uk/~jonathan) .
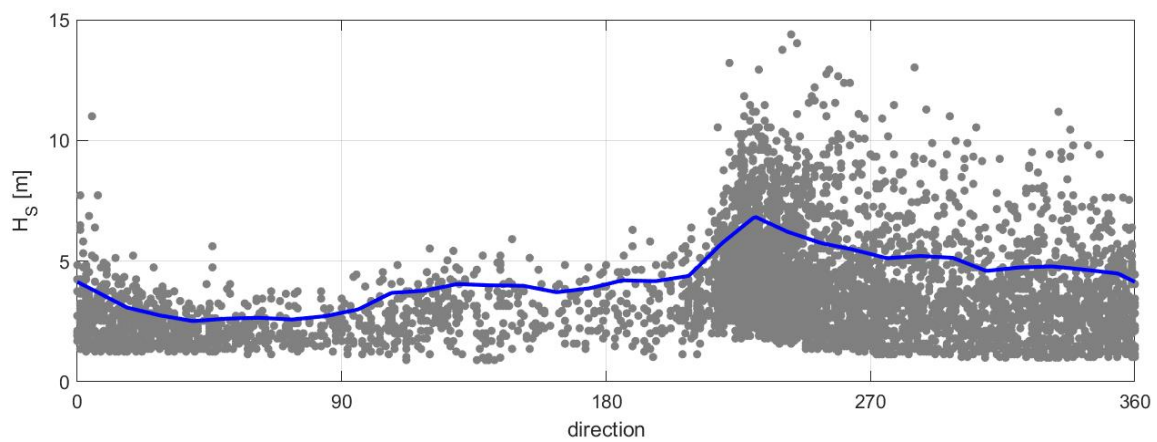
FIGURES

Figure 1. North Sea application. Storm peak significant wave height H_S (in metres) on direction (dots), together with directional extreme value threshold ψ (blue line, corresponding to $\tau = 0.8$) estimated using a P-spline covariate representation, with 30 knots evenly-spaced on the covariate domain.
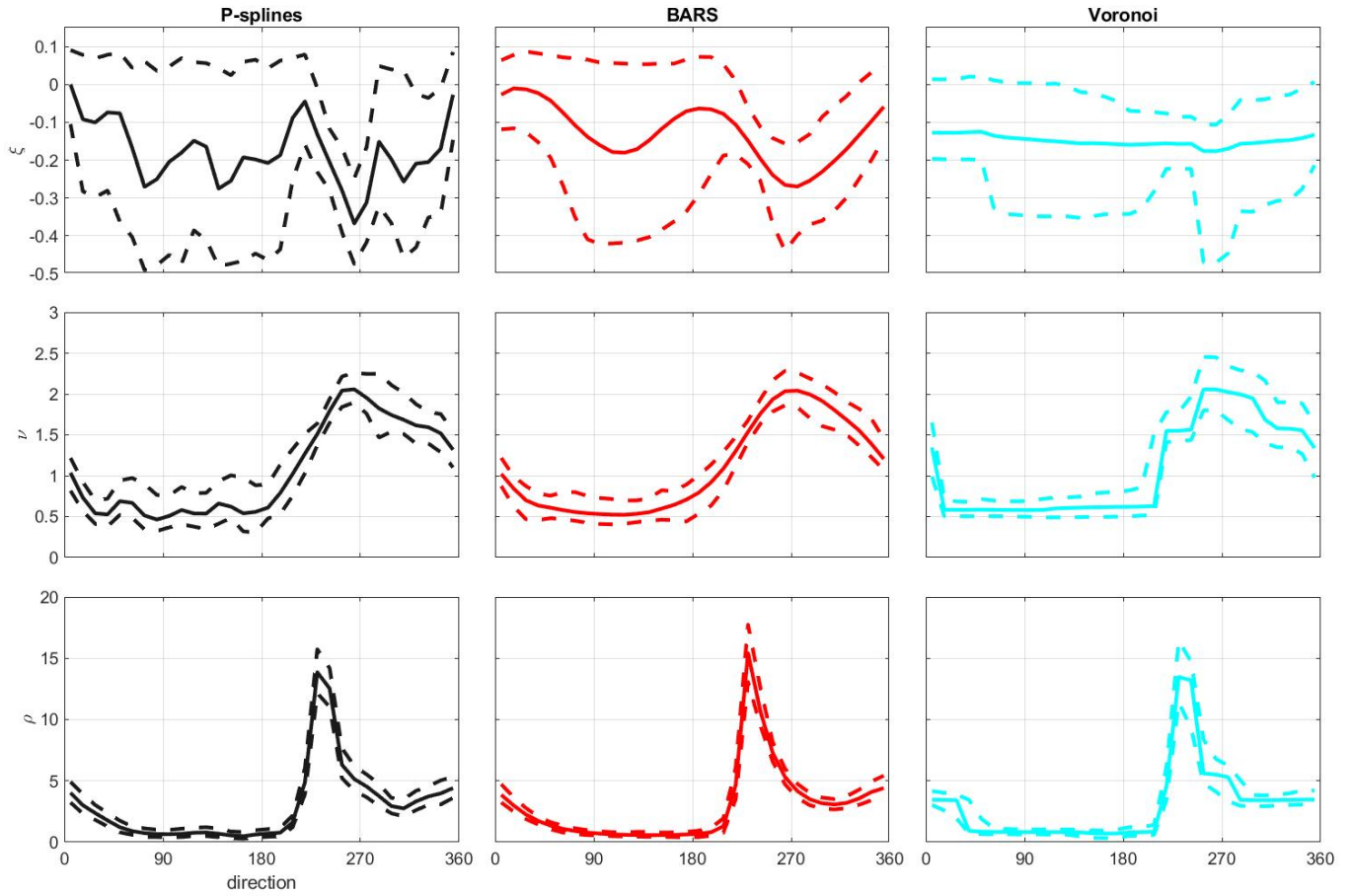
Figure 2. North Sea application. Posterior distributions of generalised Pareto shape ξ (top), scale ν (centre) and Poisson rate ρ (bottom), for each of P-splines (left), BARS (centre) and Voronoi (right) representations. Each panel shows posterior mean (full line) and 95% credible interval (dotted line) as a function of direction.
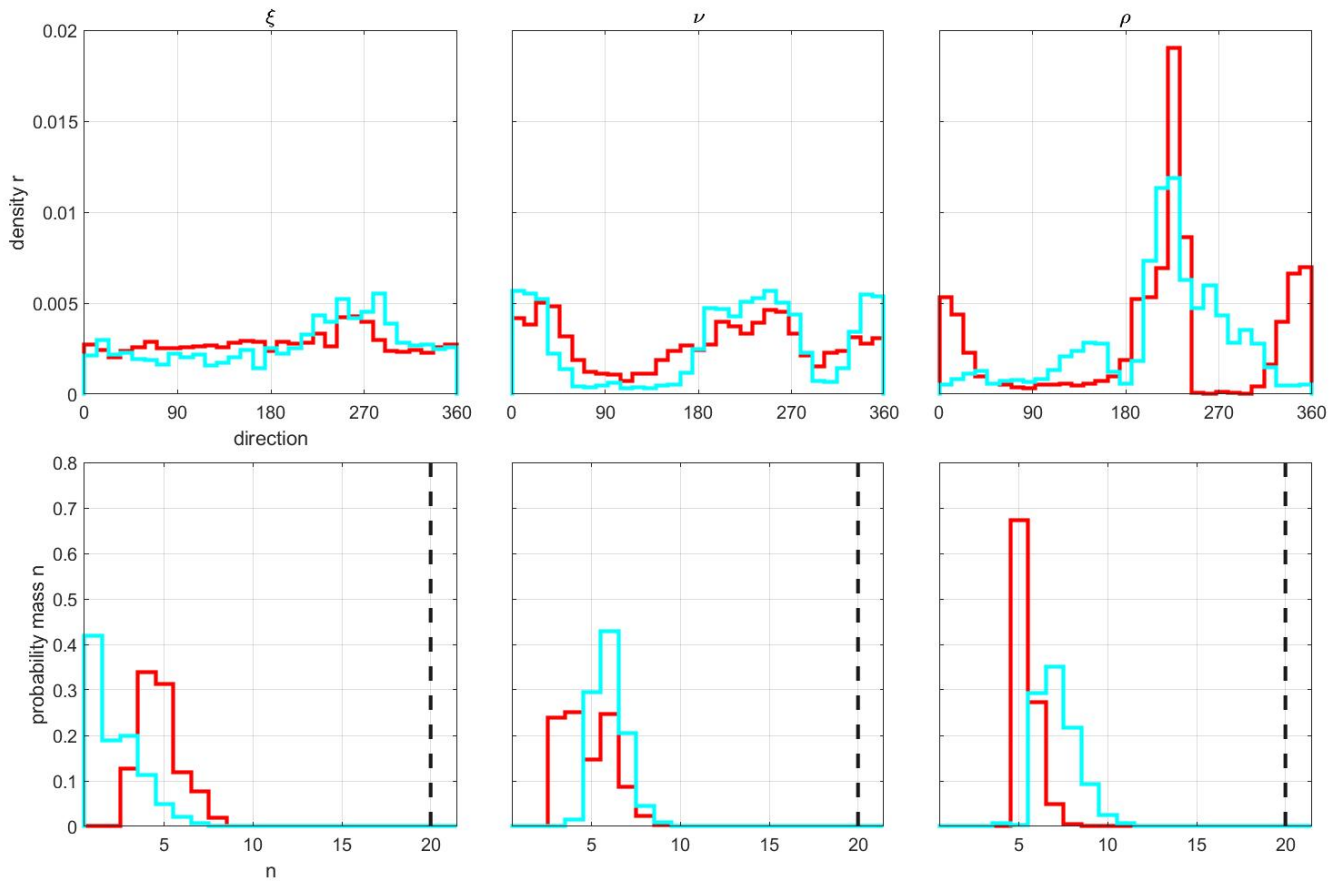
Figure 3. North Sea application. Estimated posterior density of knot and centroid locations (top), and estimated posterior density of number of knots (bottom) for each of ξ (left), ν (centre) and ρ (right). Each panel shows estimates for BARS (red) and Voronoi (cyan) representations. Vertical black lines show the number of P-spline knots used. The P-spline knot locations are evenly spaced on the covariate domain.
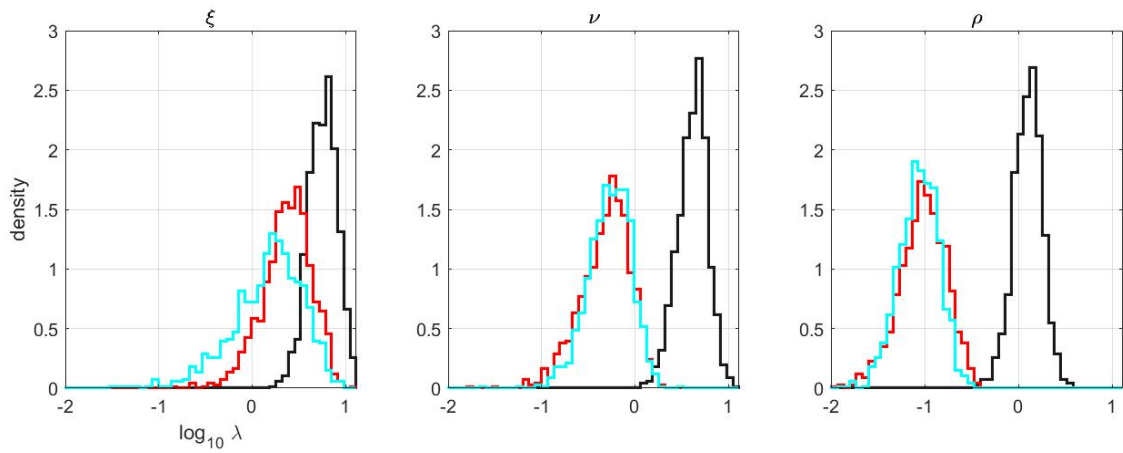
Figure 4. North Sea application. Estimated posterior density for roughness coefficient λ for each of ξ (left), ν (centre) and ρ (right). Each panel shows estimates for P-splines (black), BARS (red) and Voronoi (cyan). Note that the P-splines model exploits a difference penalty, whereas BARS and Voronoi use a (zero-order) ridge-type penalty.
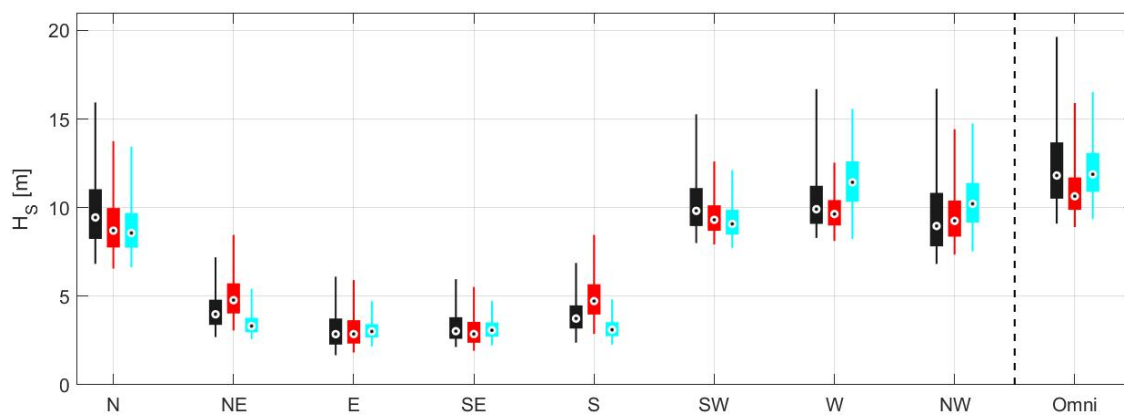
Figure 5. North Sea application. Posterior predictive distribution for M_{1000}, the 1000-year maximum storm peak H_S, summarised using box-whisker plots in terms of 2.5%, 25%, 50%, 75% and 97.5% percentiles, for (triplets of) P-splines (black, left), BARS (red, centre) and Voronoi (cyan, right) corresponding to 8 directional sectors centred on the cardinal and semi-cardinal directions, and over all directions ("Omni").
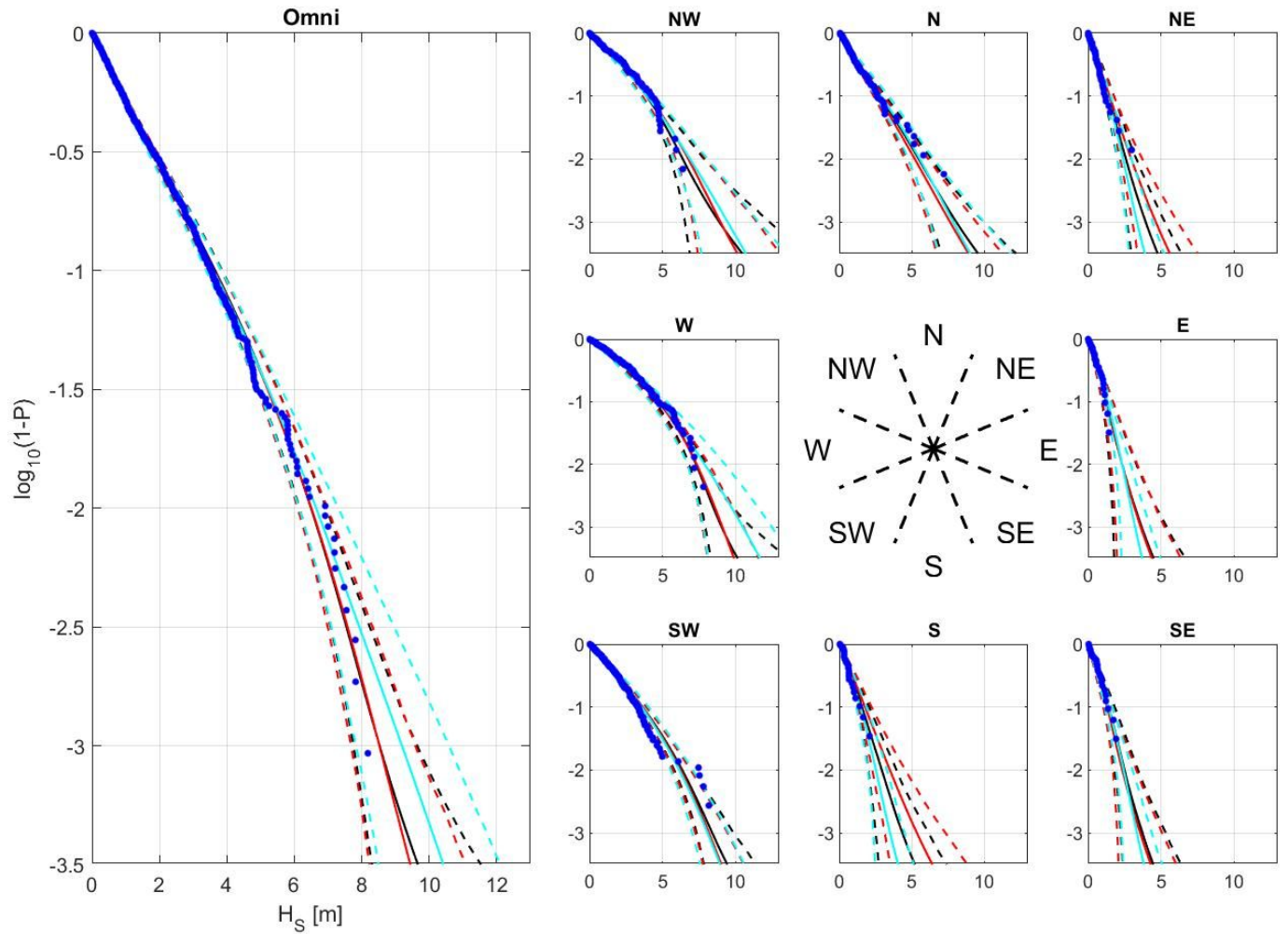
Figure 6. North Sea application. Illustration of model validation for the P-spline (black), BARS (red) and Voronoi (cyan) representations, shown in terms of estimates for the distribution of storm peak H_S corresponding to the period of the original sample, plotted as $\log_{10}(1 - P)$ to accentuate tail behaviour, for cumulative distribution function P. For comparison, the dotted blue curve is an empirical estimate for the same tail distribution obtained by sorting the sample of threshold exceedances. The black, red and cyan curves summarise the predictive distribution of the quantile estimate (for given tail probability $1 - P$) under the directional model, as the median (solid) and 2.5% and 97.5% values (dashed), estimated using numerical integration. The left hand panel corresponds to the omni-directional case, and the right hand panels to estimates for each of 8 directional octants centred on cardinal and semi-cardinal directions.
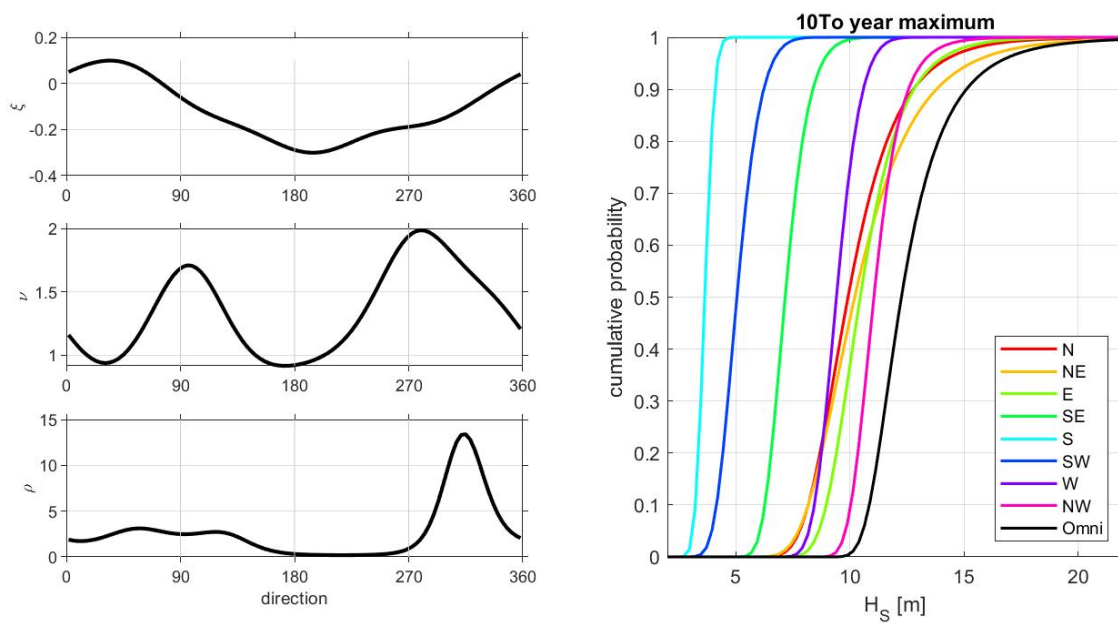
Figure 7. Simulation study. Left hand side: variation of true ξ (top), ν (centre) and ρ (bottom) with θ for one of the n_S realisations. Right hand side: the corresponding cumulative distribution of the $10T_O$-year maximum, for each of 8 directional octants centred on the cardinal and semi-cardinal directions, and over all directions ("Omni", in black).
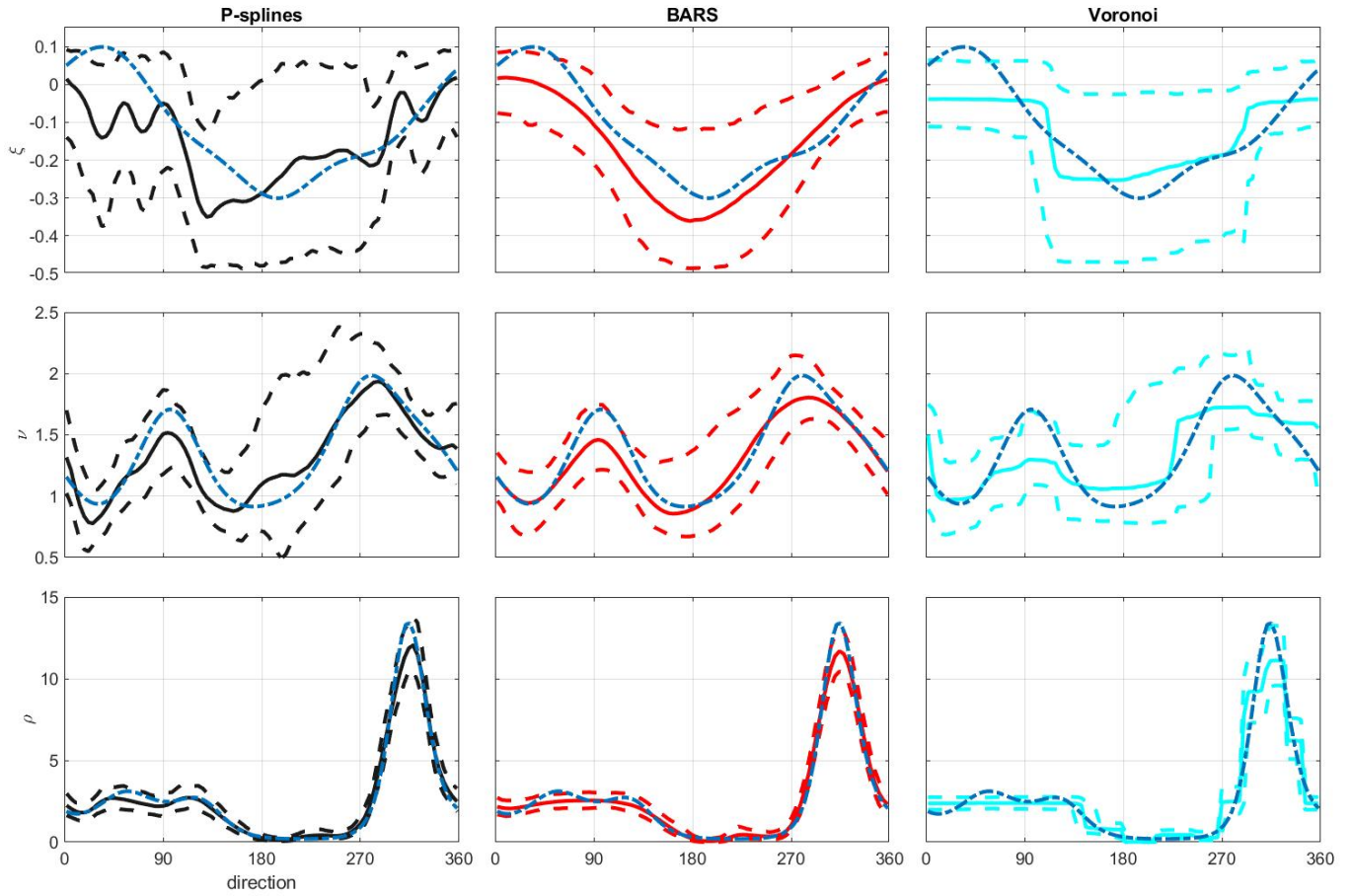
Figure 8. Simulation study. Posterior distributions of generalised Pareto shape ξ (top), scale ν (centre) and Poisson rate ρ (bottom), for each of P-splines (left), BARS (centre) and Voronoi (right) representations, for sample realisation discussed in Figure 7. Each panel shows posterior mean (full line) and 95% credible interval (dotted line) as a function of direction. The true parameter value as a function of direction (also given in Figure 7) is shown as a broken (dot-dashed) line.
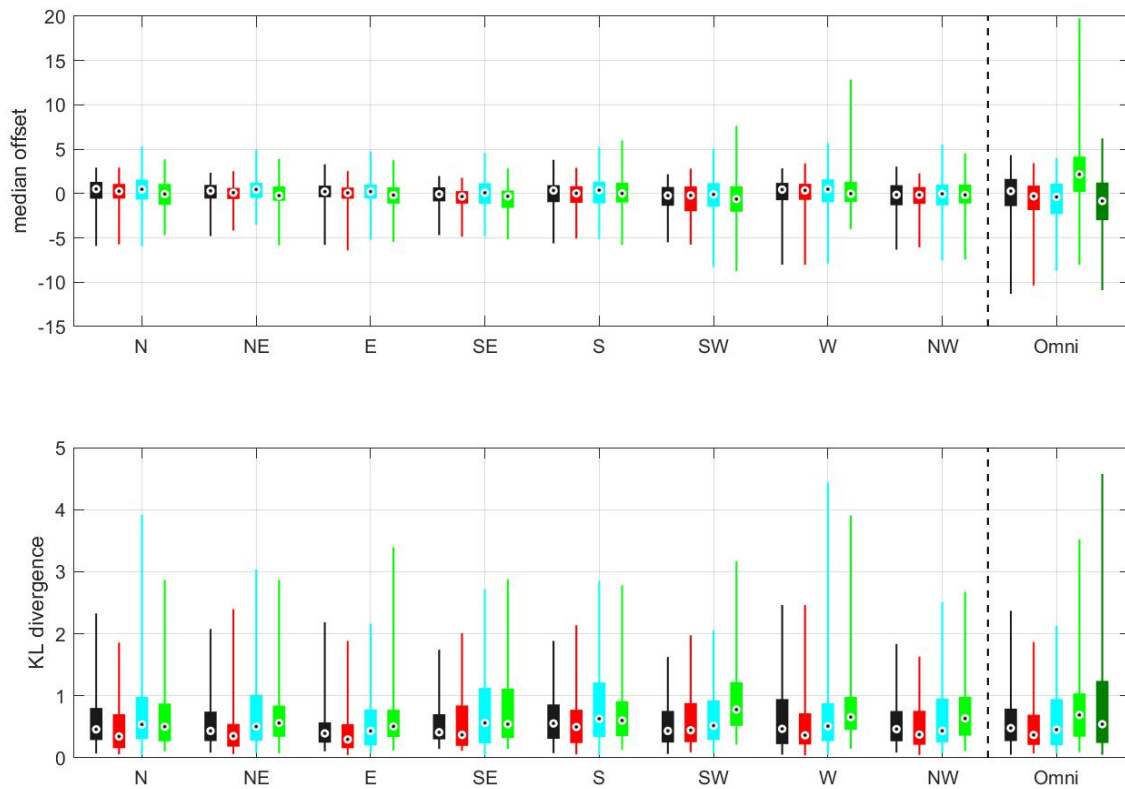
Figure 9. Simulation study. Distribution for estimated median offset (top) and Kullback-Leibler (KL) divergence (bottom) for 8 directional octants centred on the cardinal and semi-cardinal directions, and over all directions ("Omni"). Box-whisker plots are constructed using 2.5%, 25%, 50%, 75% and 97.5% percentiles of the corresponding distribution, for return period $10T_O$, for (quadruplets of) P-splines (black, left), BARS (red, centre left), Voronoi (cyan, centre right) and independent stationary fits per octant (light green, right). Omni-directional estimates of median offset and KL divergence from a stationary (covariate-free) model are also shown (dark green, far right).
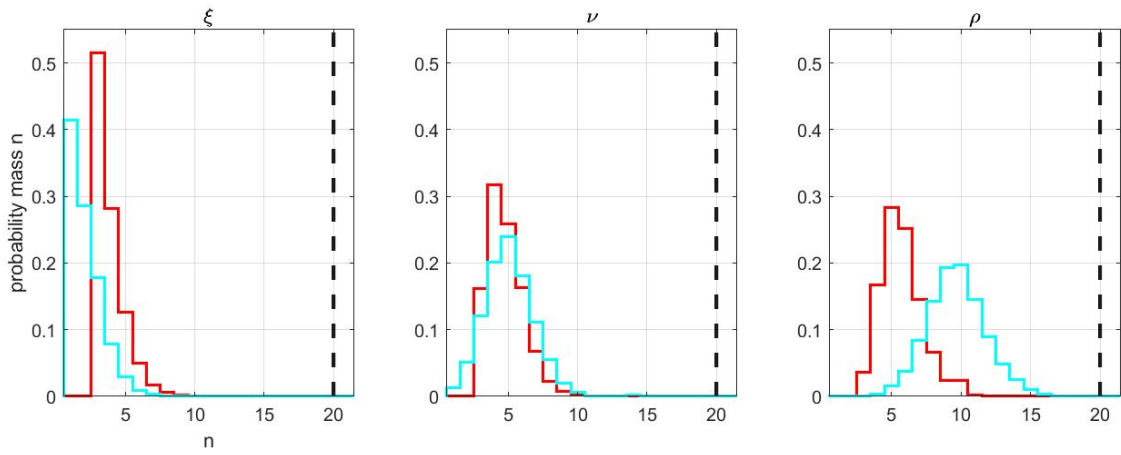
Figure 10. Simulation study. Estimated posterior density of number of knots for each of ξ (left), ν (centre) and ρ (right) from the simulation study. Each panel shows estimates for BARS (red) and Voronoi (cyan) representations. Vertical black line shows the number of P-spline knots.