A spatio-directional splines model for extreme waves in the Gulf of Mexico

Yanyun Wu[1] Philip Jonathan[2] David Randell[2]

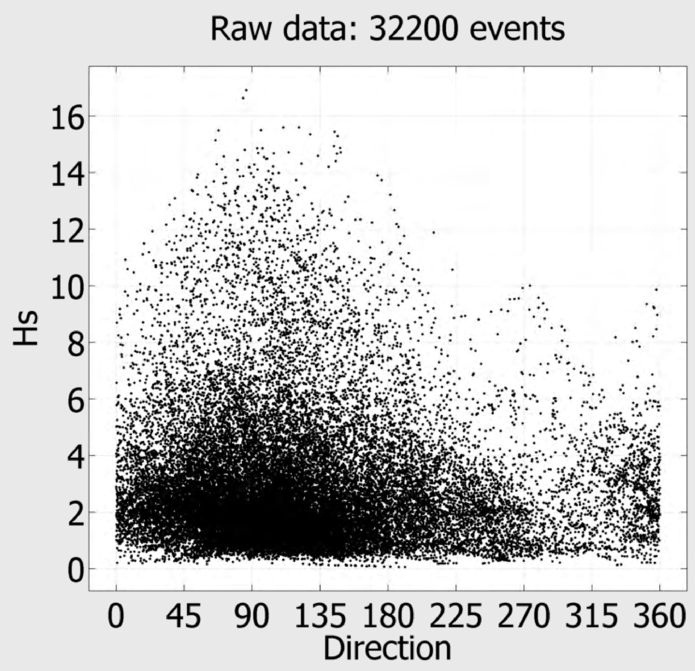
[1]Shell Research Ltd., Maths & Stats, Lancaster University; [2]Shell Research Ltd..

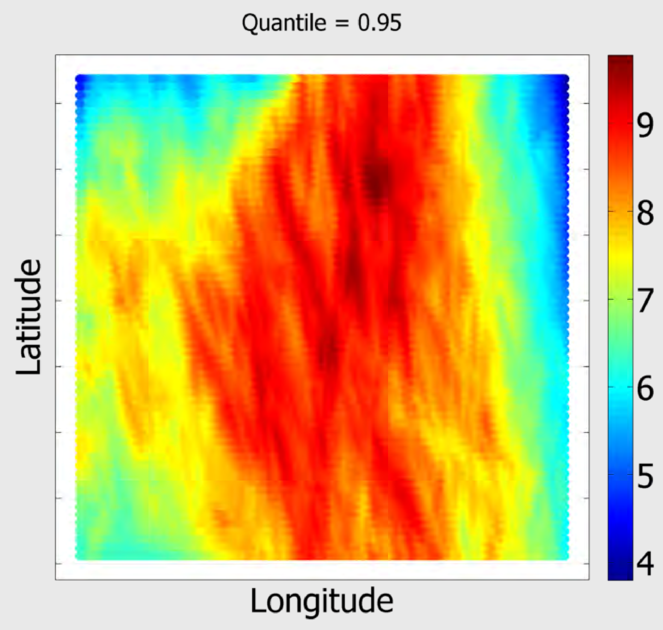
Introduction

Storm severity varies systematically with numerous covariates, including location and storm direction. Reliable estimation of design criteria requires adequate incorporation of covariate effects within extreme value models.

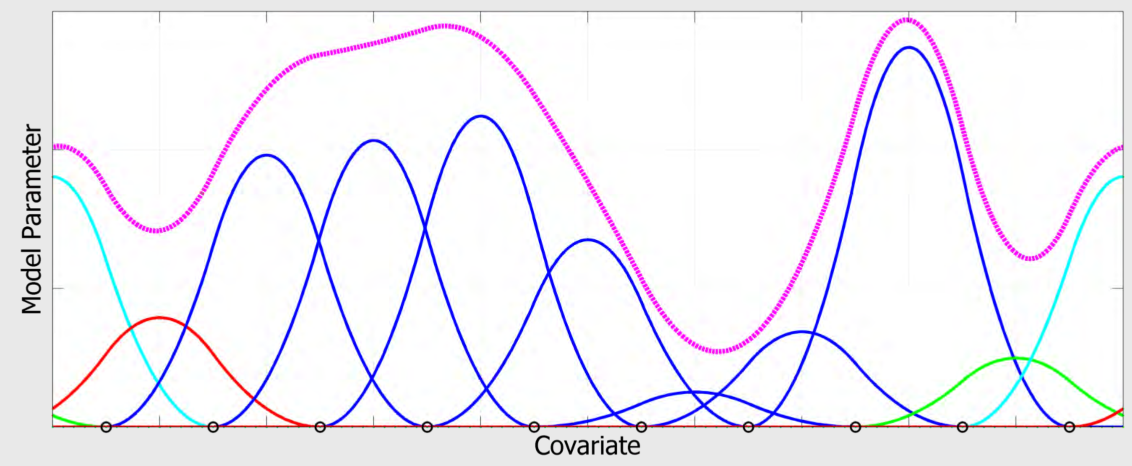
We present a spatio-directional model for extreme storm peak significant wave heights in the Gulf of Mexico based on the nonhomogeneous Poisson model for peaks over threshold. The model parameters are expressed in terms of spatio-directional p-spline forms throughout.

The model is illustrated using a neighbourhood of spatial locations from the proprietary Gulf of Mexico Oceanographic Study (GOMOS) hindcast for the US region of the Gulf of Mexico for the period of 1900 - 2005.





Modelling covariate effects using splines



Consider estimation using negative log likelihood (or other lack of fit criterion) ℓ for sample $\mathbf{y}$ and model parameter $\boldsymbol{\theta}$, expressed as $\boldsymbol{\theta} = \mathbf{B}\boldsymbol{\beta}$, where B is a basis of B-spline functions [1]. We solve for $\boldsymbol{\beta}$ by minimising the penalised negative log likelihood

$$\ell^* = \ell + \lambda \|\mathbf{D}\boldsymbol{\beta}\|^2$$

where $\mathbf{D}$ is a difference matrix and λ a roughness parameter estimated using cross validation [2].

The splines are flexible in covariate modelling:

- the model provides local fit;
- the smoothness is determined by the data;
- the model can easily incorporate periodic features.

Modelling marginal extremes with spline covariates

Quantile regression

Quantile regression [3] is used to estimate a threshold $\gamma(\tau)$ for a given quantile probability τ given a basis matrix $\mathbf{B}$ and lack of fit criterion ℓ

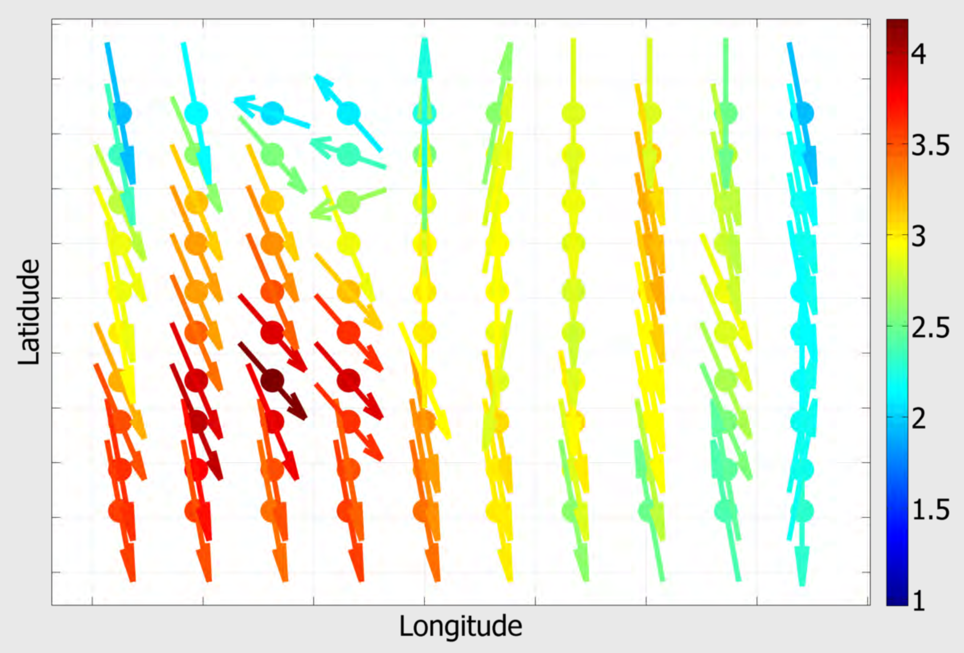
$$\ell = \tau \sum_{\varepsilon_i \geq 0}^{n} |\varepsilon_i| + (1-\tau) \sum_{\varepsilon_i < 0}^{n} |\varepsilon_i|$$

where

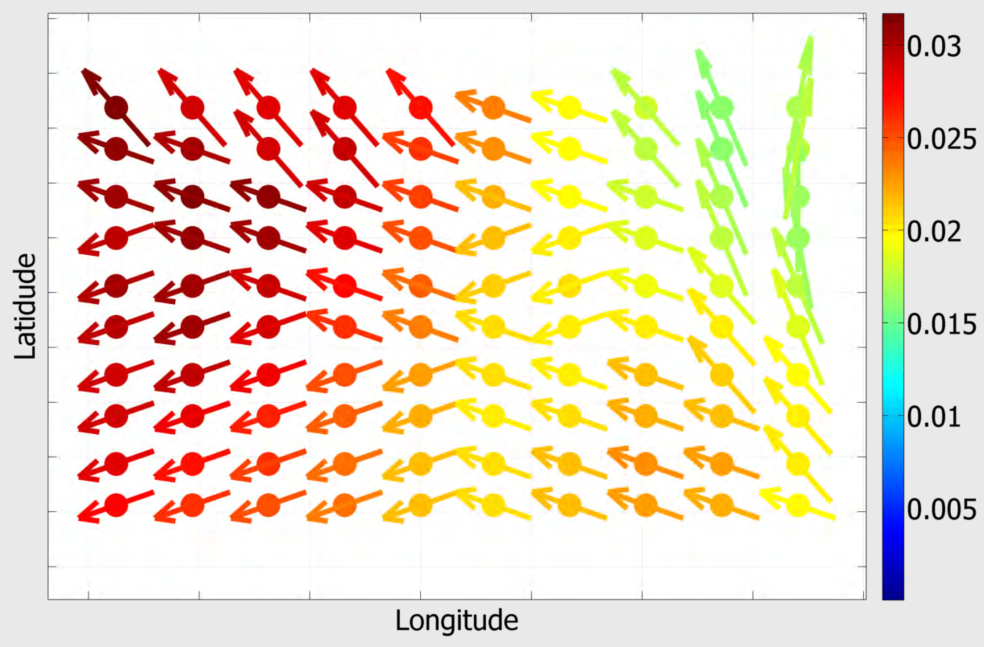
$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{B}\gamma$$

and γ minimises the penalised criterion

$$\ell^* = \ell + \lambda_\gamma \|\mathbf{D}\boldsymbol{\gamma}\|^2 .$$



Poisson Rate



The rate of threshold exceedance per covariate element is modelled using the Poisson process with rate ρ, with negative log likelihood:

$$\ell = \sum_j (\rho_j \Delta - c_j \log \rho_j)$$

where Δ is volume of covariate element j, with the number of occurrence c_j. The square root r of rate ρ is assumed to have p-spline representation $\mathbf{B}\boldsymbol{\zeta}$. The penalized likelihood is

$$\ell^* = \sum_j (r_j^2 \Delta - c_j log r_j^2) + \lambda_\zeta \|\mathbf{D}\boldsymbol{\zeta}\|^2$$

Generalised Pareto Distribution

We model the exceedances y of high threshold γ using the generalised Pareto distribution [4].

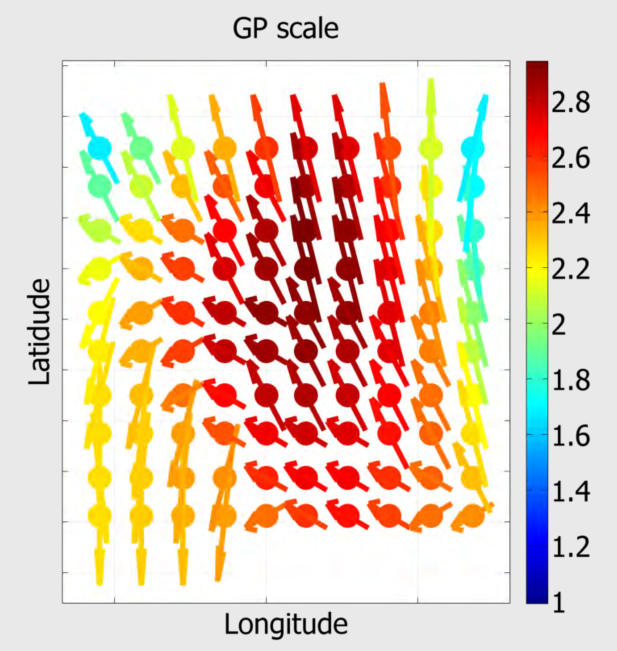
$$\mathbb{P}(Y > y | y > \gamma) = (1 + \frac{\xi}{\sigma}(y - \gamma))^{-1/\xi}, \quad y > \gamma$$

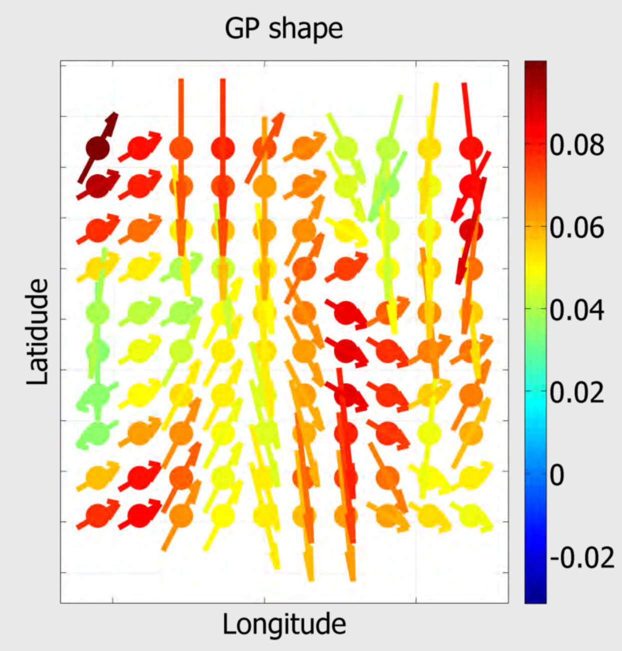
Using p-spline representations for the GP parameters,

$$\boldsymbol{\xi} = \mathbf{B}\boldsymbol{\alpha}, \quad \boldsymbol{\nu} = \mathbf{B}\boldsymbol{\beta}$$

and negative likelihood function ℓ, the penalized GP likelihood is:

$$\ell^* = \ell + \lambda_\alpha \|\mathbf{D}\boldsymbol{\alpha}\|^2 + \lambda_\beta \|\mathbf{D}\boldsymbol{\beta}\|^2 .$$





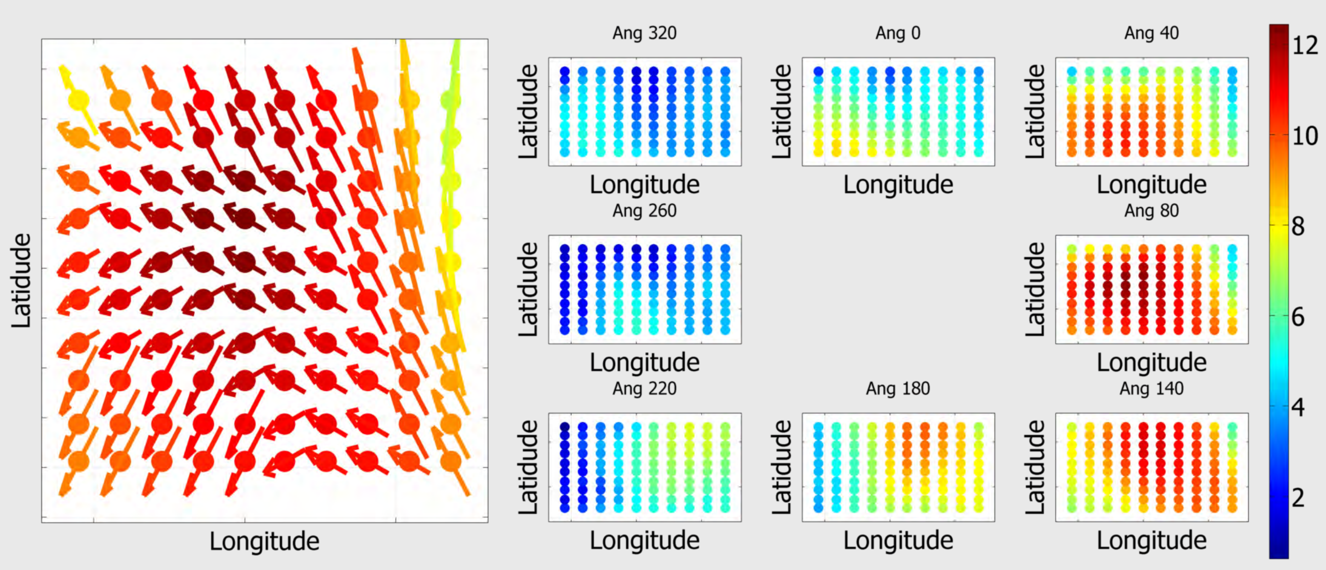
Computational considerations

- For generalised Pareto estimation, the back-fitting algorithm applies Newton-Raphson to likelihood scores $\frac{\partial \ell}{\partial \beta_a}$ for parameter β_a $(a = 1, 2, ..., p)$:

$$\left.\frac{\partial \ell}{\partial \beta_a}\right|_{\hat{\beta}} = \left.\frac{\partial \ell}{\partial \beta_a}\right|_{\beta_0} + \sum_b \left.\frac{\partial^2 \ell}{\partial \beta_a \partial \beta_b}\right|_{\beta_0} (\hat{\beta}_b - \beta_{0,b}) = 0$$

  to update current estimate $\boldsymbol{\beta}_0$ to an improved estimate $\hat{\boldsymbol{\beta}}$.
- Covariate binning is used to increase the computational speed.

Median 100-year return level



Conclusion & discussion

The spatio-directional splines model provides a feasible route to basin-wide marginal extremes modelling with covariates

- Flexible multivariate spline representation of covariates
- Quantile regression for threshold selection
- Poisson model for rate of occurrence of threshold exceedance
- Generalised Pareto model for size of occurrence of threshold exceedance

Extension to include multivariate spatial dependence

- Further computation improvements (e.g. simultaneous quantile regression, coarse and fine grid splines)
- Censored likelihood to apply component-maxima ideas to threshold exceedances
- Composite (pairwise) likelihood to approximate full (dependence) likelihood
- Mixture model to incorporate asymptotic independence and dependence

References

- Eilers, P. H. C. and Marx,B. D., 1996. Flexible Smoothing with B-splines and Penalties. Statistical Science, **11**, 89–121.
- Randell, D. et al, 2013, Modelling Covariate Effects In Extremes Of Storm Severity On The Australian North West Shelf, OMAE2013-10187.
- Koenker, R., 2005. Quantile Regression. Cambridge University Press.
- Coles, S.G., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.

yanyun.wu@shell.com