Flexible covariate representations for extremes

Slides and draft paper at *www.lancs.ac.uk/∼jonathan*

Elena Zanini, Emma Eastoe, Matthew Jones, David Randell, Philip Jonathan
Shell & Lancaster University

Thanks

- Colleagues at Shell: Vadim Anokhin, Graham Feld, Emma Ross
- Colleagues at Lancaster: Jonathan Tawn

Structural damage

Ike, Gulf of Mexico, 2008 (Joe Richard)

North Sea, Winter 2015-16 (The Inertia)

Motivation

- Rational and consistent design and assessment of marine structures
  - Reduce bias and uncertainty in estimation of structural integrity
  - Quantify uncertainty as well as possible
- Non-stationary marginal, conditional, spatial and temporal extremes
  - Multiple locations, multiple variables, time-series
  - Multidimensional covariates
- Improved understanding and communication of risk
  - Incorporation within established engineering design practices
  - Knock-on effects of improved inference

The ocean environment is an amazing thing to study ... especially if you like to combine beautiful physics, measurement and statistical modelling!

Motivation

- Environmental extremes vary smoothly with multidimensional covariates
  - Model parameters are non-stationary
- Environmental extremes exhibit spatial and temporal dependence
  - Characterise these appropriately
- Uncertainty quantification for whole inference
  - Data acquisition (simulator or measurement)
  - Data pre-processing (storm peak identification)
  - Hyper-parameters (extreme value threshold)
  - Model form (marginal measurement scale effect, spatial extremal dependence)
- Statistical and computational efficiency
  - Slick algorithms
  - Parallel computation
  - Bayesian inference

This work

Directional models for storm peak H_S

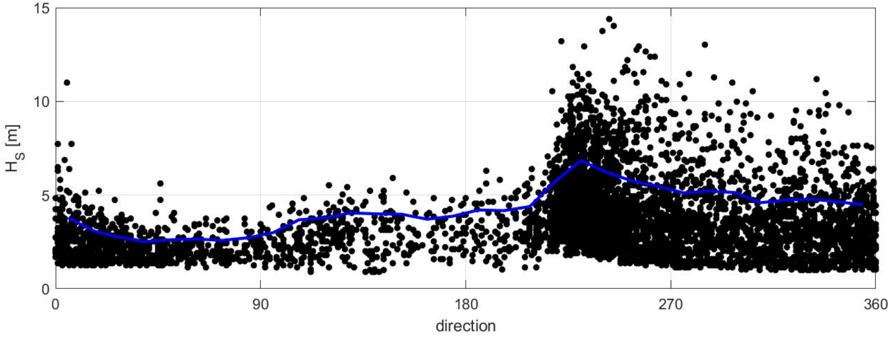
- Different covariate representations
  - Penalised B-splines (or P-splines)
  - Bayesian adaptive regression splines
  - Voronoi partition
- Generic modelling framework
- Bayesian inference

- Northern North Sea case study as motivation
- Simulation study for comparison
- Focus on the generalised Pareto (GP) inference

- Extensions to multidimensional covariates

Motivating application

Typical data for northern North Sea. Storm peak H_S on direction, with $\tau = 0.8$ extreme value threshold.



Model

Observational model

- Sample of peaks Y over threshold ψ, with covariates θ
  - θ is 1D in current work : directional
  - θ is nD later : e.g. $4D$ spatio-directional-seasonal
- Extreme value threshold ψ assumed known
  - Estimated as the $\tau = 0.8$ quantile of a directional gamma model to full data
  - Essential in general to capture uncertainty in ψ
- Y assumed to follow generalised Pareto distribution with shape ξ, (modified) scale ν (=$\sigma(1 + \xi)$)
  - ξ, ν are functions of θ

- Frigessi et al. [2002], Behrens et al. [2004], MacDonald et al. [2011], Randell et al. [2016], Northrop et al. [2017]

Generalised Pareto

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

- $\nu = \sigma(1+\xi)$
- $y > \psi, \psi \in (-\infty, \infty)$
- Shape parameter $\xi \in (-\infty, \infty)$ and scale parameter $\nu \in (0, \infty)$

Covariate representations

- Index set $\mathcal{I}_\theta = \{\theta_s\}_{s=1}^m$ on **periodic** covariate domain $\mathcal{D}_\theta$
- Each observation belongs to exactly one θ_s
- On $\mathcal{I}_\theta$, assume

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

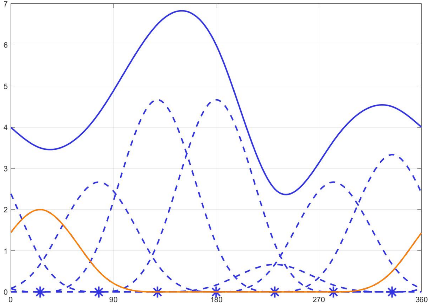
- $\eta \in (\xi, \nu)$
- $\boldsymbol{B} = \{B_{sk}\}_{s=1;k=1}^{m;n}$ basis for $\mathcal{D}_\theta$
- $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^{n}$ basis coefficients
- Inference reduces to estimating n_ξ, n_ν, $\boldsymbol{B}_\xi$, $\boldsymbol{B}_\nu$, $\boldsymbol{\beta}_\xi$ and $\boldsymbol{\beta}_\nu$

- P-splines, BARS and Voronoi are different forms of $\boldsymbol{B}$

P-splines

- n **regularly**-spaced knots on $\mathcal{D}_\theta$
- B consists of n B-spline bases
  - Order d
  - Each using $d+1$ consecutive knot locations
  - Local support
  - Wrapped on $\mathcal{D}_\theta$
  - Cox - de Boor recursion formula

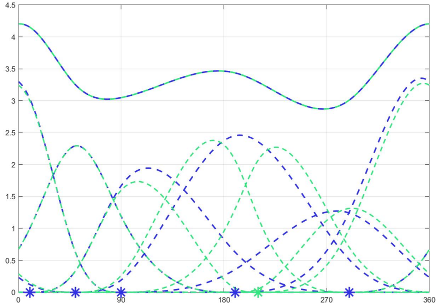
- n is fixed and "over-specified"
- Knot locations $\{r_k\}_{k=1}^n$ fixed
- Local roughness of $\boldsymbol{\beta}$ penalised



Periodic P-splines

BARS basis

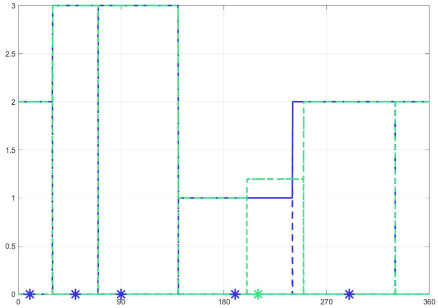
- n **irregularly**-spaced knots on $\mathcal{D}_\theta$
- $\boldsymbol{B}$ consists of n B-spline bases
- Knot locations $\{r_k\}_{k=1}^n$ can change
- Number of knots n can change



Periodic BARS knot birth and death

Voronoi partition

- n **irregularly**-spaced centroids on $\mathcal{D}_\theta$
  - Define n neighbourhoods or "cells"
- B consists of n basis functions
  - Piecewise constant on $\mathcal{D}_\theta$
  - $= 1$ "within cell", $= 0$ "outside"
- Centroid locations $\{r_k\}_{k=1}^{n}$ can change
- Number of centroids n can change



Periodic Voronoi centroid birth and death

Prior for β (all representations)

$$\text{prior density of } \boldsymbol{\beta} \quad \propto \quad \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}\boldsymbol{\beta}\right)$$

- $\boldsymbol{P} = \lambda \boldsymbol{D}'\boldsymbol{D}$, $\boldsymbol{D}$ is a $n \times n$ (wrapped) differencing matrix
- P-splines: $\boldsymbol{D}$ represents first-difference; prior equivalent to local roughness penalty
- BARS and Voronoi: $\boldsymbol{D}$ is $\boldsymbol{I}_n$; prior is "ridge-type" for Bayesian regression

Prior for λ (all representations)

$$\lambda \quad \sim \quad \text{gamma}$$

Prior for n (BARS and Voronoi)

$$n \quad \sim \quad \text{Poisson}$$

Prior for $r_k, k = 1, 2, ..., n$ (BARS and Voronoi)

$$r_k \quad \sim \quad \text{uniform}$$

Parameter set Ω

- P-splines: $\Omega = \{\boldsymbol{\beta}_\xi, \lambda_\xi, \boldsymbol{\beta}_\nu, \lambda_\nu\}$ with $n_\xi, \boldsymbol{r}_\xi, n_\nu$ and $\boldsymbol{r}_\nu$ pre-specified
- BARS and Voronoi: $\Omega = \{n_\xi, \boldsymbol{r}_\xi, \boldsymbol{\beta}_\xi, \lambda_\xi, n_\nu, \boldsymbol{r}_\nu, \boldsymbol{\beta}_\nu, \lambda_\nu\}$
- where $\boldsymbol{r} = \{r_k\}_{k=1}^n$, $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^n$,

Updating β, λ (all representations) and r (BARS and Voronoi)

- Gibbs sampling when full conditionals available
- Otherwise Metropolis-Hastings (MH) within Gibbs, using suitable proposal mechanisms, mMALA where possible
- Roberts and Stramer [2002], Girolami and Calderhead [2011], Xifara et al. [2014]

Conditional structure

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

- where $\eta \in (\xi, \nu)$

Dimension-jumping (BARS and Voronoi)

- Update n, and birth or death elements of $\boldsymbol{r}, \boldsymbol{\beta}$ using reversible-jump MCMC
- Green [1995], Richardson and Green [1997], Biller [2000], Zhou and Shen [2001], DiMatteo et al. [2001], Wallstrom et al. [2008], Costain [2008], Bodin and Sambridge [2009]

Birth-death Metropolis-Hastings acceptance probability

- Jump from current $\Omega = (n_\eta, \boldsymbol{r}_\eta, \lambda_\eta, \boldsymbol{\beta}_\eta)$ to proposed Ω^* (=$(\Omega \setminus \omega, \omega^*)$)
- $\omega = (n_\eta, \boldsymbol{\beta}_\eta, \boldsymbol{r}_\eta)$ in current and $\omega^* = (n^*_\eta, \boldsymbol{\beta}^*_\eta, \boldsymbol{r}^*_\eta)$ in proposed

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

- $f(\boldsymbol{y}|\Omega)/f(\boldsymbol{y}|\Omega^*)$ sample lik. ratio
- $f(\omega)/f(\omega^*)$ prior ratio
- $q(\omega^*|\omega)/q(\omega|\omega^*)$ proposal ratio
- Final term Jacobian for transformation

- Sample from prior!

Dimension-jumping birth for β

- Location r^+ of the new knot is sampled uniformly on $\mathcal{D}_\theta$
- Current knot locations $\boldsymbol{r} = \{r_k\}_{k=1}^n$ and proposed $\boldsymbol{r}^* = (\{r_k\}_{k=1}^n, r^+)$
- Establish bijection between augmented coefficient vector $\boldsymbol{\beta}^a = (\boldsymbol{\beta}, u_\beta)$ $(u_\beta \sim N(0, \bullet))$ for current state, and vector $\boldsymbol{\beta}^*$ for proposed
- Motivation: make $\boldsymbol{B}\boldsymbol{\beta}$ and $\boldsymbol{B}^*\boldsymbol{\beta}^*$ as similar as possible
- Regression solution is $\hat{\boldsymbol{\beta}}^* = \left[(\boldsymbol{B}^{*\prime}\boldsymbol{B}^*)^{-1}\boldsymbol{B}^{*\prime}\boldsymbol{B}\right]\boldsymbol{\beta} = \boldsymbol{G}_j\boldsymbol{\beta}$
- Set

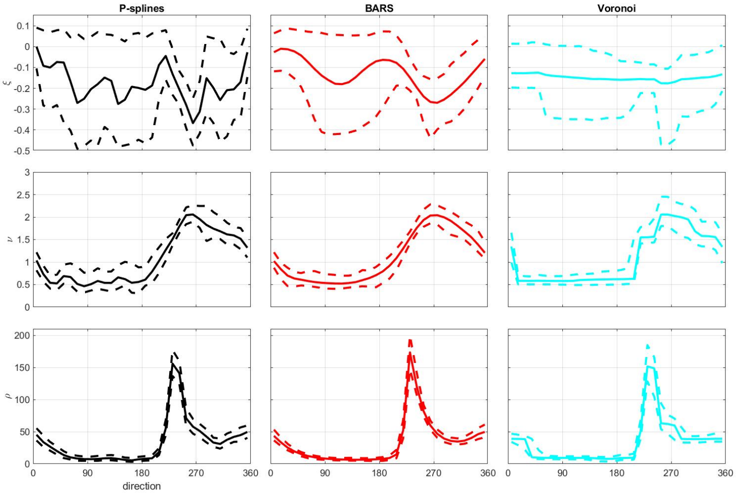
$$\boldsymbol{\beta}_j^* = \left[\begin{array}{c|c} & 0 \\ \boldsymbol{G}_j & \vdots \\ & 0 \\ & 1 \end{array}\right] \times \begin{bmatrix}\boldsymbol{\beta}_j \\ u_\beta\end{bmatrix} = \boldsymbol{F}_j\boldsymbol{\beta}_j^a.$$

- Jacobian for a birth is $|\boldsymbol{G}|$
- For death transition, essentially use $\boldsymbol{F}^{-1}$
- Zanini et al. [2019]

North Sea application

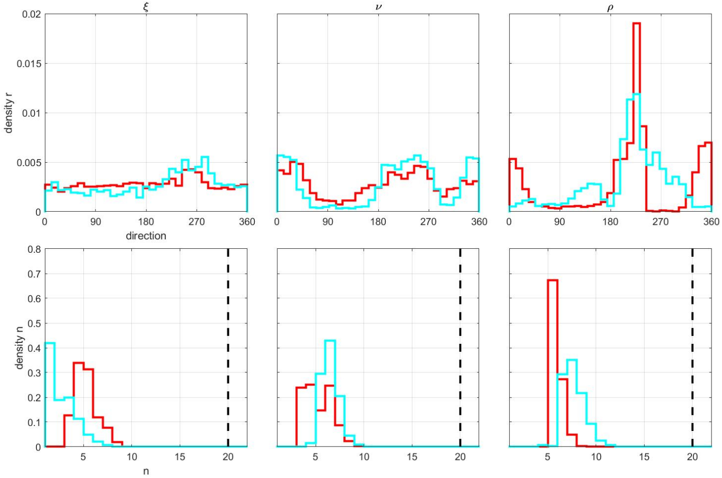
Posterior parameter estimates for ξ, ν and ρ for northern North Sea

- Note colour scheme
- Rate ρ and ν very similar
- Voronoi gives almost constant ξ
- Voronoi piecewise constant
- Land shadow effects
- General agreement

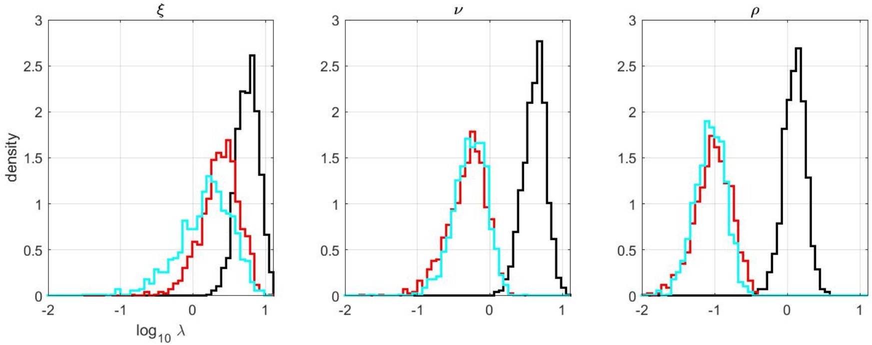


Posterior densities for locations r and numbers n

- Knot placement uniform for ξ, clear effect for ρ
- n close to 1 for Voronoi ξ
- General agreement
- Effect of different priors on n checked

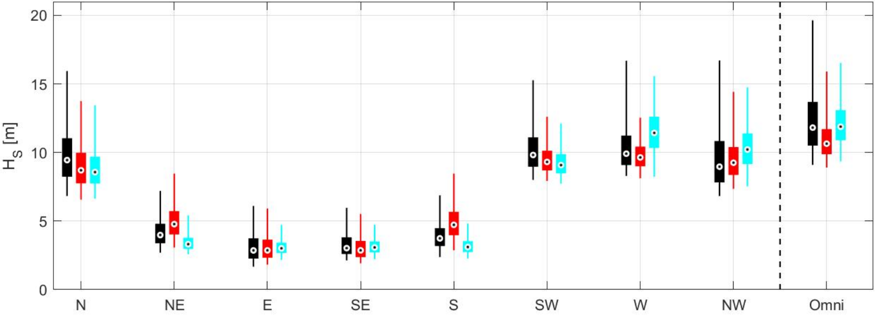


Posterior densities for penalty coefficients λ



- Ridge penalties for BARS and Voronoi, but roughness for P-splines
- λ somewhat lower for Voronoi, but also this has smaller n
- General consistency

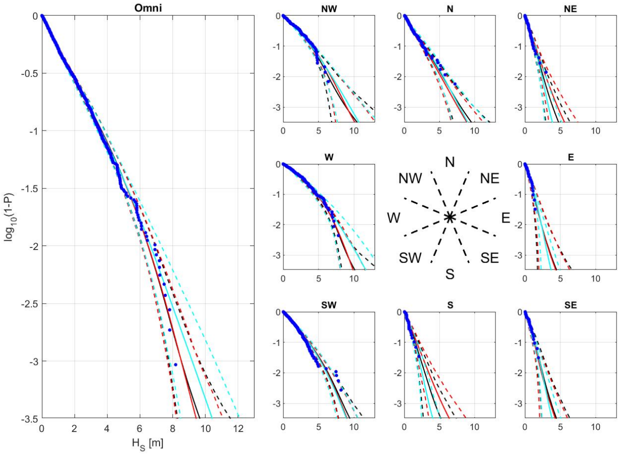
Directional posterior predictive distribution of $T = 1000$-year maximum



- Box-whiskers with 2.5%, 25%, 50%, 75% and 97.5% percentiles
- Uncertainties larger for P-splines?
- General consistency

Fit diagnostic

- Empirical tail (blue)
- Posterior means and 95% credible intervals for quantile levels from different models
- General consistency

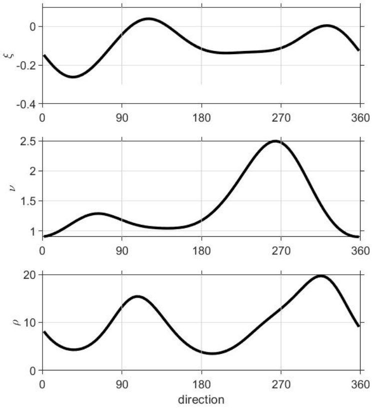


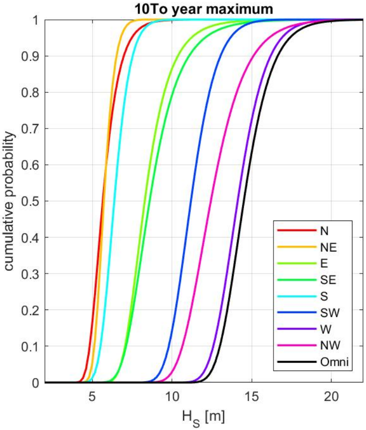
Simulation study

Set-up

- $n_S = 100$ samples, each containing exactly $n_O = 1000$ observations of threshold exceedances with a generalised Pareto distribution
- True Poisson rate ρ, shape ξ and scale ν vary systematically with covariate θ.
- Functional forms of $\xi(\theta)$, $\nu(\theta)$ and $\rho(\theta)$ generated using sum of 10 weighted (wrapped) Gaussian kernels of standard deviation 30°, randomly located on the periodic covariate domain
- Weights drawn at random from suitable distributions, so that $\xi(\theta)$, $\nu(\theta)$ and $\rho(\theta)$ like North Sea sample
- Distribution of T-year maxima ($T = 10\times$ the period of sample, T_O) estimated

Illustrative realisation

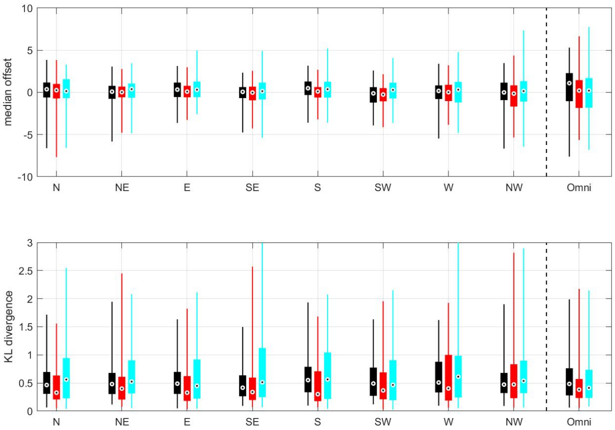




- True $\xi(\theta)$, $\nu(\theta)$ and $\rho(\theta)$ for typical realisation
- Directional distribution of $10T_O$-year maximum for 8 octants, and "omni"

Performance summary

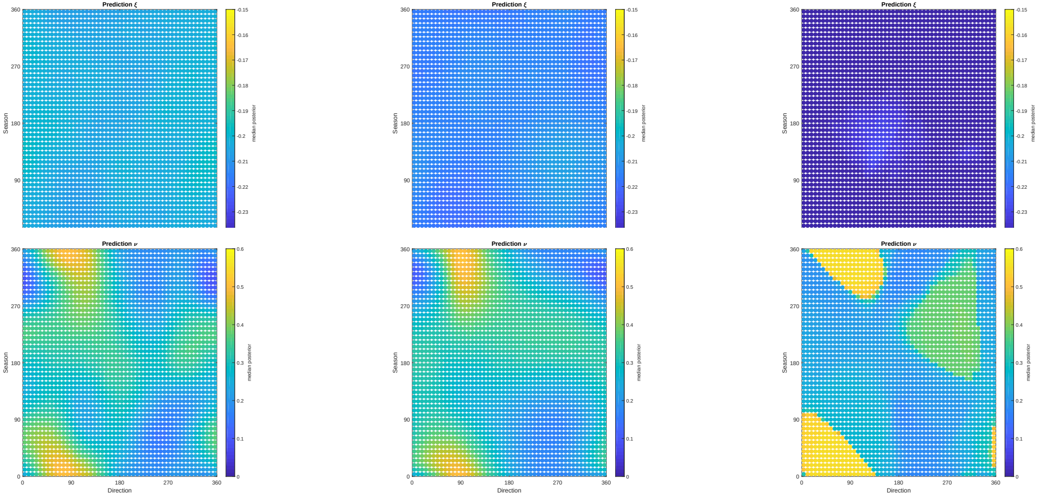
- Compare posterior predictive distribution for $10T_O$-year maximum with truth
- Median offset small
- KL divergence more variable for Voronoi
- BARS slightly better?
- General consistency



Where next?

2D covariates: a qualitative comparison for the South China Sea

P-splines: $n_\xi = 6 \times 6, n_\nu = 6 \times 6$ | BARS: $n_\xi^{mo} = 3 \times 3, n_\nu^{mo} = 4 \times 4$ | Voronoi: $n_\xi^{mo} = 1, n_\nu^{mo} = 7$



Summary

- Covariate effects important in environmental extremes
- Need to tackle big problems $\Rightarrow$ need efficient models
- Need to provide solutions as "end-user" software $\Rightarrow$ stable inference

- P-splines: straightforward, global roughness per dimension
- BARS: optimally-placed knots
- All splines: nD basis is tensor product of marginal bases
- Voronoi: piecewise constant, naturally nD

- Combinations useful
- Conditional, spatial and temporal extremes

References

C N Behrens, H F Lopes, and D Gamerman. Bayesian analysis of extreme events with threshold estimation. *Stat. Modelling*, 4:227–244, 2004.

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

Thomas Bodin and Malcolm Sambridge. Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, 178:1411–1436, 2009.

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

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

A. Frigessi, O. Haug, and H. Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5:219–235, 2002.

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

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

A. MacDonald, C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. A flexible extreme value mixture model. *Comput. Statist. Data Anal.*, 55:2137–2157, 2011.

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

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

S Richardson and P. J. Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *J. Roy. Statist. Soc. B*, 59(4):731–792, 1997.

G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and Computing in Applied Probability*, 4:337–358, 2002.

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

T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, and M Girolami. Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Stat. Probabil. Lett.*, 91(2002):14–19, 2014.

E. Zanini, E. Eastoe, M. Jones, D. Randell, and P. Jonathan. Covariate representations for non-stationary extremes. *Environmetrics (draft at www.lancs.ac.uk/∼jonathan)*, 2019.

S. Zhou and X. Shen. Spatially adaptive regression splines and accurate knot selection schemes. *J. Am. Statist. Soc.*, 96:247–259, 2001.