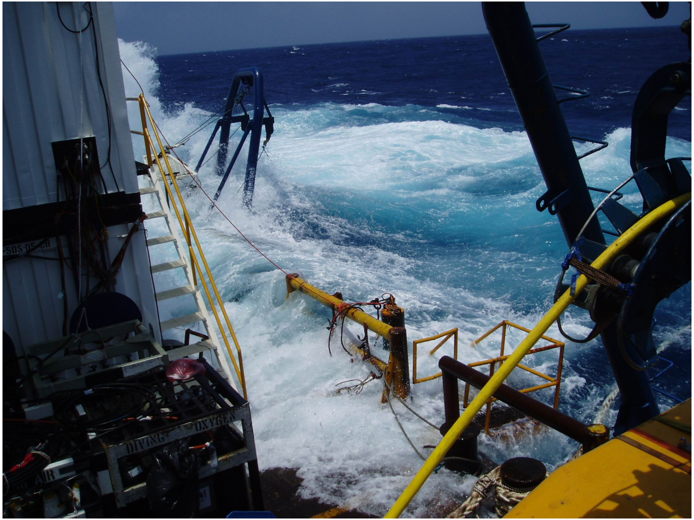
Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference

David Randell, Emma Ross, Philip Jonathan
David Randell, EVA 2017

South China Sea Storms



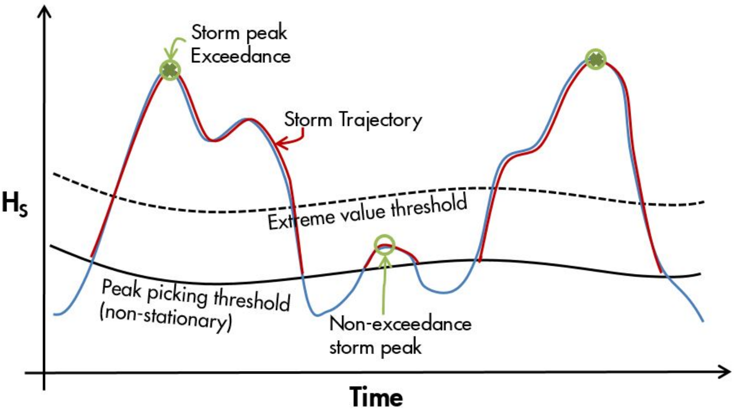
Motivation

- 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

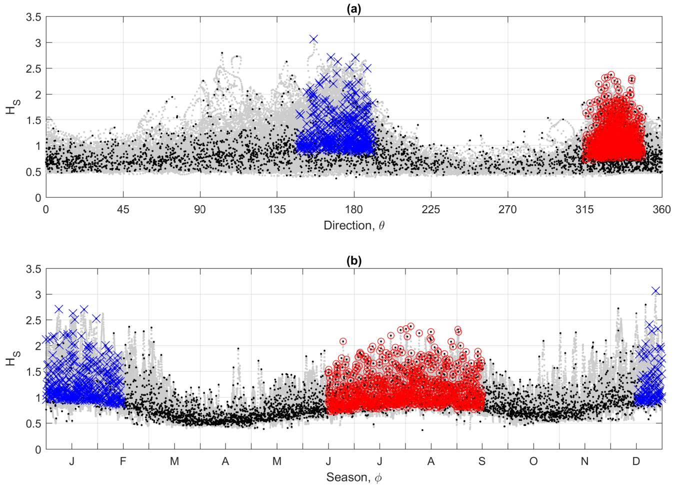
South China Sea

Motivation: storm model

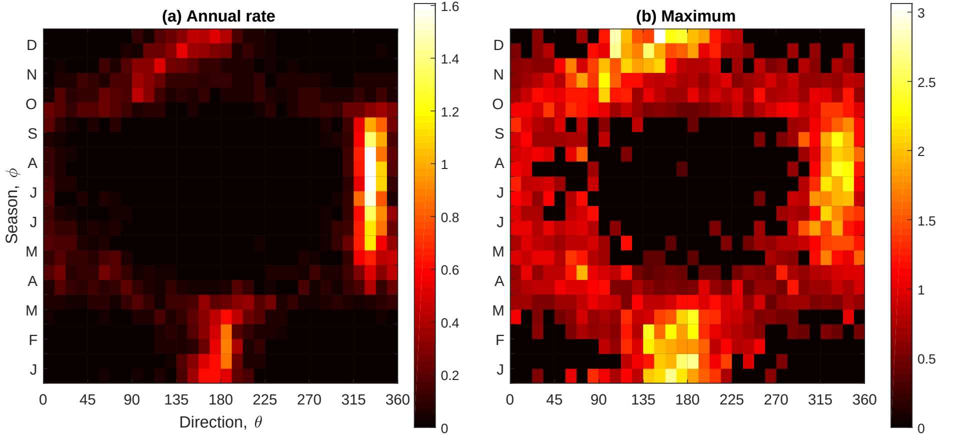
$H_S \approx 4\times$ standard deviation of ocean surface time-series at specific location corresponding to a specified period (typically three hours)



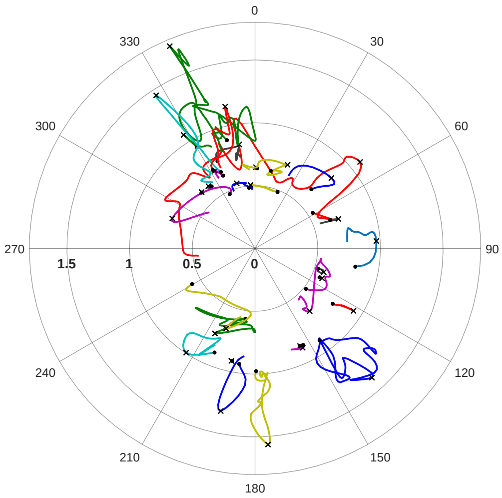
Storm peak data



Storm peak data by bin



Storm trajectories



Marginal: 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

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

Gamma-generalised Pareto model for extremes

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

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

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

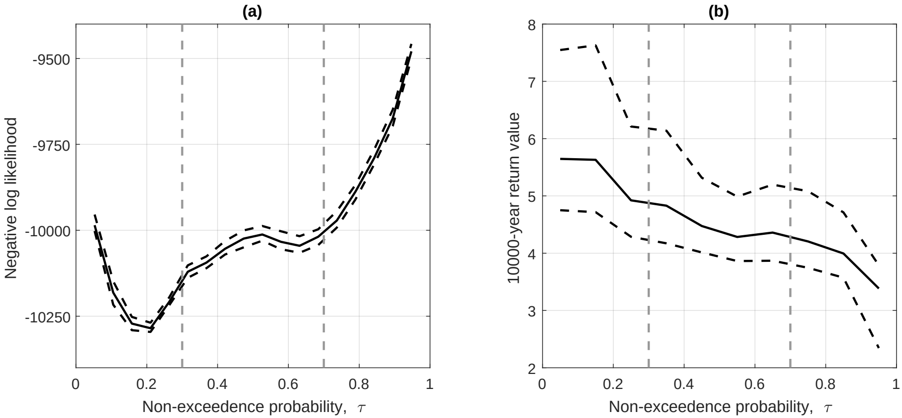
$$= \prod_{i:y_i\leq\psi} f_{TG}(y_i|\alpha,\beta,\psi) \prod_{i:y_i>\psi} f_{GP}(y_i|\xi,\sigma,\psi)$$
$$\times \quad \tau^{n_B}(1-\tau)^{(1-n_B)} \text{ where } n_B = \sum_{i:y_i\leq\psi} 1.$$

Estimate all parameters as functions of θ

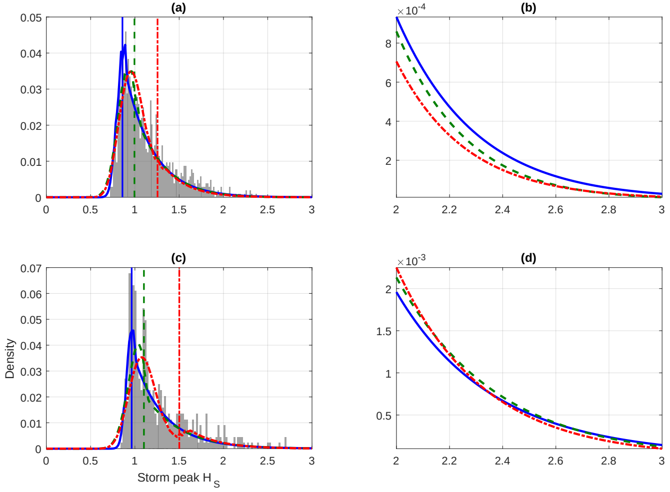
Marginal: count rate c

- 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]

Threshold effect



Threshold effect



Marginal: priors and conditional structure

Priors

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

Conditional structure

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

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

τ is not estimated

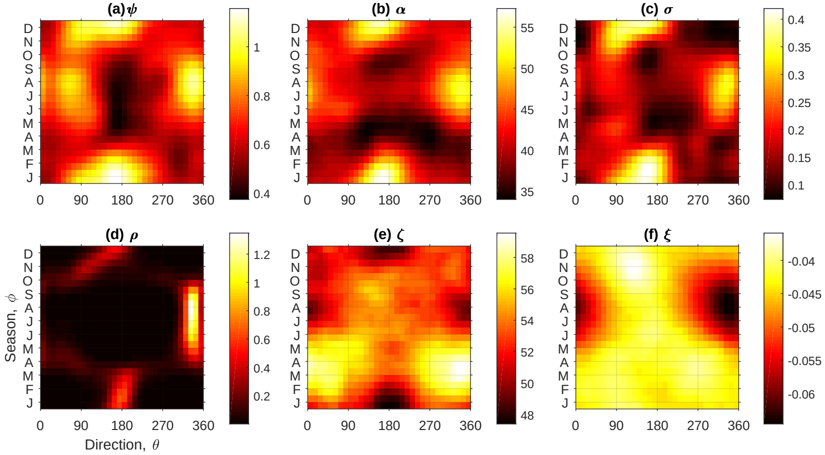
Inference

- Elements of β_η 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

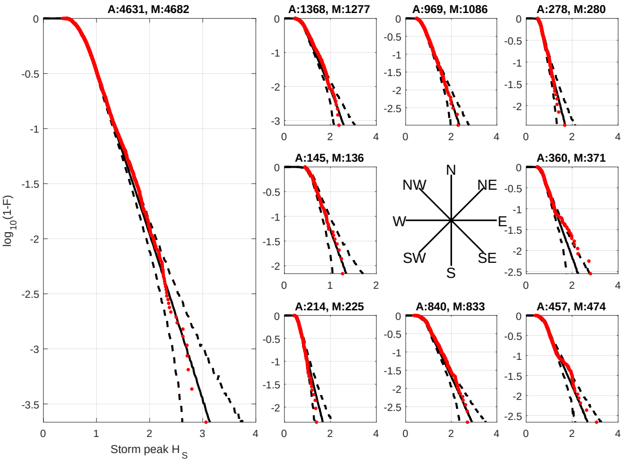
- Roberts and Stramer [2002], Girolami and Calderhead [2011], Xifara et al. [2014]

Posterior parameter estimates



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 ρ 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
  $$n_{ij} \sim Poisson(\rho_j)$$
  where ρ_j is the annual rate of occurrence.
  2. Sample $n_{ij} * T$ values from
  $$Y_{ij} \sim GammaGP(\omega_j)$$
  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 ω_i.
We estimate the cumulative distribution function $F_{M_T}(y|\omega_i)$

$$
\begin{aligned}
F_{M_T}(y|\omega_i) &= \mathbb{P}\left(M_T < y\right) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\left(k \text{ events in } S_i \text{ in } T \text{ years }\right) \times \mathbb{P}^k\left(\text{size of an event in } S_i < y\right) \\
&= \sum_{k=0}^{\infty} \frac{(T\rho_i)^k}{k!} \exp(-T\rho_i) \times F^k(y|\omega_i) \\
&= \exp\left(-T\rho_i\left(1 - F(y|\omega_i)\right)\right).
\end{aligned}
$$

Posterior predictive return values across bins

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

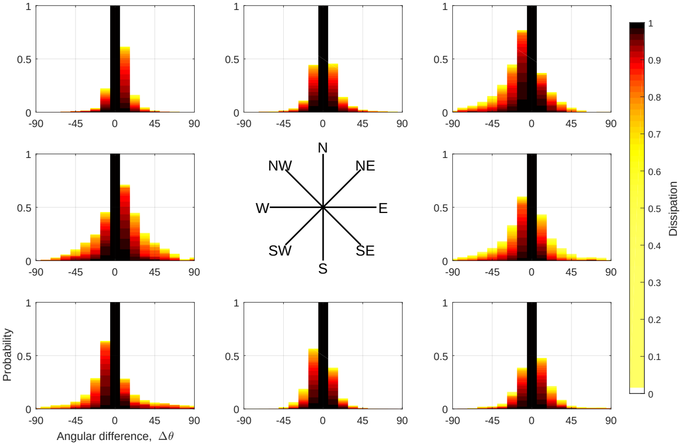
$$F_{M_T}(y|\omega) = \prod_{j=1}^{m} F_{M_T}(y|\omega_j)$$

- The final estimate for $F_{M_T}(y)$, unconditional on ω, is estimated by marginalising over ω

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

where $f(\omega)$ is the estimated posterior density for ω.

Empirical dissipation shapes



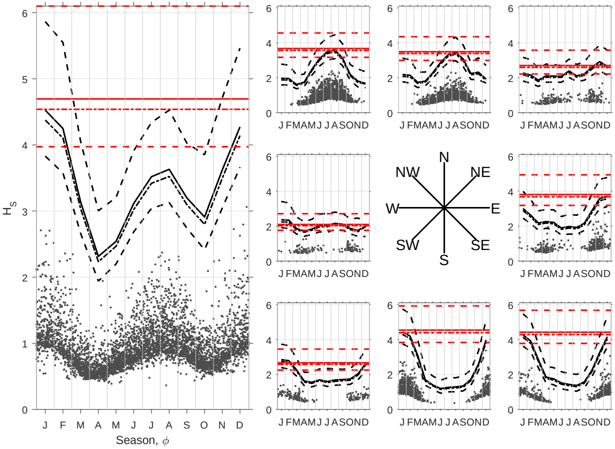
Numerical integration of return values with dissipation

- For applications, it is also necessary to estimate the distribution of return value $M_{T\mathbb{S}}(y)$ for maximum of sea state.
- We empirically estimate the storm dissipation function $\delta(\mathbb{S}; j, y)$ for sea state H_S in directional sector $\mathbb{S}$ estimated from the sample of storm trajectories.
- Next we estimate the cumulative distribution function $F_{D_\mathbb{S}}(d|\omega_j)$ of $D_\mathbb{S}$, the dissipated sea state H_S in sector $\mathbb{S}$ from a random storm dissipating from directional-seasonal bin S_j

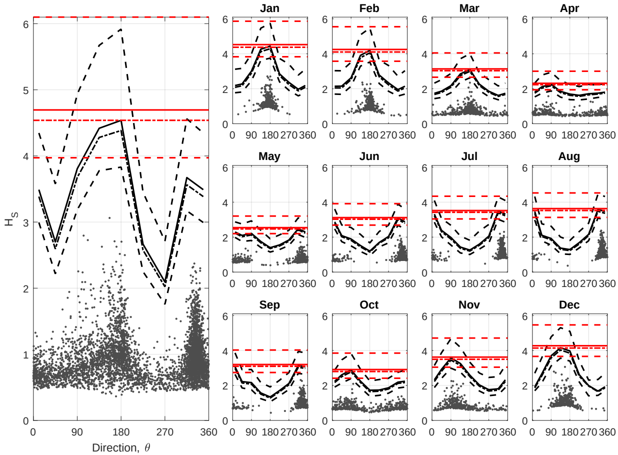
$$F_{D_\mathbb{S}}(d|\omega_j) = \mathbb{P}(D_\mathbb{S} \leq d|\omega_j) = \int_y \mathbb{P}(\delta(\mathbb{S}; j, Y) \leq d|Y = y) f(y|\omega_j) dy$$

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

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

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.

Arnoldo Frigessi, Ola Haug, and Hå vard 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.

Gareth O Roberts and Osnat Stramer. Langevin diffusions and metropolis-hastings algorithms. *Methodology and computing in applied probability*, 4(4):337–357, 2002.

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