Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference
South China Sea Storms
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 and spatial extremes
- Improved understanding and communication of risk
- Incorporation within established engineering design practices
- Knock-on effects of improved inference
South China Sea
Motivation: storm model

\[ H_s \approx 4 \times \text{standard deviation of ocean surface time-series at specific location corresponding to a specified period (typically three hours)} \]
Storm peak data
Storm peak data by bin

(a) Annual rate

(b) Maximum

Copyright of Shell
David Randell, EVA 2017
Storm trajectories
Marginal: gamma-GP model

- Sample of peaks over threshold $y$, with covariates $\theta$
  - $\theta$ is 1D in motivating example: directional
  - $\theta$ is $nD$ later: e.g. 4D spatio-directional-seasonal

- Below threshold $\psi$
  - $y$ follows truncated gamma with shape $\alpha$, scale $1/\beta$
  - Hessian for gamma better behaved than Weibull

- Above $\psi$
  - $y$ follows generalised Pareto with shape $\xi$, scale $\sigma$

- $\xi$, $\sigma$, $\alpha$, $\beta$, $\psi$ all functions of $\theta$

- $\psi$ for pre-specified threshold probability $\tau$

- Frigessi et al. [2002], Behrens et al. [2004], MacDonald et al. [2011]
Gamma-generalised Pareto model for extremes

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

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

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

$$L(\xi, \sigma, \alpha, \beta, \psi, \tau|\{y_i\}_{i=1}^n) = \prod_{i:y_i \leq \psi} f_{TG}(y_i|\alpha, \beta, \psi) \prod_{i:y_i > \psi} f_{GP}(y_i|\xi, \sigma, \psi) \times \tau^{n_B}(1 - \tau)^{(1-n_B)} \text{ where } n_B = \sum_{i:y_i \leq \psi} 1.$$ 

Estimate all parameters as functions of $\theta$
Marginal: count rate $c$

- Whole-sample rate of occurrence $\rho$ modelled as Poisson process given counts $c$ of numbers of occurrences per covariate bin

- Chavez-Demoulin and Davison [2005]
Threshold effect

(a) Negative log likelihood

(b) 100-year return value
Threshold effect

(a)

(b)

(c)

(d)

Storm peak $H_S$

Density

Copyright of Shell

David Randell, EVA 2017
Marginal: priors and conditional structure

Priors

density of $\beta_{\eta \kappa}$ $\propto \exp\left(-\frac{1}{2} \lambda_{\eta \kappa} \beta_{\eta \kappa}' \varmathbb{P}_{\eta \kappa} \beta_{\eta \kappa}\right)$

$\lambda_{\eta \kappa} \sim \text{gamma}$

Conditional structure

$f(\beta_{\eta} | \mathbf{y}, \Omega \setminus \beta_{\eta})$ $\propto$ $f(\mathbf{y} | \beta_{\eta}, \Omega \setminus \beta_{\eta}) \times f(\beta_{\eta} | \delta_{\eta}, \lambda_{\eta})$

$f(\lambda_{\eta} | \mathbf{y}, \Omega \setminus \lambda_{\eta})$ $\propto$ $f(\beta_{\eta} | \delta_{\eta}, \lambda_{\eta}) \times f(\lambda_{\eta})$

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

$\tau$ is not estimated
Inference

- Elements of $\beta_\eta$ 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]
Posterior parameter estimates

(a) \( \psi \)  
(b) \( \alpha \)  
(c) \( \sigma \)  
(d) \( \rho \)  
(e) \( \zeta \)  
(f) \( \xi \)
Validation

Compare sample with simulated values on partitioned covariate domain
Return values

To get directional return values we can do 2 main approaches

- Monte Carlo simulation: easy to understand and simple to implement but slow. 10000 year events can take over a day to compute for the complex models we fit.
- Numerical integration: much faster, 100 fold improvement in return value calculation time.
Return values from Monte Carlo simulation

- Consider directional-seasonal bin $S_j (j = 1, 2, ..., m)$ centred on location $l_j$. $S_j$ is sufficiently small that all model parameters $\rho$ are assumed constant within it.

- For each realisation $i$, for each covariate bin $j$, with $\omega_j = \{\alpha_j, \zeta_j, \xi_j, \nu_j, \psi_j\}$
  1. Sample the number of storms
     \[ n_{ij} \sim \text{Poisson}(\rho_j) \]
     where $\rho_j$ is the annual rate of occurrence.
  2. Sample $n_{ij} \times T$ values from
     \[ Y_{ij} \sim \text{GammaGP}(\omega_j) \]
     where $T$ is the return period.

- $T$-year return values in $S_j$ are then found by taking maximum over in each realisation and then finding the empirical cdf

- Bins can be combined by taking maximum over bins.
Numerical integration of return values storm peaks

We define $F(y|\omega_i)$ to be the cumulative distribution function of any storm peak event given $\omega_i$. We estimate the cumulative distribution function $F_{Mt}(y|\omega_i)$

$$F_{Mt}(y|\omega_i) = P(M_T < y)$$

$$= \sum_{k=0}^{\infty} P(k \text{ events in } S_i \text{ in } T \text{ years}) \times P^k (\text{size of an event in } S_i < y)$$

$$= \sum_{k=0}^{\infty} \frac{(T\rho_i)^k}{k!} \exp(-T\rho_i) \times F^k(y|\omega_i)$$

$$= \exp (-T\rho_i (1 - F(y|\omega_i))).$$
Posterior predictive return values across bins

Since storm peak events are independent given covariates, we combine by taking the product

$$F_{MT}(y|\omega) = \prod_{j=1}^{m} F_{MT}(y|\omega_j)$$

The final estimate for $F_{MT}(y)$, unconditional on $\omega$, is estimated by marginalising over $\omega$

$$F_{MT}(y) = \int_{\omega} F_{MT}(y|\omega) f(\omega) d\omega$$

where $f(\omega)$ is the estimated posterior density for $\omega$. 
Empirical dissipation shapes
For applications, it is also necessary to estimate the distribution of return value \( M_{TS}(y) \) for maximum of sea state.

We empirically estimate the storm dissipation function \( \delta(S; j, y) \) for sea state \( H_S \) in directional sector \( S \) estimated from the sample of storm trajectories.

Next we estimate the cumulative distribution function \( F_{D_S}(d|\omega_j) \) of \( D_S \), the dissipated sea state \( H_S \) in sector \( S \) from a random storm dissipating from directional-seasonal bin \( S_j \)

\[
F_{D_S}(d|\omega_j) = P(D_S \leq d|\omega_j) = \int_y P(\delta(S; j, Y) \leq d|Y = y)f(y|\omega_j)dy
\]

where \( f(y|\omega_j) \) is the marginal directional density of storm peak \( H_S \) in directional-seasonal bin \( S_j \) corresponding to cumulative distribution function \( F(y|\omega_j) \).
10000 years return values by direction
10000 year return values by season
Summary

- Evidence for covariate effects in marginal extremes of ocean storms
  - Modelling non-stationarity essential for understanding extreme ocean storms, and estimating marine risk well
  - Non-parametric P-spline flexible basis for covariate description
  - Essential that non-stationary models are used for marginal, conditional and spatial extremes inference of ocean environment
  - Cradle-to-grave uncertainty quantification

- Numerical integration of return value provides a much faster way to estimate return values without the need to resort to Monte Carlo simulation.

- Looking at way of modelling dissipation to avoid the empirical resampling

- Paper accepted Ocean Engineering on Monday!
  http://www.lancs.ac.uk/ jonathan/RssEAMrgBys17.pdf
References


