Conditional extremes with covariates

Philip Jonathan, Kevin Ewans, David Randell

June 2011

Objectives

Problem structure:

- Bivariate sample $\{\dot{x}_{ij}\}_{i=1,j=1}^{n,2}$ of random variables $\{\dot{X}_{ij}\}_{i=1,j=1}^{n,2}$
- Covariate $\{\theta_{ij}\}_{i=1,j=1}^{n,2}$ associated with each individual

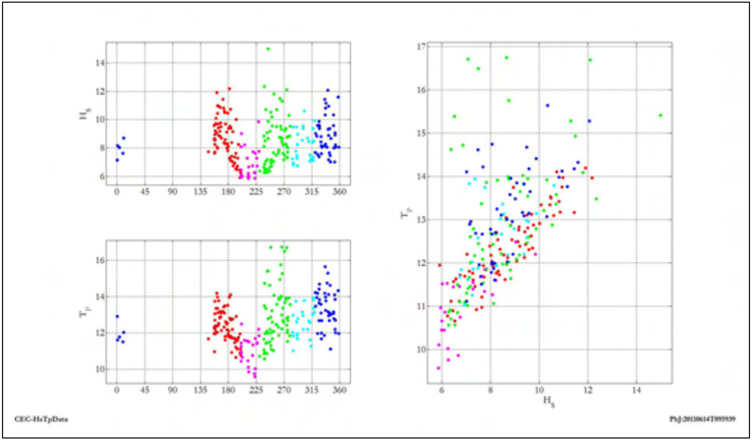
- For some choices of variables $\dot{X}$, e.g. $\dot{X}_1 = H_S$, $\dot{X}_2 = T_P$, $\theta_{i1} \triangleq \theta_{i2} \quad \forall i$
- For other choices, e.g. $\dot{X}_1 = H_S$, $\dot{X}_2 =$WindSpeed, $\theta_{i1} \neq \theta_{i2}$ in general

Objective:

- Objective is to model the **joint distribution** of extremes of $\dot{X}_1$ and $\dot{X}_2$ **as a function of θ**

Location

Exploratory data analysis



- Spread of T_P vs H_S different for different directions

Outline of method

- Follows Heffernan & Tawn (2004)
- Model $\dot{X}_1$ and $\dot{X}_2$ marginally **as a function of θ**
    - **Quantile regression** (QR) below threshold
    - **Generalised Pareto** (GP) above threshold
- Transform to standard Gumbel variates X_1 and X_2
- Model X_2 given large values of X_1 using extension of **conditional extremes model** incorporating **covariate θ**
- **Simulate** for long return periods
    - Generate samples of joint extremes on Gumbel scale
    - Transform to original scale

Marginal model for threshold exceedences

For sufficiently large threshold, $\forall ij$, the $\dot{X}_{ij}$s are marginally independently distributed according to:

$$\mathbf{Pr}(\dot{X}_{ij} > \dot{x}_{ij} | \dot{X}_{ij} > \psi_{ij}(\tau_{j*})) = (1 + \frac{\xi_{ij}}{\zeta_{ij}}(\dot{x}_{ij} - \psi_{ij}(\tau_{j*})))^{-\frac{1}{\xi_{ij}}}$$

where:

- $\psi_{ij}(\tau_{j*}) = \psi_j(\theta_{ij}, \tau_{j*})$ is a quantile threshold associated with cumulative probability τ_{j*}
- $\xi_{ij} = \xi_j(\theta_{ij})$ and $\zeta_{ij} = \zeta_j(\theta_{ij})$
- ψ_j, ξ_j and ζ_j are smooth functions $\forall j$
- Fourier forms estimated by maximising roughness-penalised likelihood

Use diagnostics to select an appropriate threshold level τ_{j*}:

- Q-Q plot
- Stability of $\xi_j(\theta)$ with θ $\forall j$

Unconditional marginal CDF

The unconditional cumulative distribution function for threshold excesses is:

$$\begin{aligned} F_{ij}(\dot{x}_{ij}) &= Pr(\dot{X}_{ij} \leqslant \dot{x}_{ij}) \\ &= 1 - (1-\tau_{j*})(1 + \frac{\xi_{ij}}{\zeta_{ij}}(\dot{x}_{ij} - \psi_{ij}(\tau_{j*})))^{-\frac{1}{\xi_{ij}}} \quad \dot{x}_{ij} > \psi_{ij}(\tau_{j*}) \\ &= \tau_L + (\tau_H - \tau_L)\frac{(\dot{x}_{ij} - \psi_{ij}(\tau_L))}{(\psi_{ij}(\tau_H) - \psi_{ij}(\tau_L))} \quad \dot{x}_{ij} \leqslant \psi_{ij}(\tau_{j*}) \end{aligned}$$

where $\forall j$, $\{\tau_d\}_{d=1}^{D}$ is a set of threshold probabilities for which quantile thresholds $\psi_j(\theta, \tau_d)$ have been estimated, and:

$$H = \arg\min_{d}\{\psi_{ij}(\tau_d) \geqslant \dot{x}_{ij}\}$$

with $L = H - 1$ and $K = j*$

Typically we would have $\{\tau_d\}_{d=1}^{D} = 0.1, 0.2, ..., 0.9$ say, and evaluate quantile regressions for each. We would choice the smallest value for which GP gives good marginal fit, then use quantiles corresponding to smaller values to approximate the CDF

Quantile regression with Fourier parameterisation

- Data $\{\theta_i, y_i\}_{i=1}^n$
- τ^{th} conditional quantile function $Q_y(\tau|\theta) = \psi(\tau, \theta)$, where:

$$\psi(\tau,\theta) = \sum_{k=0}^{p} a_{\tau\psi k}\cos(k\theta) + b_{\tau\psi k}\sin(k\theta) \text{ and } b_{\tau\psi 0} \triangleq 0$$

- Estimated by minimising criterion Q_τ with respect to $\{a_{\tau\psi k}, b_{\tau\psi k}\}_{k=0}^p$:

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

in terms of residuals:

$$r_i = y_i - \psi(\tau, \theta_i) \text{ for } i = 1, 2, ..., n$$

Roughness-penalised quantile regression

Use penalised criterion Q_τ^* instead of Q_τ :

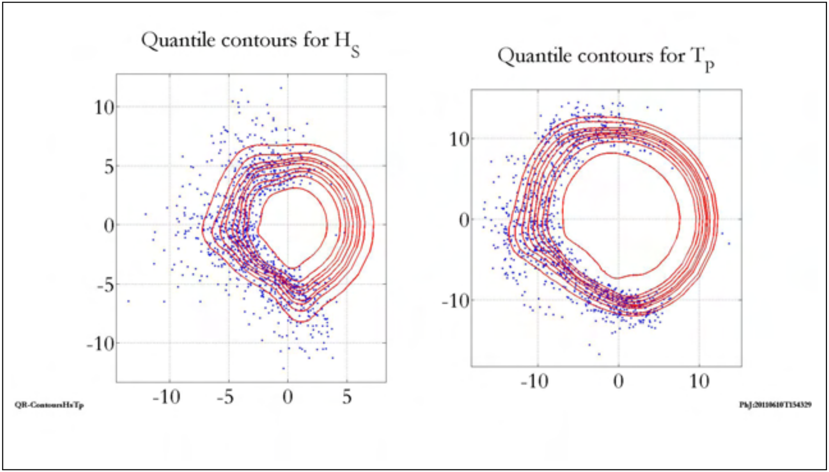
$$Q_\tau^* = Q_\tau + \lambda R_{Q\tau}$$

where parameter roughness R_ψ with respect to x is defined by:

$$\begin{aligned} R_{Q\tau} &= \int_0^{2\pi} (\psi_\tau(x)'')^2 dx \\ &= \sum_{k=0}^{p} k^4 (a_{\tau\psi k}^2 + b_{\tau\psi k}^2) \end{aligned}$$

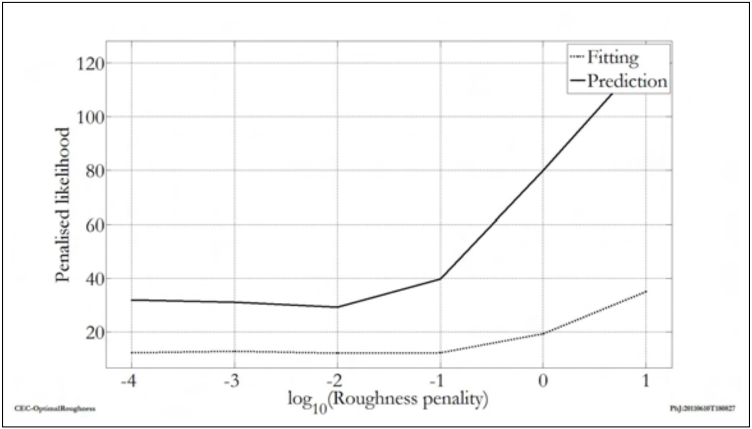
Solved using linear programming

Regression quantiles



- Transform directions to **uniform prior** using QR estimation
- Deciles to 80%

Cross-validatory choice of QR roughness penalty, λ



- Penalty of approximately 0.1 appropriate

Transformation to Gumbel scale

Transform sample $\{\dot{x}_{ij}\}_{i=1,j=1}^{n,2}$ to sample $\{x_{ij}\}_{i=1,j=1}^{n,2}$ on Gumbel scale using probability integral transform:

$$\exp(-\exp(-x_{ij})) = Pr(X_{ij} \leqslant x_{ij}) = Pr(\dot{X}_{ij} \leqslant \dot{x}_{ij}) \text{ from above}$$

Model form

On Gumbel scale, by analogy with Heffernan & Tawn (2004) we propose the following conditional extremes model:

$$(X_k|X_j = x_j, \Phi = \phi) = \alpha_\phi x_j + x_j^{\beta_\phi}(\mu_\phi + \sigma_\phi Z) \textbf{ for } x_j > \psi_j^G(\theta_j, \tau_{j*}^G)$$

where:

- $\psi_j^G(\theta_j, \tau_{j*}^G)$ is a high directional quantile of X_j on Gumbel scale, above which the model fits well
- $\alpha_\phi \in [0,1]$, $\beta_\phi \in (-\infty, 1]$, $\sigma_\phi \in [0, \infty)$
- Z is a random variable with **unknown** distribution G
- Z will be assumed to be approximately Normally distributed for the purposes of parameter estimation

Settings:

- In a (H_S, T_P) case, $\phi \triangleq \theta_j \triangleq \theta_k$, and dependence is assumed a function of absolute covariate
- In a $(H_S, WindSpeed)$ case, $\phi = \theta_k - \theta_j$, and dependence is assumed a function of relative covariate

Fourier parameterisation of conditional model

Defining $\{\eta\}_{r=1}^{4}$ to be $\{\alpha, \beta, \mu, \sigma\}$, we assume Fourier form with ϕ:

$$\eta_r(\phi) \quad = \quad \sum_{s=0}^{p} a_{\eta_r s} \cos(s\phi) + b_{\eta_r s} \sin(s\phi) \text{ and } b_{\eta_r 0} \triangleq 0$$

Parameter roughness R_{η_r} with respect to ϕ is defined by:

$$R_{\eta_r} \quad = \quad \int_0^{2\pi} (\eta_r^{''}(\phi))^2 d\phi = \sum_{s=0}^{p} s^4 (a_{\eta_r s}^2 + b_{\eta_r s}^2)$$

Total solution roughness $R_\eta(\underline{\omega})$ (for $\underline{\omega}$ s.t. $\sum_{r=0}^{4} \omega_r = 1$ in general):

$$R_\eta = R(\underline{\omega}) \quad = \quad \sum_{r=0}^{4} \omega_r R_{\eta_r}$$

Since it is reasonable to expect that $\alpha_\phi \in [0, 1]$, $\beta_\phi \in [0, 1)$, $\mu_\phi \in [-\frac{1}{2}, \frac{1}{2}]$ (residual mean should be near zero) and $\sigma_\phi \in (0, 1]$ (σ_ϕ just relative scale, absolute scale given by $\mathrm{Var} Z$), we set $\omega_1 = \omega_2 = \omega_3 = \omega_4 = \frac{1}{4}$ for simplicity. We therefore have only one overall roughness tuning parameter.

Penalised likelihood

For sample $\{x_{ik}, x_{ij}, \phi_i\}_{i=1}^{m}$ corresponding to threshold exceedences $\{x_{ij}\}_{i=1}^{m}$ of ψ_j^G, negative log likelihood ℓ is given by:

$$\ell = \sum_{i=0}^{n} \log s_i + \frac{(x_{ik} - m_i)^2}{2s_i^2}$$

where:

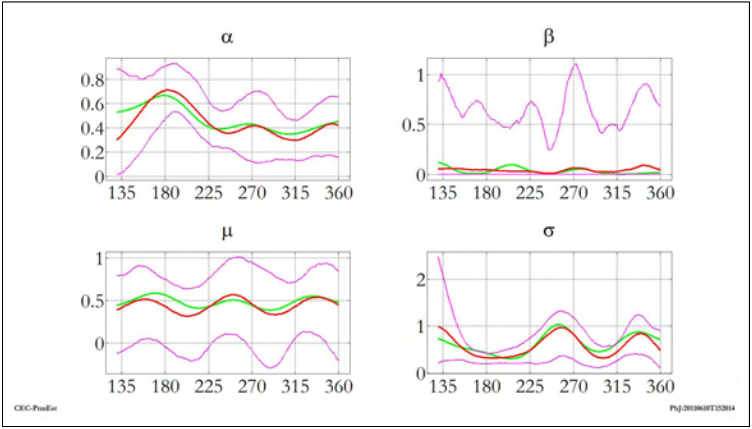
$$m_i = m_i(x_{ij}, \phi_i) = \alpha(\phi_i)x_{ij} + \mu(\phi_i)x_i^{\beta(\phi_i)}$$
$$s_i = s_i(x_{ij}, \phi_i) = \sigma(\phi_i)x_{ij}^{\beta(\phi_i)}$$

Penalised negative log likelihood ℓ^* is given by

$$\ell^* = \ell + \lambda R_\eta$$

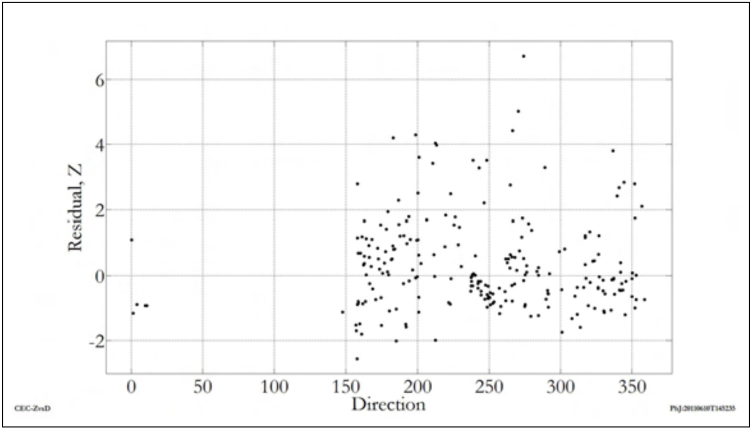
Imposing non-negativity: We choose to express $\sqrt{\alpha}$, $\sqrt{\beta}$ and $\sqrt{\sigma}$ as Fourier series so that their squares are non-negative. Roughness penalty estimated using cross-validation.

Parameter estimates

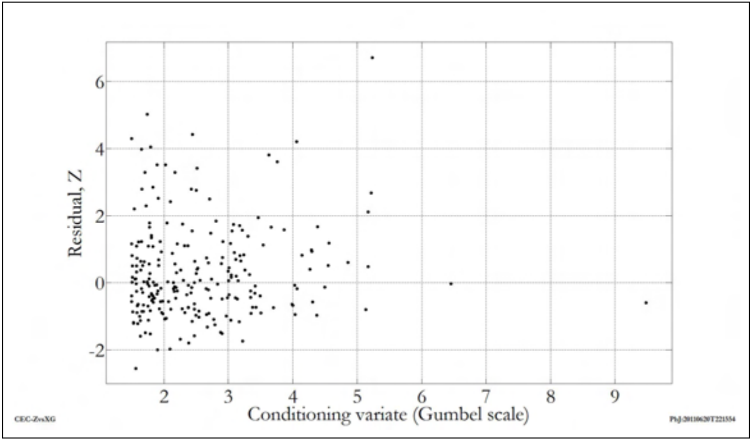


- MLE in green; 1000 bootstrap resamples (median in red, 95% band in magenta)

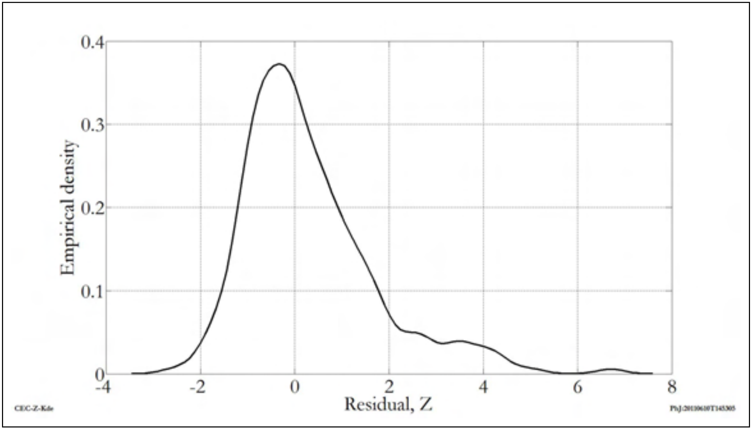
Residuals with direction



Residuals with conditioning variate



Kernel density estimate for residuals



Limit assumption

The limit assumption required to justify the conditional model is:

$$\mathbf{Pr}\left(\frac{x_j^{-\beta_\phi}(X_k - \alpha_\phi x_j) - \mu_\phi}{\sigma_\phi} \leqslant z | X_j = x_j, \Phi = \phi\right) \to G(z) \textbf{ as } x_j \to \infty$$

Estimating T_P associated with extreme quantile of H_S

Given parameter estimates and sample of residuals:

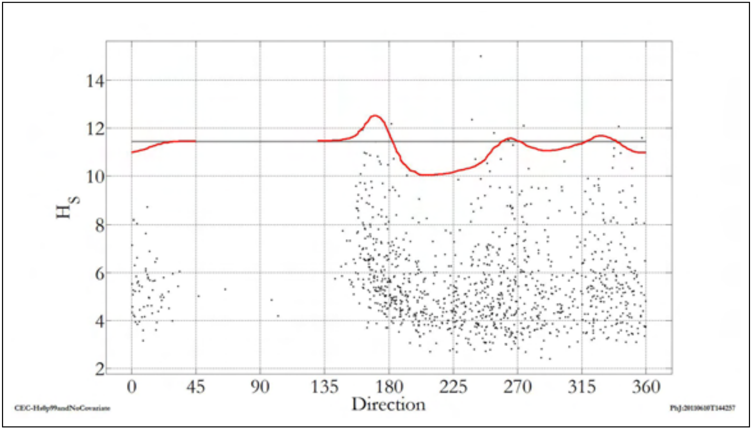
- Estimate quantiles of T_P given any quantile of H_S on Gumbel scale

$$(T_P|H_S = h, \Theta = \theta) = \widehat{\alpha}_\theta h + h^{\widehat{\beta}_\theta}(\widehat{\mu}_\theta + \widehat{\sigma}_\theta Z) \textbf{ for } h > \psi^G(\theta, \tau^G_{j*})$$

- Transform to original scale

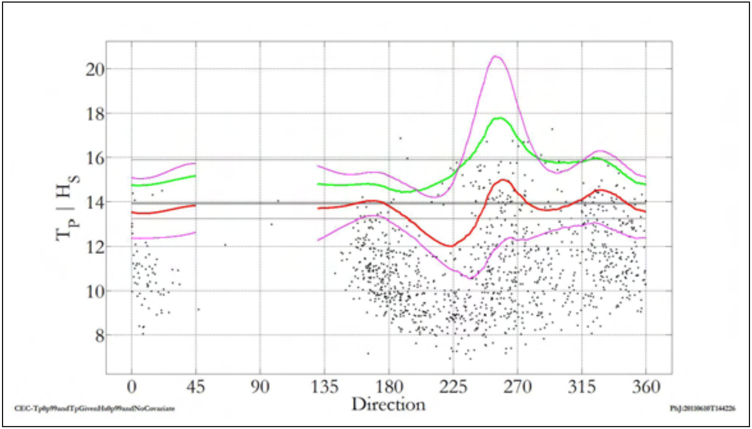
Compare with model ignoring covariate effects

Conditioning variate H_S with tail probability $= 0.01$



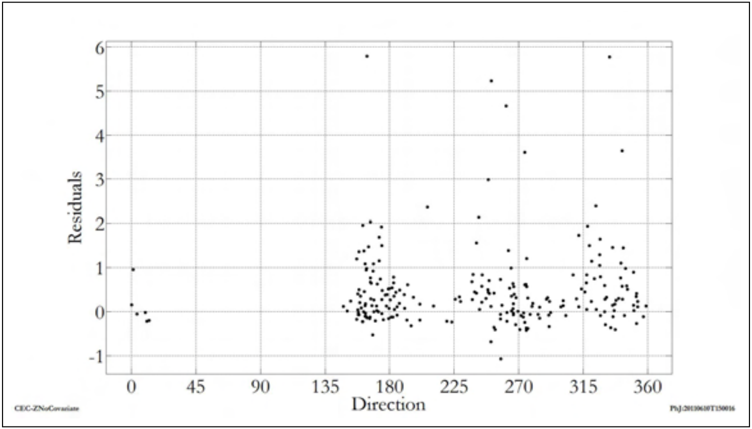
- Exceedence probability $= 0.01$ with covariate (red) and without (grey)

Conditional T_P corresponding to H_S with tail probability $= 0.01$



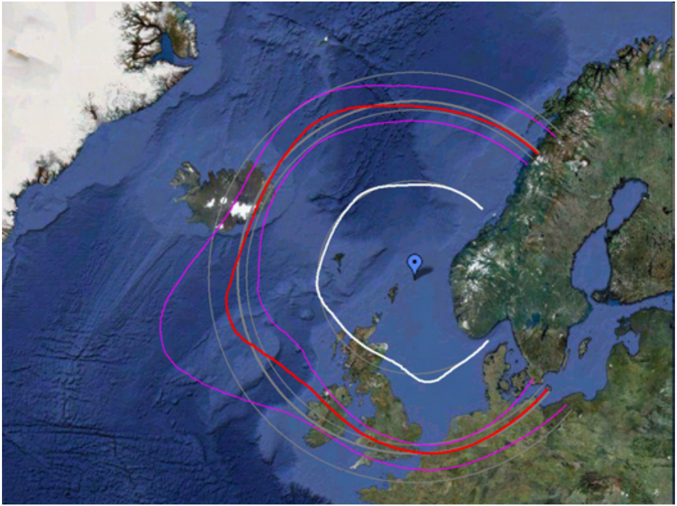
- With covariate (median (red), 95% band (magenta)), without (grey)
- T_P with exceedence probability $= 0.01$ shown in green

Residuals ignoring covariates



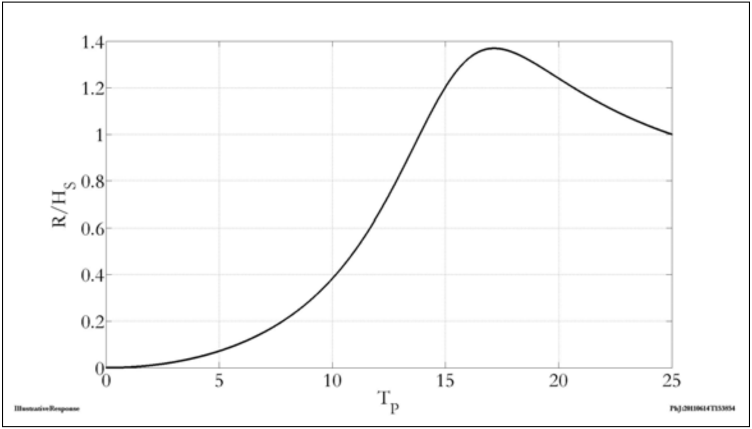
- Directional variation clear

Conditional T_P corresponding to H_S with tail probability $= 0.01$



- With covariate (median (red), 95% band (magenta)), without (grey)
- H_S with exceedence probability $= 0.01$ shown in white

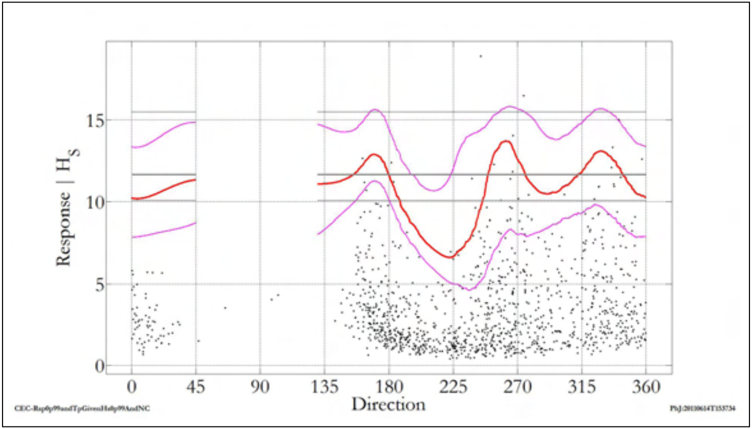
Illustrative response transfer function



- Characteristic of roll or heave response of floating structure

$$\frac{R}{H_S} = \frac{1}{\sqrt{(1-\omega^2) + (k\omega)^2}}, \quad \omega = \frac{2\pi}{T_P}$$

Conditional extreme response: with direction



- Response with covariate effect (median in red, 95% limits in magenta) and without (grey) for H_S with tail probability $= 0.01$

Simulation under the model

The procedure for simulating from the conditional extremes model with covariates is as follows:

- Sample a value θ_{sj}
- Sample a value ϕ_s
- Sample a value x_{sj} of X_j from its Gumbel distribution

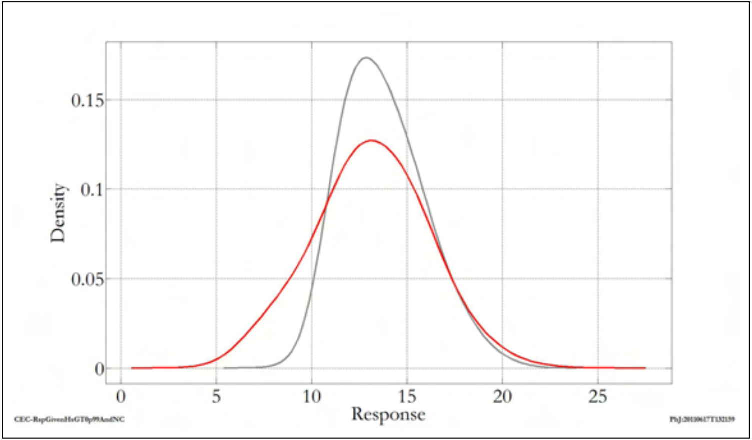
  If $x_{sj} > \psi_j^G(\theta_{sj} : \tau_{j*}^G)$:
  - Sample x_{sk} from the estimated conditional model

  Else:
  - Sample a pair of values $\{x_{sk}, x_{sj}\}$ from the subset of the original sample (of non-exceedences of ψ_j^G)
- Transform from Gumbel to original scale

  If $x_{sj} > \psi_j^G(\theta_{sj}, \tau_{j*}^G)$:
  - Apply probability integral transform

  Else:
  - Find pair $\{\dot{x}_{sk}, \dot{x}_{sj}\}$ corresponding to $\{x_{sk}, x_{sj}\}$ in original data

Conditional extreme response: kernel density estimate



- Response density with covariate effect (red) and without (grey) for exceedences of H_S with tail probability $= 0.01$

Conclusions and references

Conclusions

- Extension of conditional extremes model to include covariate effects
- Requires approach to marginal estimation with covariate (QR used here)
- Makes engineering application of conditional extremes model feasible, particularly for floating structures

References

- 2004: Heffernan, JE and Tawn, JA: A conditional approach for multivariate extreme values,**J. R. Statist. Soc. B** , v66, p497.
- 2005: Koenker, R: **Quantile regression**, Cambridge University Press.

Thanks for listening

philip.jonathan@shell.com