Modelling covariate effects in extremes of storm severity on the Australian North West Shelf

David Randell, Philip Jonathan, Kevin Ewans, Yanyun Wu
david.randell@shell.com

Shell Technology Centre Thornton, Chester, UK

SIMRIDE March 2013
Acknowledgements

Thanks for work done by summer students

- Emma Ross
- Kaylea Haynes
Outline

1 Background
   - Motivation
   - Australian North West Shelf

2 Extreme Value Analysis: Challenges

3 Modelling Covariates
   - Model Components
   - P-Splines
   - Quantile regression models threshold
   - Poisson models rate of threshold exceedances
   - GP models size of threshold exceedances
   - Return Values

4 Other Applications and Developments
Contents

1 Background
   • Motivation
   • Australian North West Shelf

2 Extreme Value Analysis: Challenges

3 Modelling Covariates
   • Model Components
   • P-Splines
   • Quantile regression models threshold exceedances
   • Poisson models rate of threshold exceedances
   • GP models size of threshold exceedances
   • Return Values

4 Other Applications and Developments
Motivation

- **Rational** design an assessment of marine structures:
  - Reducing *bias* and *uncertainty* in estimation of structural reliability.
  - Improved understanding and communication of risk.
  - Climate change.

- Other applied fields for extremes in industry:
  - Corrosion and fouling.
  - Finance.
  - Network traffic.
Katrina in the Gulf of Mexico.
Katrina damage.
Platform in a Northern North Sea storm.
Platform in the Southern North Sea.
A wave seen from a ship.
Australian North West Shelf
Model storm peak significant wave height $H_S$.
Wave climate is dominated by westerly monsoonal swell and tropical cyclones.
Cyclones originate from Eastern Indian Ocean and in the Timor and Arafura Sea area is also a region of cyclogenesis.
Storm Peak $H_S$ by Direction

Raw data: 6156 events
Quantiles of storm peak $H_S$ Spatially
1 Background
   - Motivation
   - Australian North West Shelf

2 Extreme Value Analysis: Challenges

3 Modelling Covariates
   - Model Components
   - P-Splines
   - Quantile regression models threshold exceedances
   - Poisson models rate of threshold exceedances
   - GP models size of threshold exceedances
   - Return Values

4 Other Applications and Developments
Extreme Value Analysis: Challenges

- **Covariate** effects:
  - Location, direction, season, time ...
  - Multiple covariates in practice.

- **Cluster** dependence:
  - e.g. storms independent, observed (many times) at many locations.
  - e.g. dependent occurrences in time.
  - estimated using e.g. extremal index (Ledford and Tawn 2003)

- **Scale** effects:
  - Modelling $X^2$ gives different estimates c.f. modelling $X$. (Reeve et al. 2012)

- **Threshold** estimation.

- **Parameter** estimation.

- **Measurement** issues:
  - Field measurement uncertainty greatest for extreme values.
  - Hindcast data are simulations based on pragmatic physics, calibrated to historical observation.
Multivariate extremes:
- Waves, winds, currents, forces, moments, displacements, ...
- Componentwise maxima $\Leftrightarrow$ max-stability $\Leftrightarrow$ multivariate regular variation:
  - Assumes all components extreme.
  - $\Rightarrow$ Perfect independence or asymptotic dependence only.
- Extremal dependence:
  - Assumes regular variation of joint survivor function.
  - Gives rise to more general forms of extremal dependence.
  - $\Rightarrow$ Asymptotic dependence, asymptotic independence (with +ve, -ve association).
- Conditional extremes:
  - Assumes, given one variable being extreme, convergence of distribution of remaining variables.
  - Not equivalent to extremal dependence.
  - Allows some variables not to be extreme.
- Inference:
  - ... a huge gap in the theory and practice of multivariate extremes ...
    (Beirlant et al. 2004)
Contents

1. Background
   - Motivation
   - Australian North West Shelf

2. Extreme Value Analysis: Challenges

3. Modelling Covariates
   - Model Components
   - P-Splines
   - Quantile regression models threshold
   - Poisson models rate of threshold exceedances
   - GP models size of threshold exceedances
   - Return Values

4. Other Applications and Developments
Sample \( \{ \hat{z}_i \}_{i=1}^n \) of \( n \) storm peak significant wave heights observed at locations \( \{ \hat{x}_i, \hat{y}_i \}_{i=1}^n \) with storm peak directions \( \{ \hat{\theta}_i \}_{i=1}^n \).

- **Model Components**
  1. **Threshold** function \( \phi \) above which observations \( \hat{z} \) are assumed to be extreme estimated using quantile regression.
  2. **Rate of occurrence** of threshold exceedances modelled using Poisson Process model with rate \( \rho(\Delta) = \rho(\theta, x, y) \).
  3. **Size of occurrence** of threshold exceedance using a generalised Pareto (GP) model with shape and scale parameters \( \xi \) and \( \sigma \).
Rate of occurrence and size of threshold exceedance are functionally independent (Chavez-Demoulin and Davison 2005).

Equivalent to non-homogeneous Poisson point process model (Dixon et al. 1998).

Smooth functions of covariates are estimated using P-splines (Eilers and Marx 2010).
Physical considerations suggest that we should expect the model parameters $\phi, \rho, \xi$ and $\sigma$ to vary smoothly with respect to covariates $\theta, x, y$.

$n$ dimensional basis matrix $B$ formulated using Kronecker products of marginal basis matrices

$$B = B_\theta \otimes B_x \otimes B_y$$

Roughness is defined

$$R = \beta' P \beta$$

where $P$ is penalty matrix formed by taking differences of neighbouring $\beta$. 
P-Splines

- Wrapped bases allows for periodic covariates such as seasonality or direction.

- High dimensional bases can easily be constructed although number of parameters problematic.

- Strength of roughness penalty is controlled by roughness coefficient $\lambda$: cross validation is used to choose $\lambda$ optimally.
Quantile regression models threshold

- Estimate smooth quantile \( \phi(\theta_i, x_i, y_i; \tau) \) for non-exceedance probability \( \tau \) of storm peak \( H_S \).

Spline basis:

\[
\psi(\tau, \theta) = \sum_{k=0}^{p} \Phi_{\theta_k} \beta_{\tau_k}
\]

- Estimated by minimising penalised criterion \( \ell^*_\phi \) with respect to basis parameters:

\[
\ell^*_\phi = \{ \tau \sum_{r_i \geq 0} |r_i| + (1 - \tau) \sum_{r_i < 0} |r_i| \} + \lambda_\phi R_\phi
\]

for \( r_i = z_i - \phi(\theta_i, x_i, y_i; \tau) \) for \( i = 1, 2, \ldots, n \), and roughness \( R_\phi \) controlled by roughness coefficient \( \lambda_\phi \).

- Quantile regression with P-splines can be formulated and solved as a linear program (Bollaerts et al. 2006).
Marginal 50% Quantile Threshold

Spatio-directional threshold estimation

Value vs. Covariate 1: Direction

Value vs. Covariate 2: Longitude

Value vs. Covariate 3: Latitude
Spatio-Directional 50% Quantile Threshold
Cross Validation for Penalty

QR lack of fit as a function of penalty

David Randell (Shell)
Modelling covariate effects
SIMRIDE March 2013

Lack of fit

log_{10}(penalty)

2400
2600
2800
3000
3200
3400
3600
3800
-2
0
2
4
6
8
Poisson models rate of threshold exceedances

- Rate of occurrence of threshold exceedances is estimated by minimizing the roughness penalised log likelihood

\[ \ell^*_\rho = \ell_\rho + \lambda_\rho R_\rho \]

- (Negative) penalised Poisson log-likelihood for **rate of occurrence** of threshold excesses:

\[ \ell_\rho = -\sum_{i=1}^{n} \log \rho(\theta_i, x_i, y_i) + \int \rho(\theta, x, y) d\theta dx dy \]

- \( \lambda_\rho \) is estimated using cross validation.
Marginal Rate of Threshold Exceedances

Spatio-directional threshold exceedence rate

Covariate 1: Direction

Covariate 2: Longitude

Covariate 3: Latitude

David Randell (Shell)
Modelling covariate effects
Spatio-Directional Rate of Threshold Exceedances

David Randell (Shell)

Modelling covariate effects

SIMRIDE March 2013
Cross Validation for Penalty

Poisson lack of fit as a function of penalty

Lack of fit

log_{10}(penalty)
Generalised Pareto density (and negative conditional log-likelihood) for sizes of threshold exceedances:

$$
\ell_{\xi,\sigma} = \sum_{i=1}^{n} \log \sigma_i + \frac{1}{\xi_i} \log \left( 1 + \frac{\xi_i}{\sigma_i} (z_i - \phi_i) \right)
$$

Parameters: shape $\xi$, scale $\sigma$.

Threshold $\phi_i$ set prior to estimation.

Smoothness is imposed by minimising the roughness penalised log-likelihood.

$$
\ell^*_{\xi,\sigma} = \ell_{\xi,\sigma} + \lambda_\xi R_\xi + \lambda_\sigma R_\sigma
$$

$\lambda_\xi$ and $\lambda_\sigma$ are estimated using cross validation. In practice set $\lambda_\xi = \kappa \lambda_\sigma$ for fixed $\kappa$. 

Marginal Rate GP Shape and Scale

Spatio-directional GP shape

Spatio-directional GP scale
Spatio-Directional Scale of GP Exceedances

David Randell (Shell)
Spatio-Directional Shape of GP Exceedances

David Randell (Shell)

Modelling covariate effects

SIMRIDE March 2013
Cross Validation for Penalty

Generalised Pareto lack of predictive fit as a function of penalty
The return value $z_T$ of storm peak significant wave height corresponding to some return period $T$, expressed in years, can be evaluated in terms of estimates for model parameters $\phi$, $\rho$, $\xi$ and $\sigma$

$$z_T = \phi - \frac{\sigma}{\xi} \left(1 + \frac{1}{\rho} \left(\log(1 - \frac{1}{T})\right)^{-\xi}\right)$$

- $z_{100}$ corresponds to the 100–year return value, often denoted by $H_{S100}$. 
Marginal 100-year Return Value $H_{S_{100}}$
Spatio-Directional 100-year Return Value $H_{S100}$
Contents

1 Background
  • Motivation
  • Australian North West Shelf

2 Extreme Value Analysis: Challenges

3 Modelling Covariates
  • Model Components
  • P-Splines
  • Quantile regression models threshold
  • Poisson models rate of threshold exceedances
  • GP models size of threshold exceedances
  • Return Values

4 Other Applications and Developments
Other Applications and Developments

- Spatio-directional models for other ocean basins
  - North Sea
  - Gulf of Mexico

- Spatio-temporal splines for non-stationary extreme values
  - Almost all current EVA assumes data are steady state
  - Climate change means this is no longer reliable.
  - Using GCM, RCM as well as historical hindcasts.

- Incorporation of uncertainty
  - Spatial block bootstrapping allows quick estimates of parameter uncertainty
  - Bayesian estimation.

- Incorporation of spatial dependency
  - Composite likelihood: model (asymptotically dependent) componentwise–maxima.
  - Censored likelihood: allows extension from block-maxima to threshold exceedances.


Thank You

david.randell@shell.com