Efficient adaptive covariate modelling for extremes

Slides at *www.lancs.ac.uk/∼jonathan*

Matthew Jones, David Randell, Emma Ross, Elena Zanini, Philip Jonathan

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

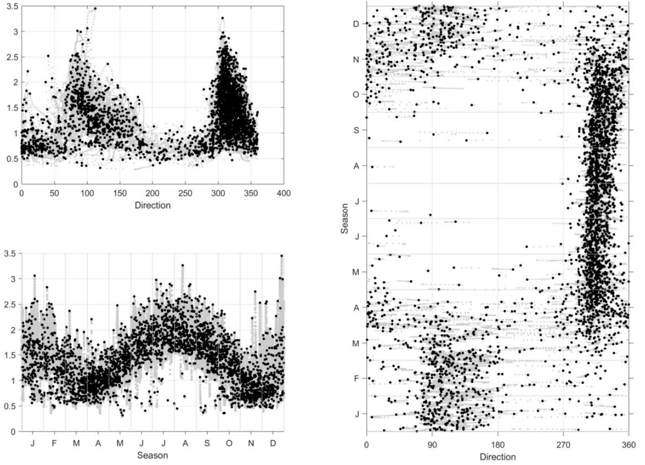
The ocean environment is an amazing thing to study ... especially if you like to combine beautiful physics, measurement and statistical modelling!

Fundamentals

- 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

A typical sample

Typical data for South China Sea location. Sea state (grey) and storm peak (black) H_S on season and direction



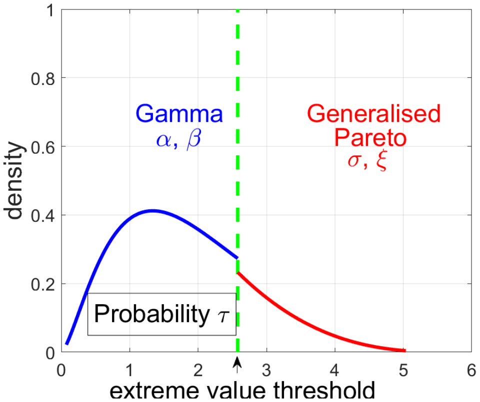
Outline

Directional-seasonal covariate models for H_S^{sp}

- Introductory example using P-splines
- Adaptive splines
- Partition models

- South China Sea example as "connecting theme"
- Focus on the generalised Pareto (GP) inference

Simple gamma-GP model



Simple gamma-GP model

- Sample of peaks over threshold y, with covariates θ
  - θ is 1D in motivating example : directional
  - θ is nD later : e.g. $4D$ spatio-directional-seasonal
- Below threshold ψ
  - y follows truncated gamma with shape α, scale $1/\beta$
  - Hessian for gamma better behaved than Weibull
- Above ψ
  - y follows generalised Pareto with shape ξ, scale σ
- ξ, σ, α, β, ψ all functions of θ
- ψ for pre-specified threshold probability τ
  - Generalise later to estimation of τ

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

Simple gamma-GP model

- Density is $f(y|\xi, \sigma, \alpha, \beta, \psi, \tau)$

$$= \begin{cases} \tau \times f_{TG}(y|\alpha, \beta, \psi) & \text{for } y \leq \psi \\ (1-\tau) \times f_{GP}(y|\xi, \sigma, \psi) & \text{for } y > \psi \end{cases}$$

- Likelihood is $\mathcal{L}(\xi, \sigma, \alpha, \beta, \psi, \tau | \{y_i\}_{i=1}^{n})$

$$= \prod_{i:y_i \leq \psi} f_{TG}(y_i|\alpha, \beta, \psi) \prod_{i:y_i > \psi} f_{GP}(y_i|\xi, \sigma, \psi)$$

$$\times \quad \tau^{n_B}(1-\tau)^{(1-n_B)} \text{ where } n_B = \sum_{i:y_i \leq \psi} 1.$$

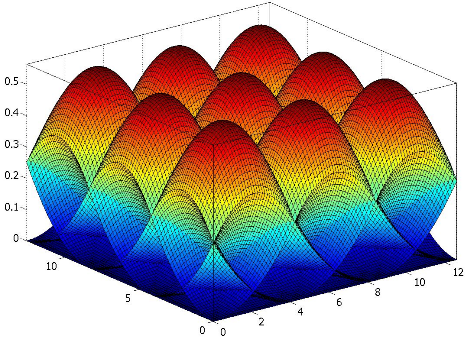
Estimate all parameters as functions of θ

Standard P-spline model

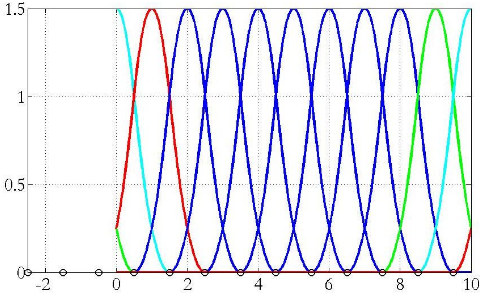
- Physical considerations suggest $\alpha, \beta, \rho, \xi, \sigma, \psi$ and τ vary smoothly with covariates θ
- Values of $\eta \in \{\alpha, \beta, \rho, \xi, \sigma, \psi, \tau\}$ on some index set of covariates take the form $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\beta}_\eta$
  - For nD covariates, $\boldsymbol{B}$ takes the form of tensor product $\boldsymbol{B}_{\theta_n} \otimes ... \otimes \boldsymbol{B}_{\theta_\kappa} \otimes ... \otimes \boldsymbol{B}_{\theta_2} \otimes \boldsymbol{B}_{\theta_1}$
- Spline roughness with respect to each covariate dimension κ given by quadratic form $\lambda_{\eta\kappa}\beta'_{\eta\kappa}\boldsymbol{P}_{\eta\kappa}\boldsymbol{\beta}_{\eta\kappa}$
- $\boldsymbol{P}_{\eta\kappa}$ is a function of stochastic roughness penalties $\delta_{\eta\kappa}$

- Brezger and Lang [2006]

P-splines



Kronecker product



Periodic P-splines

Priors and conditional structure

Priors

$$\begin{aligned}
\text{density of } \boldsymbol{\beta}_{\eta\kappa} &\propto \exp\left(-\frac{1}{2}\lambda_{\eta\kappa}\boldsymbol{\beta}'_{\eta\kappa}\boldsymbol{P}_{\eta\kappa}\boldsymbol{\beta}_{\eta\kappa}\right) \\
\lambda_{\eta\kappa} &\sim \text{gamma} \\
(\text{ and } \tau &\sim \text{beta, when } \tau \text{ estimated })
\end{aligned}$$

Conditional structure

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

$$\eta\in\Omega=\{\alpha,\beta,\rho,\xi,\sigma,\psi,\tau\}$$

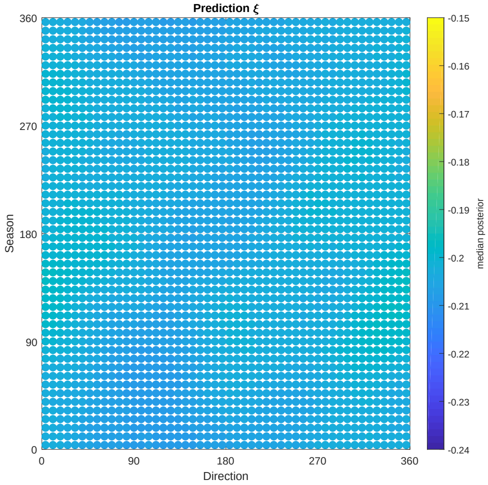
Inference

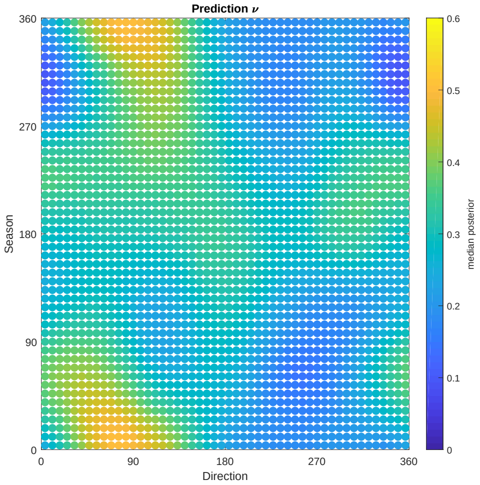
- Elements of $\boldsymbol{\beta}_\eta$ highly interdependent, correlated proposals essential for good mixing
- "Stochastic analogues" of IRLS and back-fitting algorithms for maximum likelihood optimisation used previously
- Estimation of different penalty coefficients for each covariate dimension

- 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]

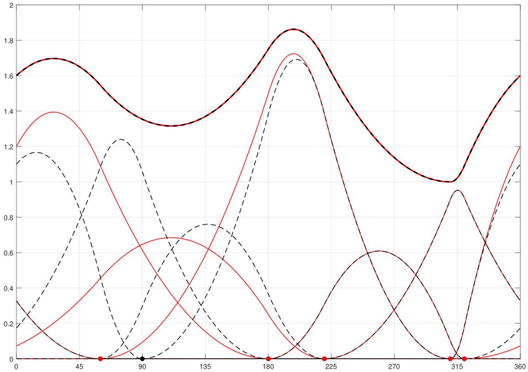
p-splines: GP parameter estimates





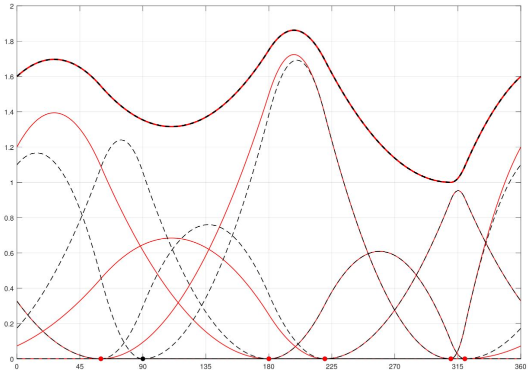
Inference with adaptive splines

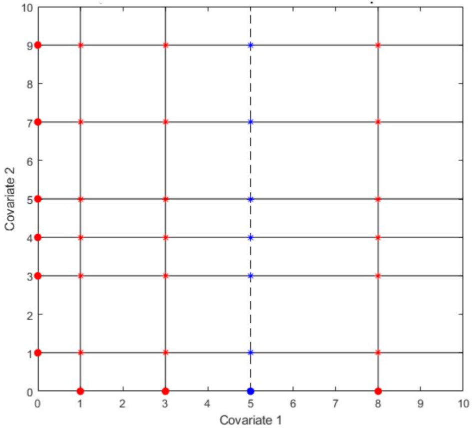
- Advantages
  - Arbitrary location of knots, and number of knots
- Estimate number, location, coefficient of knots
- Reversible-jump MCMC:
  - Birth-death
  - Split-combine (local birth-death)
  - Detailed balance



- Biller [2000], Zhou and Shen [2001], DiMatteo et al. [2001], Wallstrom et al. [2008]

Inference with adaptive splines : e.g. birth-death





Inference with adaptive bases: birth-death

Acceptance probability

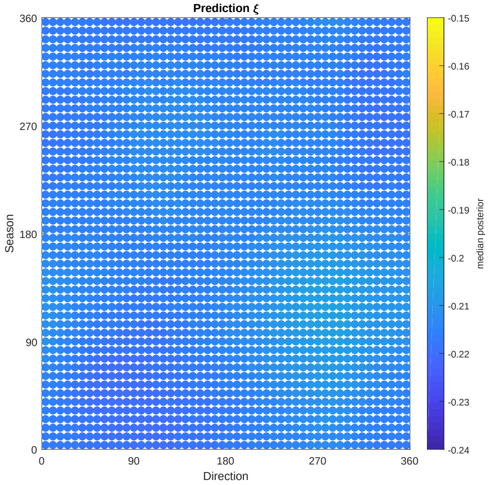
$$\alpha(m'|m) = \min\left\{1, \frac{f(m')}{f(m)} \times \frac{f(y|m')}{f(y|m)} \times \frac{q(m|m')}{q(m'|m)} \times \left|\frac{\partial m'}{\partial m}\right|\right\}$$

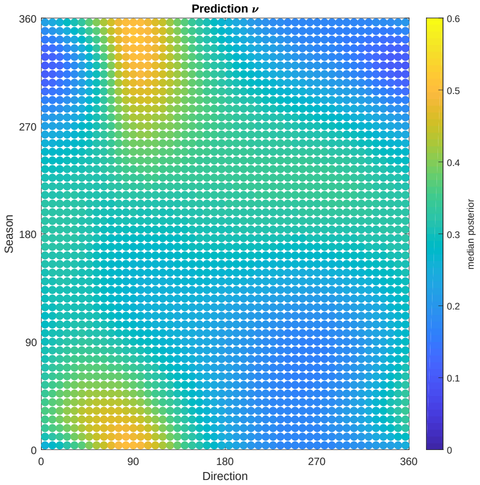
Dimension-jumping proposals: β_1 (p-vector) $\rightarrow$ β_2 (($p+1$)-vector)

$$\begin{aligned} \eta &= B_1\beta_1 = B_2\beta_2^* \\ \Rightarrow \hat{\beta}_2^* &= \left[(B_2'B_2)^{-1}B_2'B_1\right]\beta_1 = G\beta_1 \end{aligned}$$

$$\begin{aligned} \beta_2 &= \left[\begin{array}{c|c} & 0 \\ G & \vdots \\ & 0 \\ & 1 \end{array}\right] \times \begin{bmatrix} \beta_1 \\ u \end{bmatrix} \\ u &\sim N(0, \bullet) \end{aligned}$$

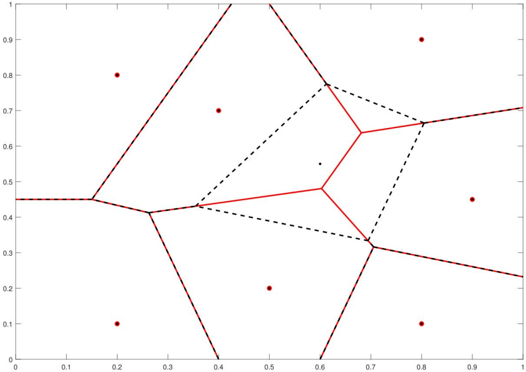
Adaptive splines: GP parameter estimates





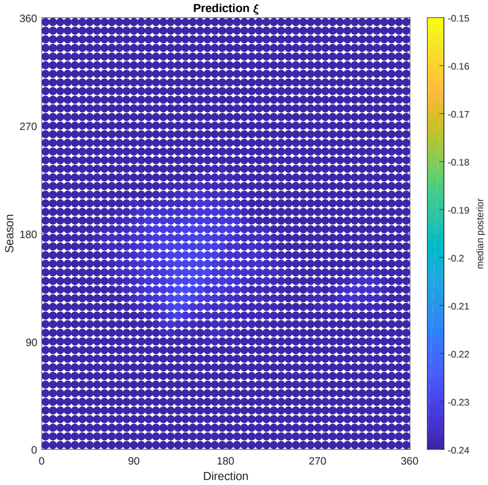
Partition model

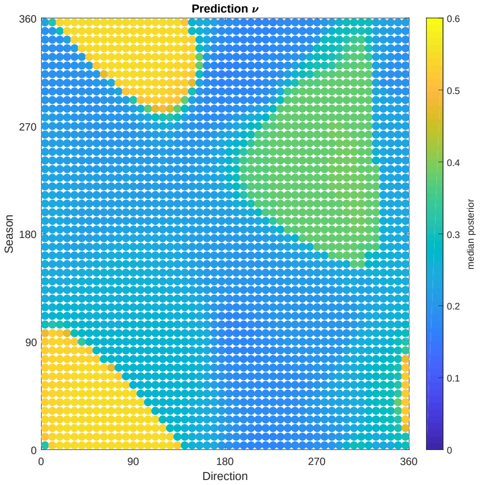
- Pros & cons
  - Naturally local, nD
  - Piecewise constant
- Estimate
  - Number of cells
  - Centroid locations
  - Cell coefficients
- Reversible-jump MCMC
  - Birth-death
  - Detailed balance



- Green [1995], Heikkinen and Arjas [1998], Denison et al. [2002], Costain [2008], Bodin and Sambridge [2009]

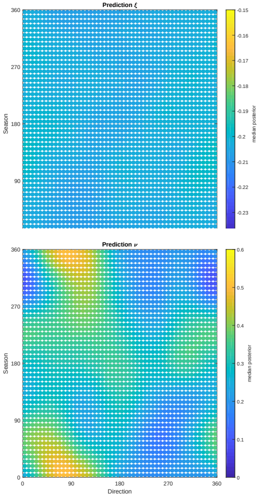
Partition model: GP parameter estimates



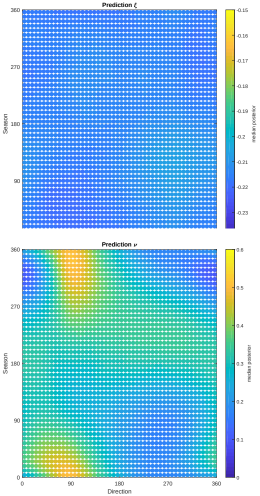


Qualitative comparison of different estimates

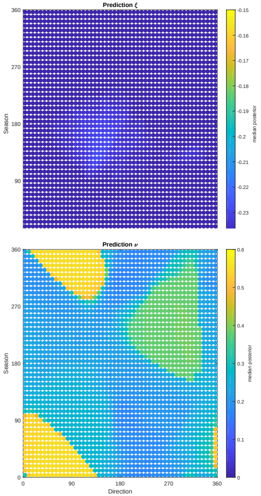
P-splines: $n_\xi = 6 \times 6$, $n_\nu = 6 \times 6$



Adaptive splines: $n_\xi^{mo} = 3 \times 3$, $n_\nu^{mo} = 4 \times 4$



Partition: $n_\xi^{mo} = 1$, $n_\nu^{mo} = 7$



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
- Adaptive splines: optimally-placed knots
- All splines: nD basis is tensor product of marginal bases
- Partition: piecewise constant, naturally nD
- Partition mixture model

- 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.

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

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

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

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.

Juha Heikkinen and Elja Arjas. Non-parametric bayesian estimation of a spatial poisson intensity. *Scand. J. Stat.*, 25:435–450, 1998.

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.

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

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.

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

Supporting material

Partition model: ψ

Partition model: ξ and ν traces