Bayesian smoothing for extremes

Slides and draft article at:
www.lancs.ac.uk/∼jonathan

David Randell, **Philip Jonathan**, **Kathryn Turnbull**, **Matthew Jones**
EVAN 2015 Santander

Acknowledgement

- Shell: Kevin Ewans, Graham Feld, Shejun Fan, Michael Vogel.
- Academic colleagues: Lancaster, Durham.
- Ex-tropical storm Henri, EasyJet, Melisa and EVAN 2015 organisers for making talk possible remotely!

Motivation

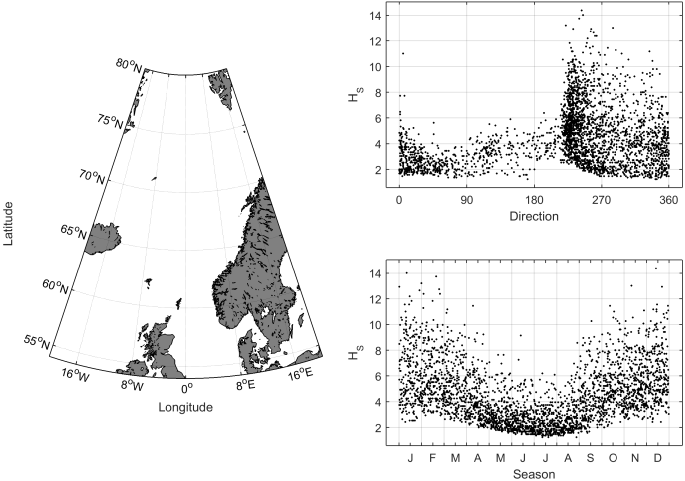
- Rational and consistent design an assessment of marine and coastal structures
  - Reduce bias and uncertainty in estimation of **return values**.
- Non-stationary **marginal** and **conditional** extremes
  - Multiple locations, multiple variables, time-series,
  - **Multidimensional** covariates.
- Improved understanding and communication of risk
  - Incorporation within well-established engineering design practices,
  - **"Knock-on" effects** of "improved" inference.

Motivation

- Environmental extremes vary continuously with multidimensional covariates
  - **Parameters of EV model should be functions of covariates**.
- Inferences should reflect all estimation uncertainty
  - Typically **threshold pre-specified** since sample size changes,
  - Need to estimate a **whole sample model**.
- Piecewise models have been used in the past
  - Empirical evidence that Weibull distribution is reasonable for body of storm severity,
  - Generalised Pareto already used for tail of distribution,
  - **(Truncated) Weibull - generalised Pareto** candidate.
- Need statistical and computational efficiency
  - Slick spline methods (GLAMs),
  - Bayesian inference.

Application: directional-seasonal

Storm peak significant wave height at northern North Sea location; clear directional and seasonal variability in storm severity; directional variability more dramatic at around 225°; seasonal variability more gradual.



Two-part model for y

- Sample of peaks y over threshold ψ, with direction θ, season ϕ
- For $y \leq \psi$, y follows truncated Weibull distribution with shape γ, scale α
- For $y > \psi$, y follows generalised Pareto with shape ξ, scale σ
- ξ, σ, α, γ, ψ all functions of θ and ϕ
- ψ specified with (stationary) threshold probability τ

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

Two-part model for y

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

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

- Hence $\quad \psi|\tau, \alpha, \gamma = \alpha\left(-\log(1-\tau)\right)^{\gamma}$
- No imposition of continuity in the density at ψ
- Other possibilities below threshold

Model for count rate c

- Whole-sample rate of occurrence ρ modelled as Poisson process given counts c of numbers of occurrences per covariate bin
- Chavez-Demoulin and Davison [2005]

Penalised spline smoothers

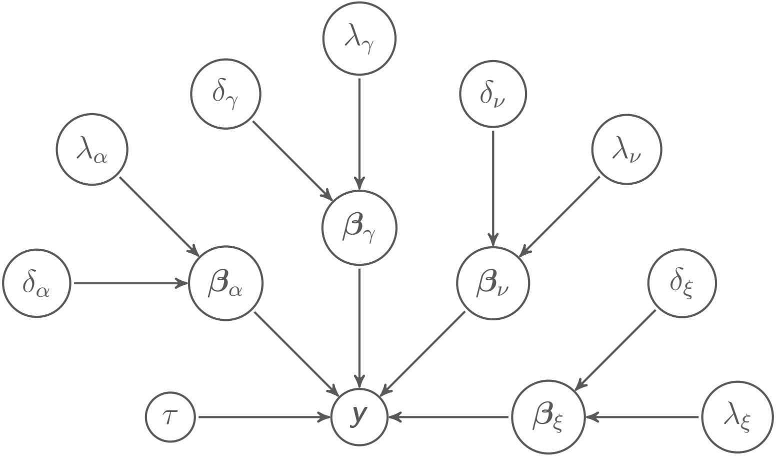
- Physical considerations suggest parameters $\alpha, \gamma, \rho, \xi$ and σ vary smoothly with covariates θ, ϕ
- Values of $\eta \in \{\alpha, \gamma, \rho, \xi, \sigma\}$ on some index set of covariates take the form $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\beta}_\eta$
- $\boldsymbol{B}$ takes the form of a tensor product $\boldsymbol{B}_\phi \otimes \boldsymbol{B}_\theta$
- Spline roughness with respect to covariate $\kappa \in \{\theta, \phi\}$ 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** $\delta_{\eta\kappa}$
- $\delta_{\eta\kappa}$ not estimated here, but could be in principle

- Eilers and Marx [2010], Brezger and Lang [2006]

Prior specification

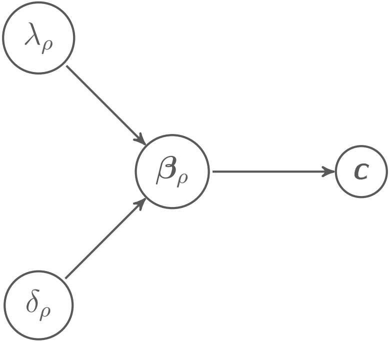
$$
\begin{aligned}
\text{density of } \boldsymbol{\beta}_{\eta\kappa} &\propto \exp\left(-\frac{1}{2}\lambda_{\eta\kappa}\boldsymbol{\beta}'_{\eta\kappa}\boldsymbol{P}_{\eta\kappa}\boldsymbol{\beta}_{\eta\kappa}\right) \\
\lambda_{\eta\kappa} &\sim \text{gamma} \\
\tau &\sim \text{beta}
\end{aligned}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

DAG for y



λ_γ
δ_γ
δ_ν
λ_α
λ_ν
$\boldsymbol{\beta}_\gamma$
δ_α
$\boldsymbol{\beta}_\alpha$
$\boldsymbol{\beta}_\nu$
δ_ξ
τ
$\boldsymbol{y}$
$\boldsymbol{\beta}_\xi$
λ_ξ

DAG for c



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 θ and ϕ

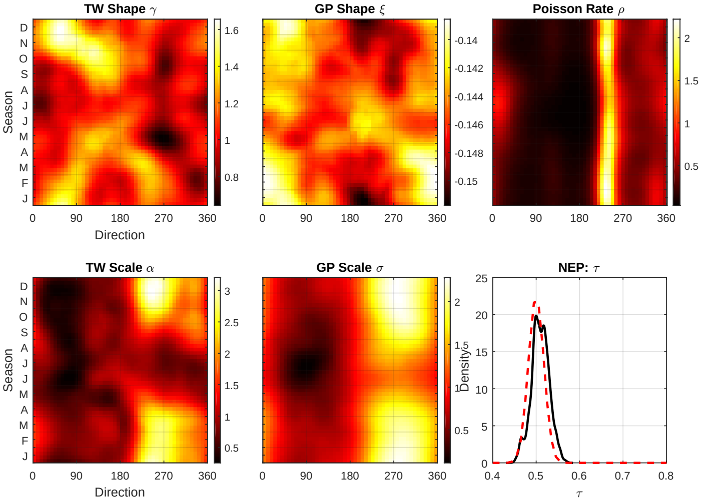
Inference

- Sampling from **full conditionals**
  - Gibbs sampling when full conditionals available,
  - Metropolis-Hastings (MH) within Gibbs otherwise, using suitable proposal mechanisms.
- Conjugacy for $\lambda_{\eta\theta}, \lambda_{\eta\phi}$, so Gibbs sampling
- Simple Gaussian random walk MH for τ
- Correlated Gaussian random walk MH for $\boldsymbol{\beta}_\eta$ when $\eta \in \{\alpha, \gamma\}$
- Correlated Gaussian random walk MH with drift (mMALA) for $\boldsymbol{\beta}_\eta$ when $\eta \in \{\rho, \sigma(\nu), \xi\}$

- Roberts and Stramer [2002], Girolami and Calderhead [2011], Xifara et al. [2014]

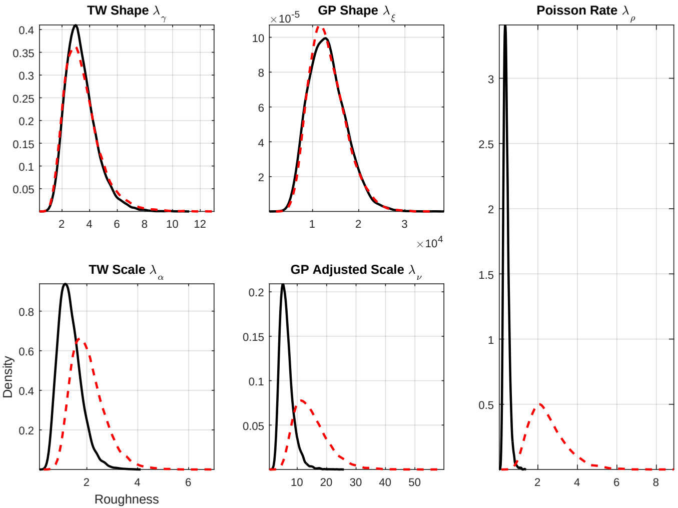
Parameter estimates

Prior for τ in red, posterior in black; all parameters except ξ and τ suggest strong directional variation; seasonal variation less pronounced but clear for α, σ (and ρ); ξ effectively constant; sample not informative about τ.



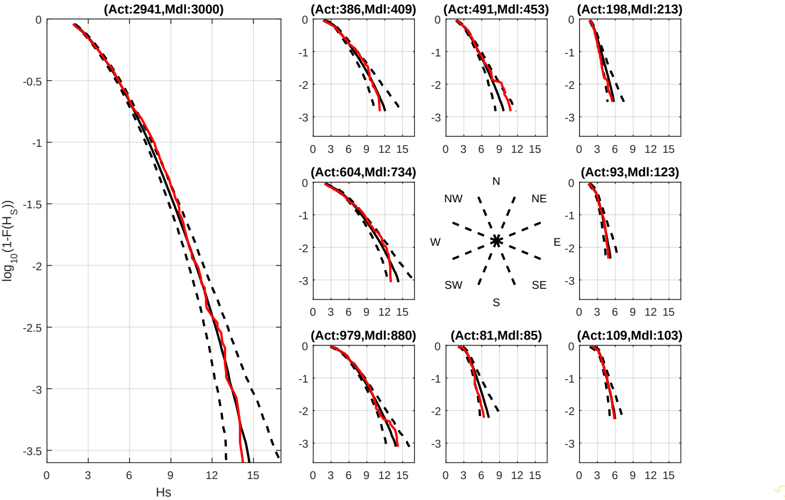
Roughness penalty estimates

Posterior directional roughness penalties in black, posterior seasonal roughness penalties in red; shape penalities $(\lambda_\gamma, \lambda_\xi)$ similar for direction and season; scale penalties $(\lambda_\gamma, \lambda_\xi)$ and rate penalty λ_ρ greater in season than direction $\Rightarrow$ rougher solution in direction than season as expected.



Validation

Estimates for the distribution of H_S (in metres) corresponding to the period of the original sample; empirical estimate in red; predictive distribution from multiple realisations under the directional-seasonal model (median with 2.5% and 97.5% values in black; omni-directional omni-seasonal case on lhs; omni-seasonal estimates for each of 8 directional octants on rhs; titles give numbers of actual events and the median number of events simulated.

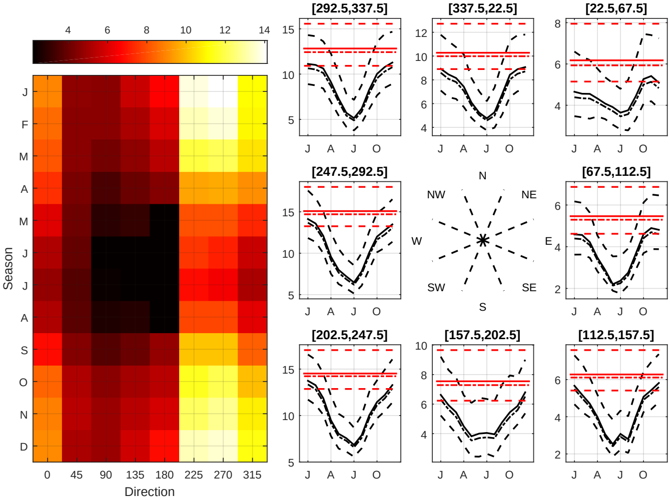


Return values

- Estimated by **simulation** from posterior
- Maximum value in a simulation period of 100–years

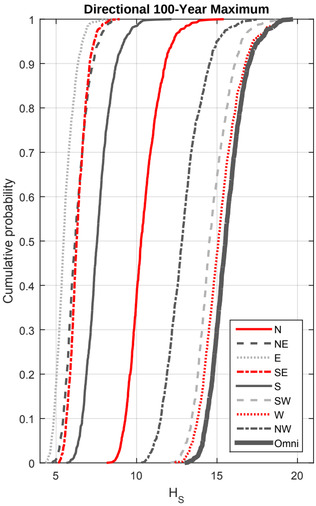
Return values

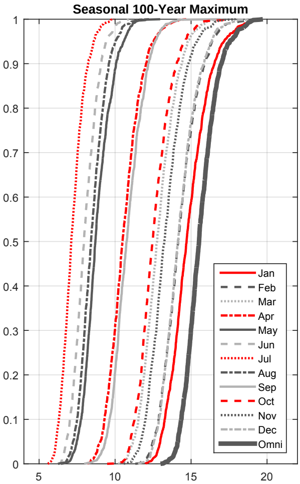
Predictive distribution of the 100-year maximum (in metres); directional and seasonal variability of the median estimate on lhs; seasonal variation of predictive distribution for directional octants (2.5%, 37%, median and 97.5% values) in black; corresponding omni-seasonal estimates in red; large difference between S and SW; smooth seasonal variation.



Return values

Predictive cumulative distribution functions for 100-year maximum H_S (in metres); directional (octant) on lhs; seasonal (monthly) on rhs; W and SW dominate directionally; Dec, Jan and Feb seasonally.





Summary

General

- Modelling non-stationarity essential for understanding extremes and development of design conditions
- Non-parametric covariate models flexible, but need to estimate roughness
- Penalised splines computationally efficient for smoothing over higher dimensional domains

Bayesian penalised splines for smoothing in extremes

- Roughness penalty coefficients estimated using Gibbs sampling
- Different roughness penalty coefficients for each covariate dimension
- Correlated MCMC proposals exploiting gradient
- Computationally more efficient and more stable than optimisation to point solution

References

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

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

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

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

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.

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.

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

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.