Joint modelling of wave spectral parameters for extreme sea states

Philip Jonathan, Jan Flynn & Kevin Ewans
Shell Technology Centre Thornton, UK
Shell International Exploration and Production, NL

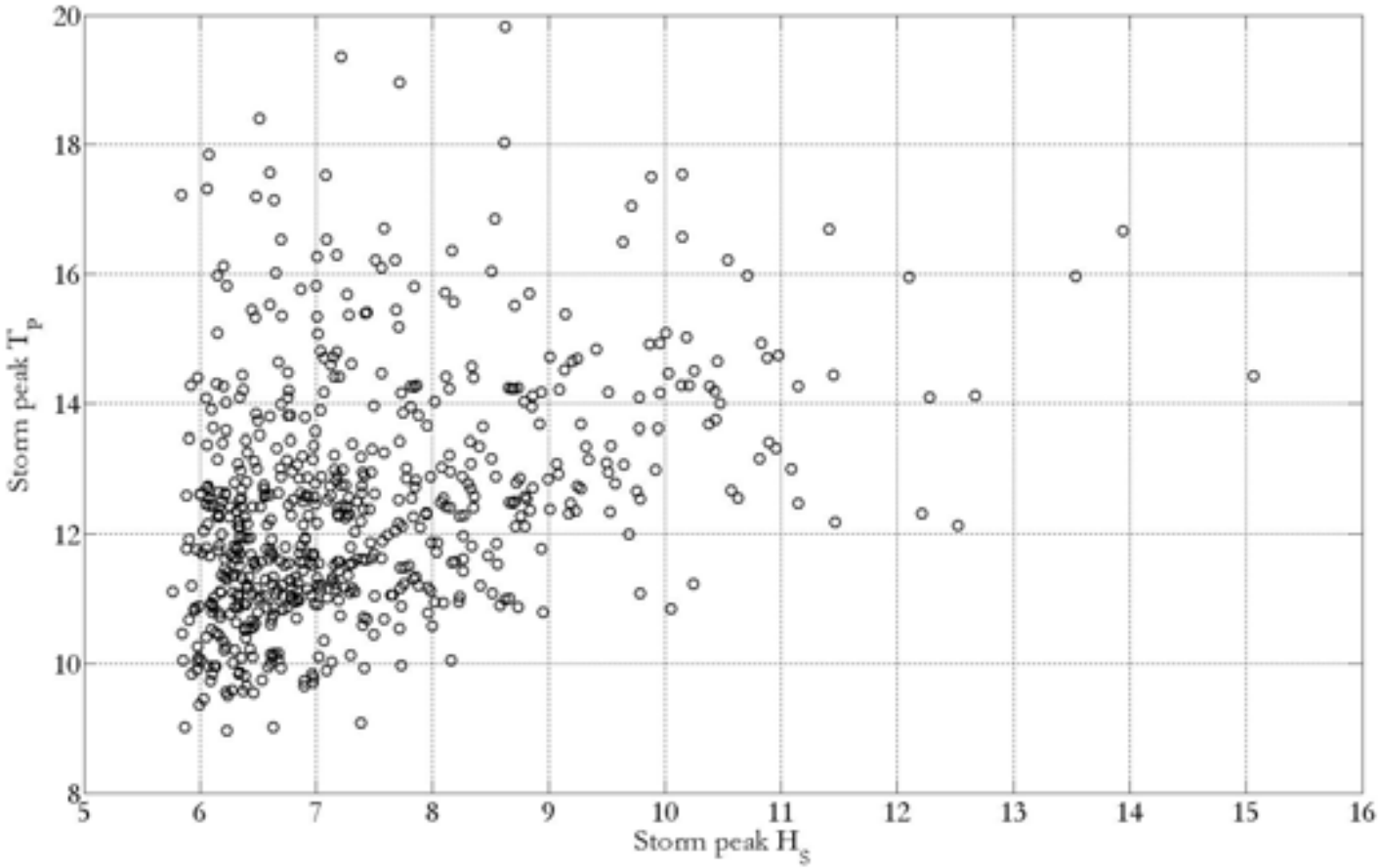
Hindcast-Forecast Workshop
Halifax
19 October 2009

Motivation

- Careful statistical description central to understanding extreme ocean environments
- Need to understand joint extremes behaviour
- Structure variable (e.g. "response-based") analysis is one possibility
- Direct joint modelling using conditional approach offers another option

- Heffernan and Tawn [2004] is the key reference
- Similarity to work of Haver [1985]

Typical challenge

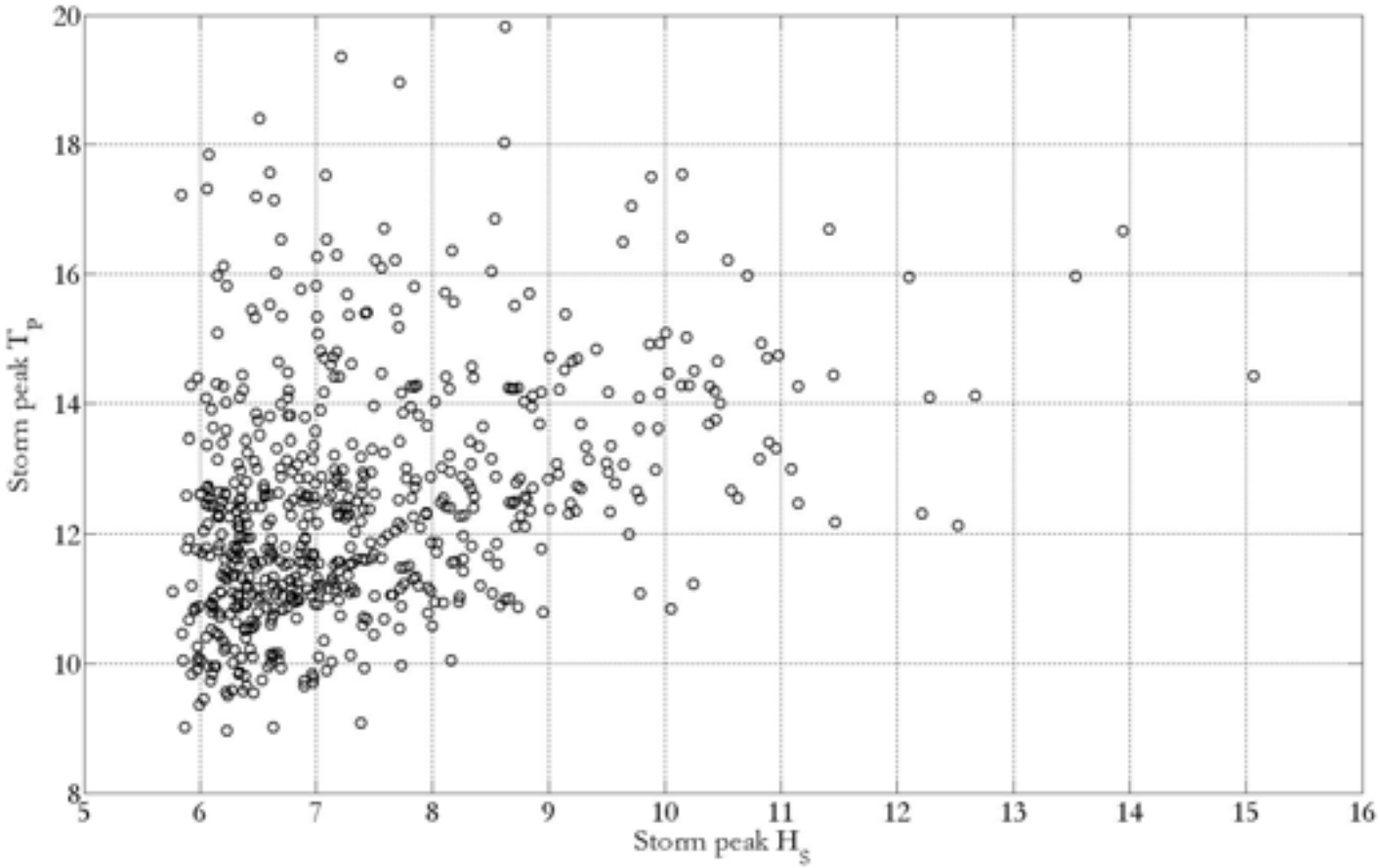


- What is the probability of $H_S^{sp} > 17m$ and $T_P^{sp} < 14s$?
- Estimate a 95% uncertainty interval for T_P^{sp} for H_S^{sp} at its 100-year level

Current work

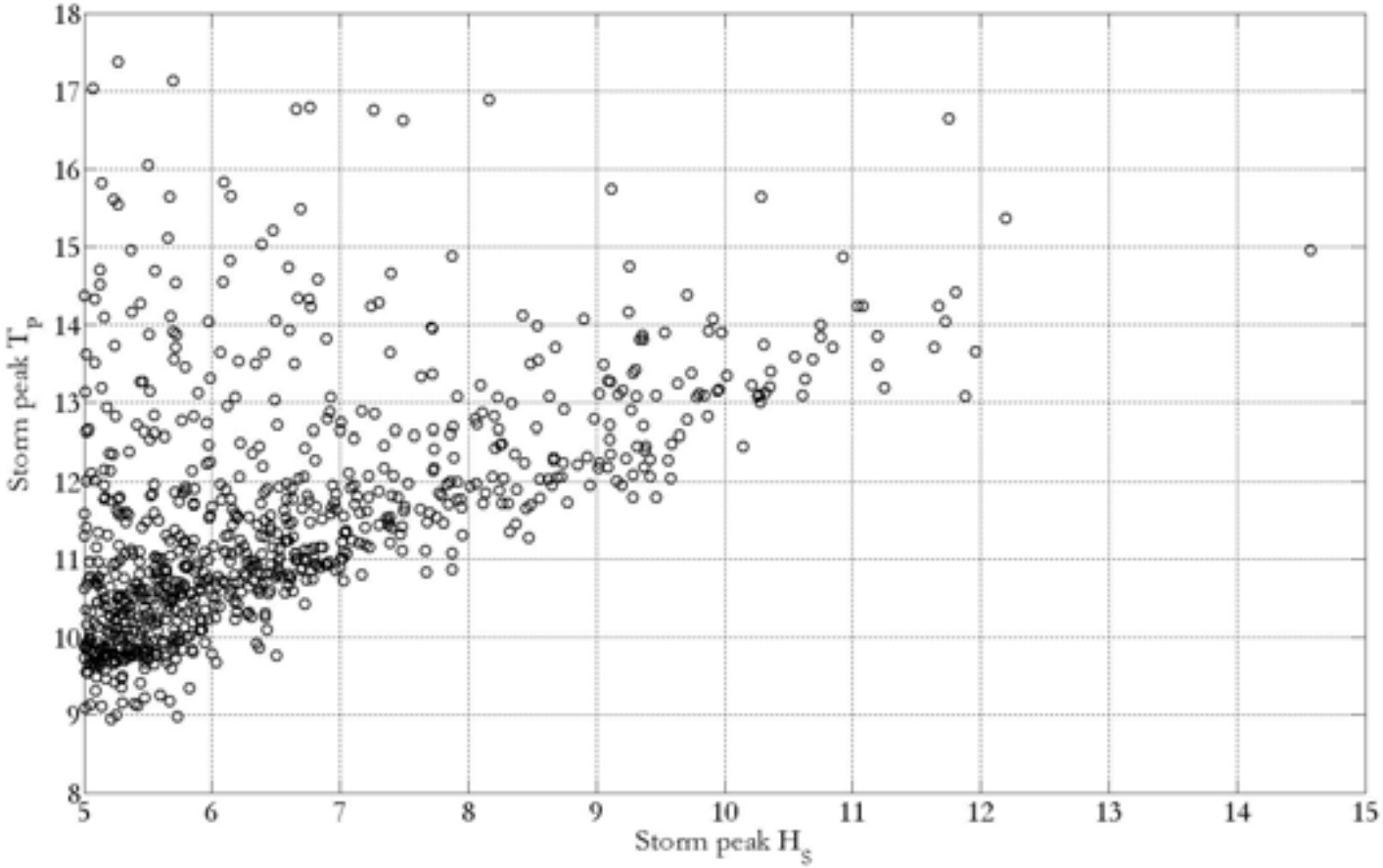
- Conditional model used for H_S and T_P
- Evaluation of conditional approach for distributions with difference extremal dependence structures
- Comparison with Haver
- Four applications considered

Measured Northern North Sea



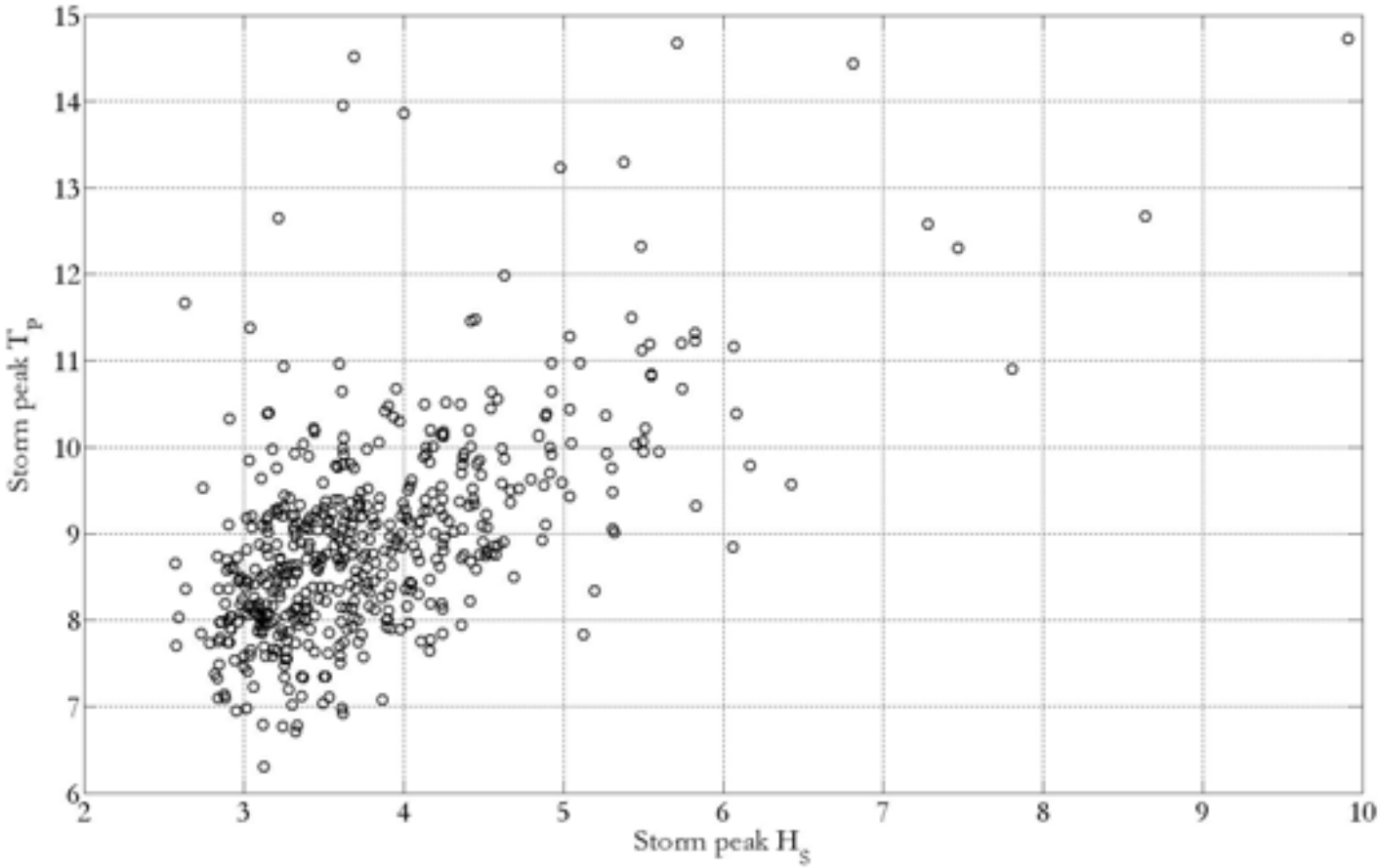
- 620 storm peak pairs for (March 1973, December 2006), laser-measured
- Same location as HndNNS

Hindcast Northern North Sea



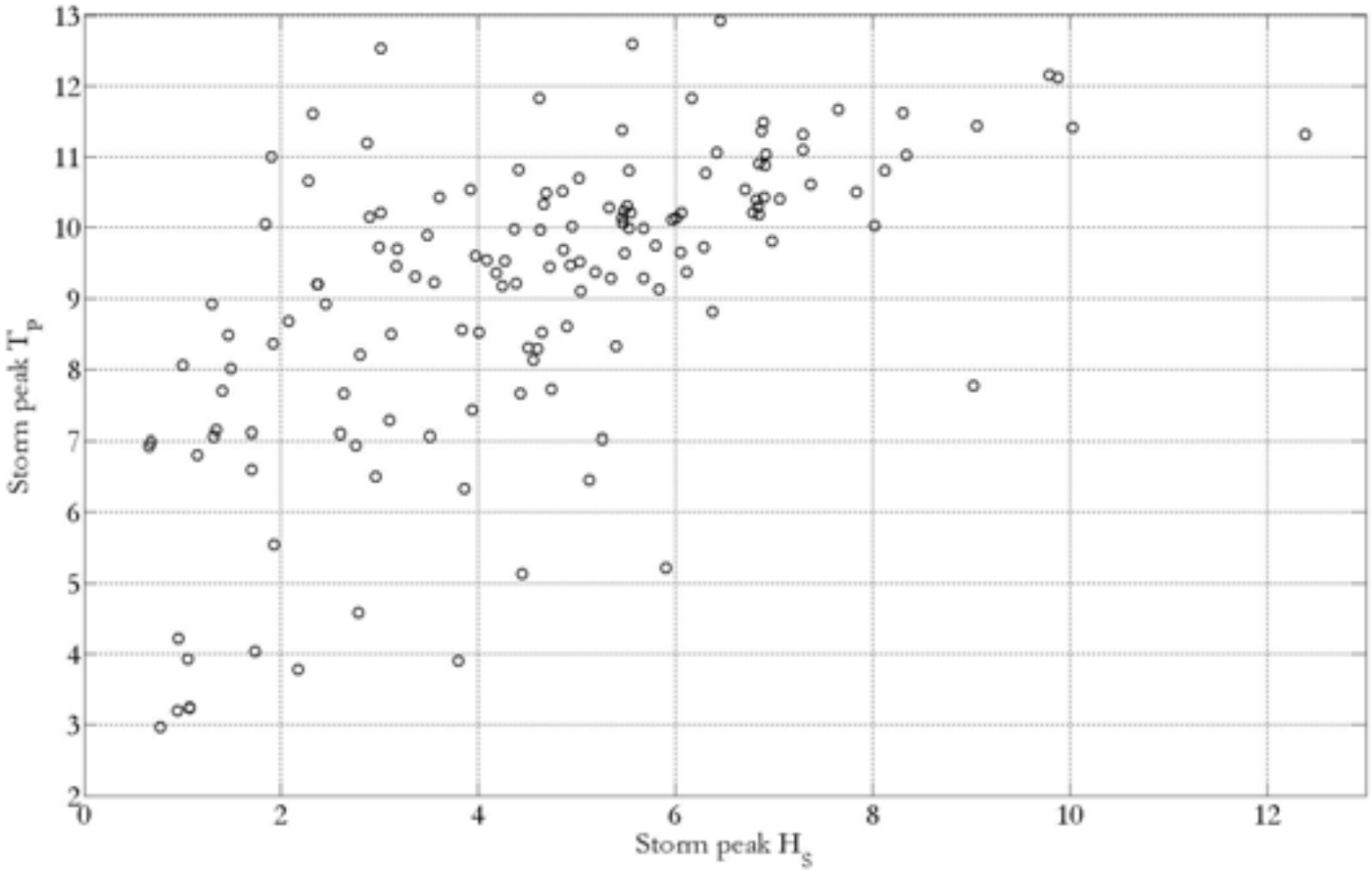
- 827 storm peak pairs for (November 1964, April 1998)
- Same location as MsrNNS

Measured Gulf of Mexico



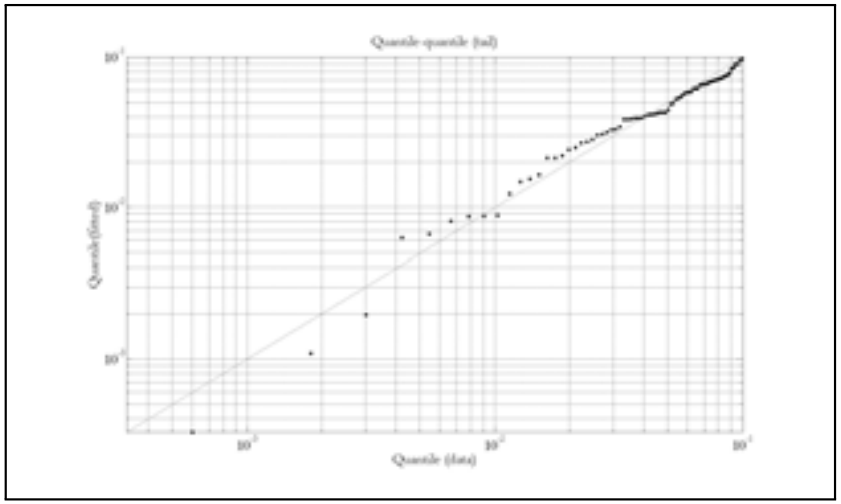
- National Data Buoy Center measurements from buoy 42002
- 505 pairs for (January 1980, December 2007)

Hindcast North West Shelf

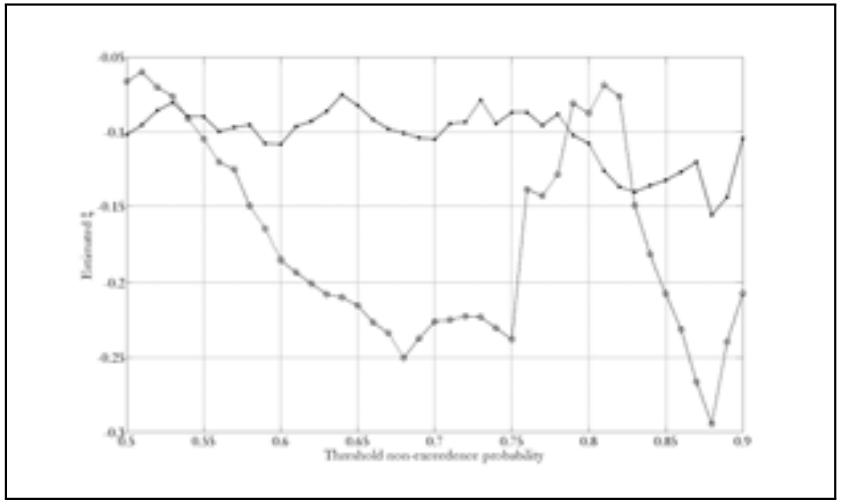


- 145 storm peak pairs for (February 1970, April 2006)

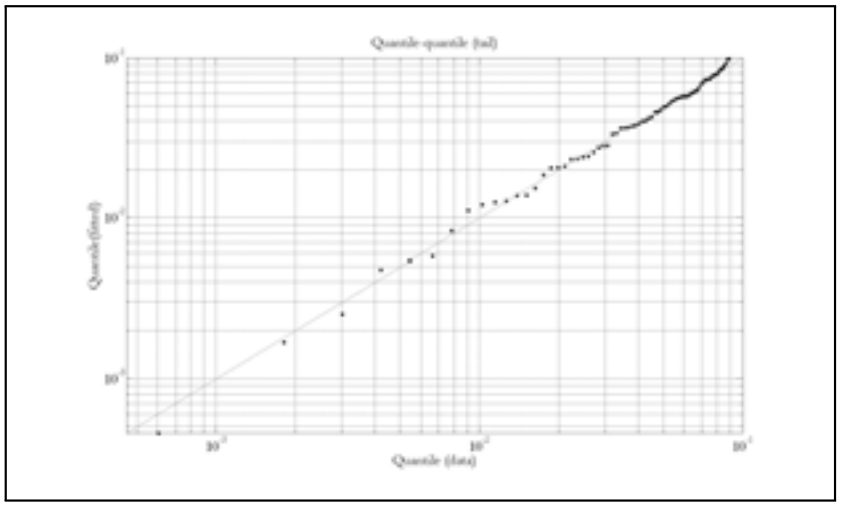
Generalised Pareto tails



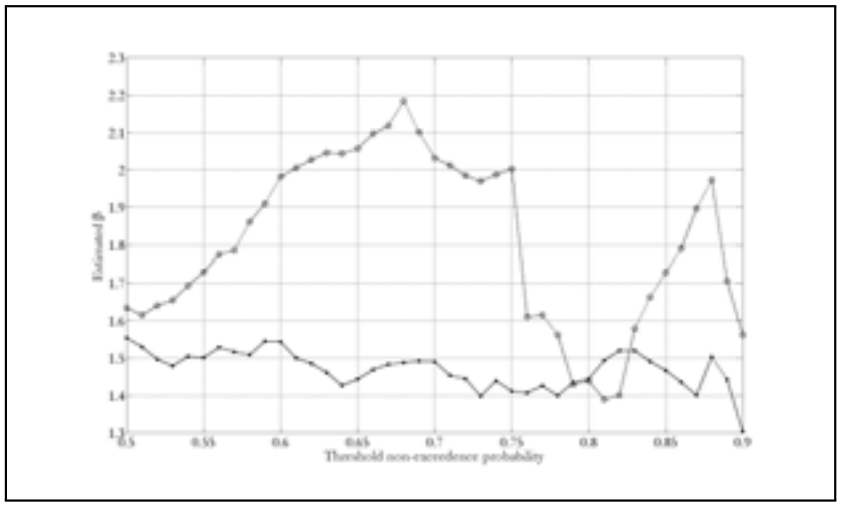
QQ: MsrNNS H_S



MsrNNS GP shape ξ with threshold u for H_S (dots) and T_P



QQ: MrsNNS T_P



MsrNNS GP scale β with threshold u for H_S (dots) and T_P

Basic idea

- For pairs (X, Y), both marginally standard Gumbel-distributed and positively-associated
- $(Y|X = x) = ax + x^b Z$ for x above a sufficiently large threshold u
- (a, b) are location and scale parameters to be estimated
  - $a \in [0, 1]$ and $b \in (-\infty, 1]$
- Z is a standardised variable, independent of X, converging with increasing x to a non-degenerate limiting distribution G
- The form of distribution G is not specified by theory (so parameters of Z treated as nuisance parameters in model)

- Strong theoretical motivation
- Easily implemented and expanded to large problems (applications with 100s of variables in literature)
- Extremal dependence characterised by a, b and G

Procedure in a nut shell

- Transform original variables (X^*, Y^*) to (X, Y) having standard Gumbel marginals (using probability integral transform)
- Estimate parameters of conditional models $(Y|X = x)$ and $(X|Y = y)$ for large thresholds x and y
- Simulate using fitted model to characterise extremal behaviour of the joint distribution
  - Simulate arbitrarily long return periods from conditional models
  - Transform to original scale

Fit Generalised Pareto model to tails

- GP form appropriate to model peaks over threshold u^*
- Fit the GP distribution (e.g. to sample $\{x_i^*\}_{i=1}^n$ of X^* here)
- $F_{GP}(x^*; \xi, \beta, u^*) = 1 - (1 + \frac{\xi}{\beta}(x^* - u^*))_+^{-\frac{1}{\xi}}$
- Diagnostic: ξ versus (varying) u^* should be constant for GP data
- Quantify uncertainty due to threshold selection

Transform from GP to Gumbel marginals (and back)

- Estimate cumulative probabilities $\{F_{GP}(x_i^*; \hat{\xi}, \hat{\beta}, u)\}_{i=1}^n$ from GP fit
- Standard Gumbel has cumulative distribution function:
  - $F_{Gmb}(x) = \exp(-\exp(-x))$
- Define transformed sample $\{x_i\}_{i=1}^n$ such that:
  - $F_{Gmb}(x_i) = F_{GP}(x_i^*; \hat{\xi}, \hat{\beta}, u)$ for $i = 1, 2, 3, ..., n$
- Similarly, given a value of x (of Gumbel variate), we can calculate the corresponding value x^* on the original GP scale

Fit conditional model

- $(Y|X = x) = ax + x^b Z$ for sufficiently large x
- $\hat{a}$, $\hat{b}$ and sample from G are estimated using regression
- For simplicity and computational ease during model fitting only, G is assumed to be Gaussian
  - mean μ_Z and variance σ_Z^2 treated as nuisance parameters
- Fitted values:
  - $\hat{z}_i = \frac{(y_i - \hat{a}x_i)}{x_i^{\hat{b}}}$, $i = 1, 2, 3, ..., n$
  - $\{\hat{z}_i\}_{i=1}^n$ provide an estimate of a sample from distribution G for subsequent simulations
- Diagnostic: for good model fit:
  - $\{\hat{z}_i\}_{i=1}^n$ and $\{x_i\}_{i=1}^n$ are not obviously dependent
  - Varying u should not overly affect values $\hat{a}$, $\hat{b}$ and subsequent estimates
- Bootstrap resampling to estimate uncertainty of estimates $\hat{a}$, $\hat{b}$, etc.

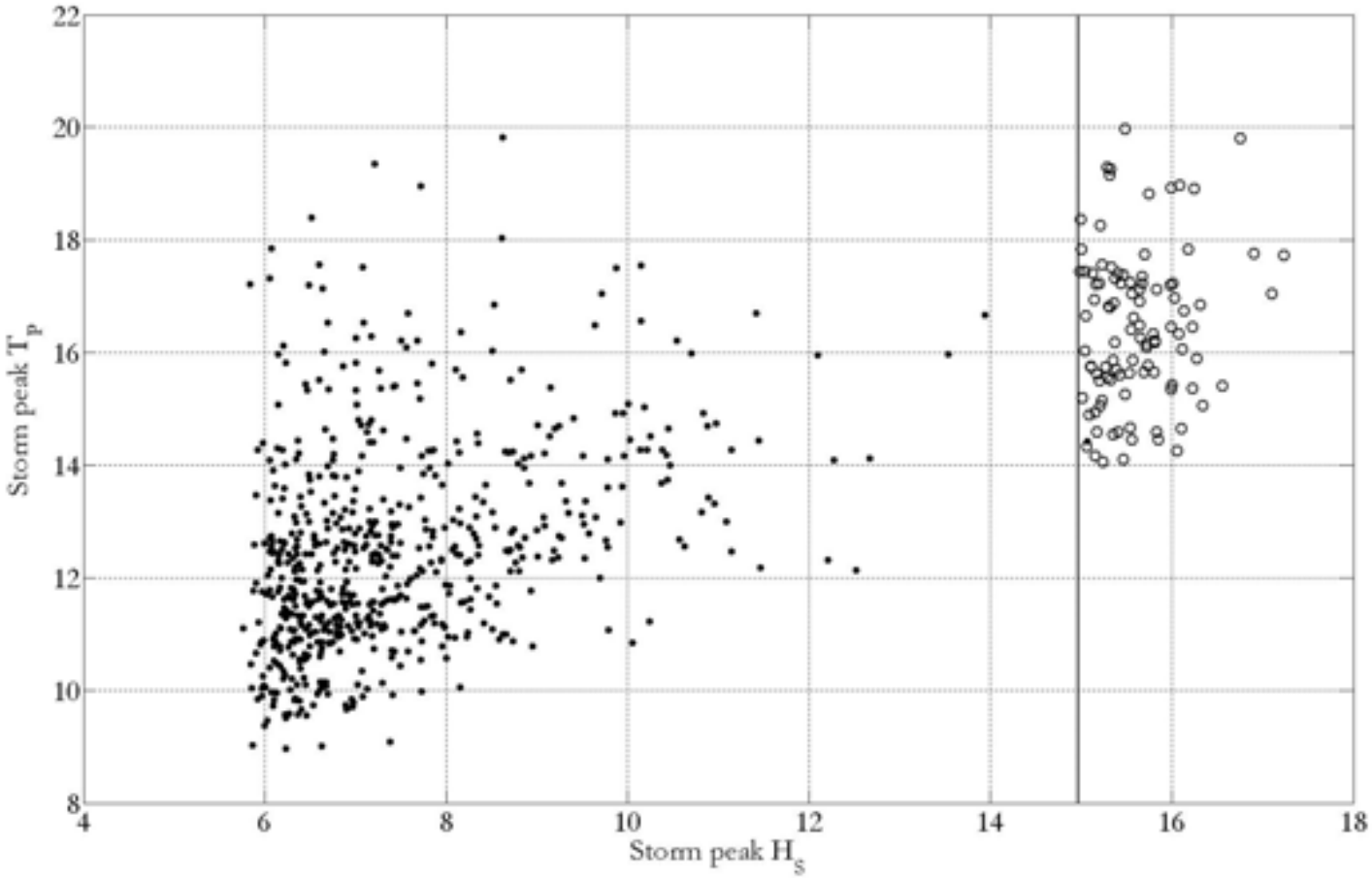
Simulate long return periods

To simulate a random drawing from conditional distribution $Y|X > u$:

- Draw value x of X at random from standard Gumbel distribution, given that the value exceeds threshold u
- Draw value z of Z at random from $\{\hat{z}_i\}_{i=1}^{n}$
- Calculate $y|x = \hat{a}x + x^{\hat{b}}z$
- Transform (x, y) to (x^*, y^*) using probability integral transform and marginal GP models

Using simulation, estimates for various extremal statistics (e.g. values associated with long return periods) can be obtained

Illustrative simulation

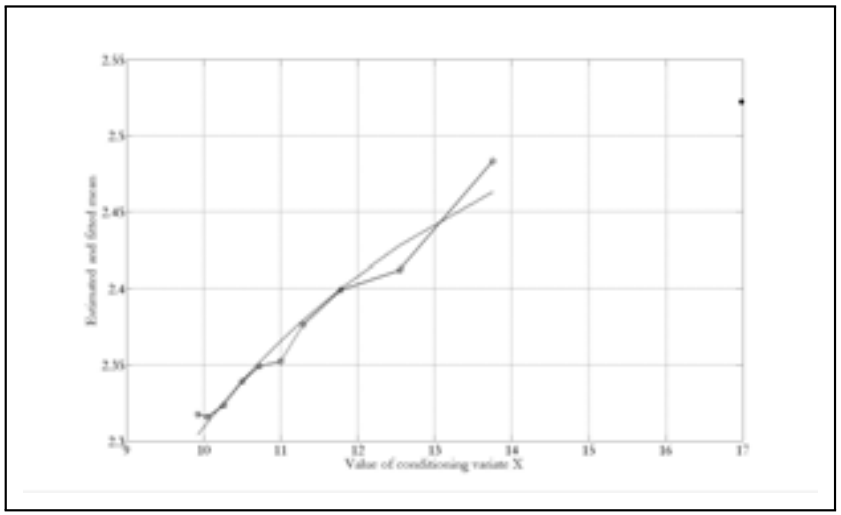


Illustrative simulation from conditional model for measured NNS application, for exceedences of H_{S10} (shown as a vertical line at 15.9m)

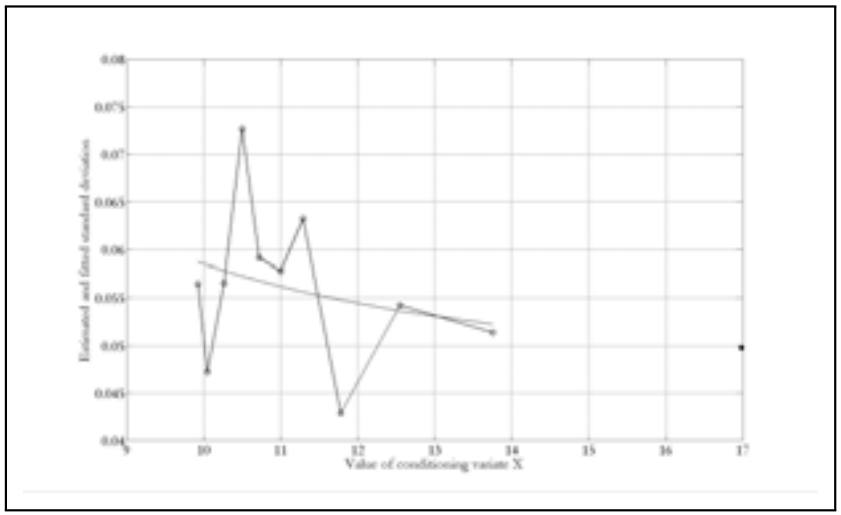
Outline of Haver model

- Developed explicitly for joint modelling of $X(H_S)$ and $Y(T_P)$ for large values of X
- Assumes that $Y|X = x$ follows a log-normal distribution:
  - $f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \frac{1}{x} \exp(-\frac{(\log x - \mu(x))^2}{2\sigma^2(x)})$
  - Estimated parameters $\mu(x)$ and $\sigma(x)$ then regressed on x
  - Forms $\mu = \zeta_1(x+1)^{-2} + \zeta_2 \log_e(x+1) + \zeta_3$ and $\sigma = \zeta_4 x^{\zeta_5}$ recommended based on fit performance
  - Empirical extrapolations to large x (beyond sample)
- Assumes Weibull model for X over a suitably large threshold
  - We consider a modified approach in which the marginal model for X is assumed GP

Extrapolating μ and σ in Haver model



$\mu(x)$



$\sigma(x)$

- Justification for functional forms of $\mu(x)$ and $\sigma(x)$ is empirical
- Haver model yields more biased and variable estimates of extreme quantiles (but could be tuned)
- Idea: constrain Haver model (e.g. moment-based) to improve asymptotic performance

Evaluation of conditional and Haver models

Four distributions used for simulations:

- Distribution D1: Multivariate extreme value distribution with exchangeable logistic dependence
- Distributions D2 and D3: Multivariate distribution with Normal dependence transformed marginally to standard Gumbel
- Distribution D4: Asymptotic conditional form

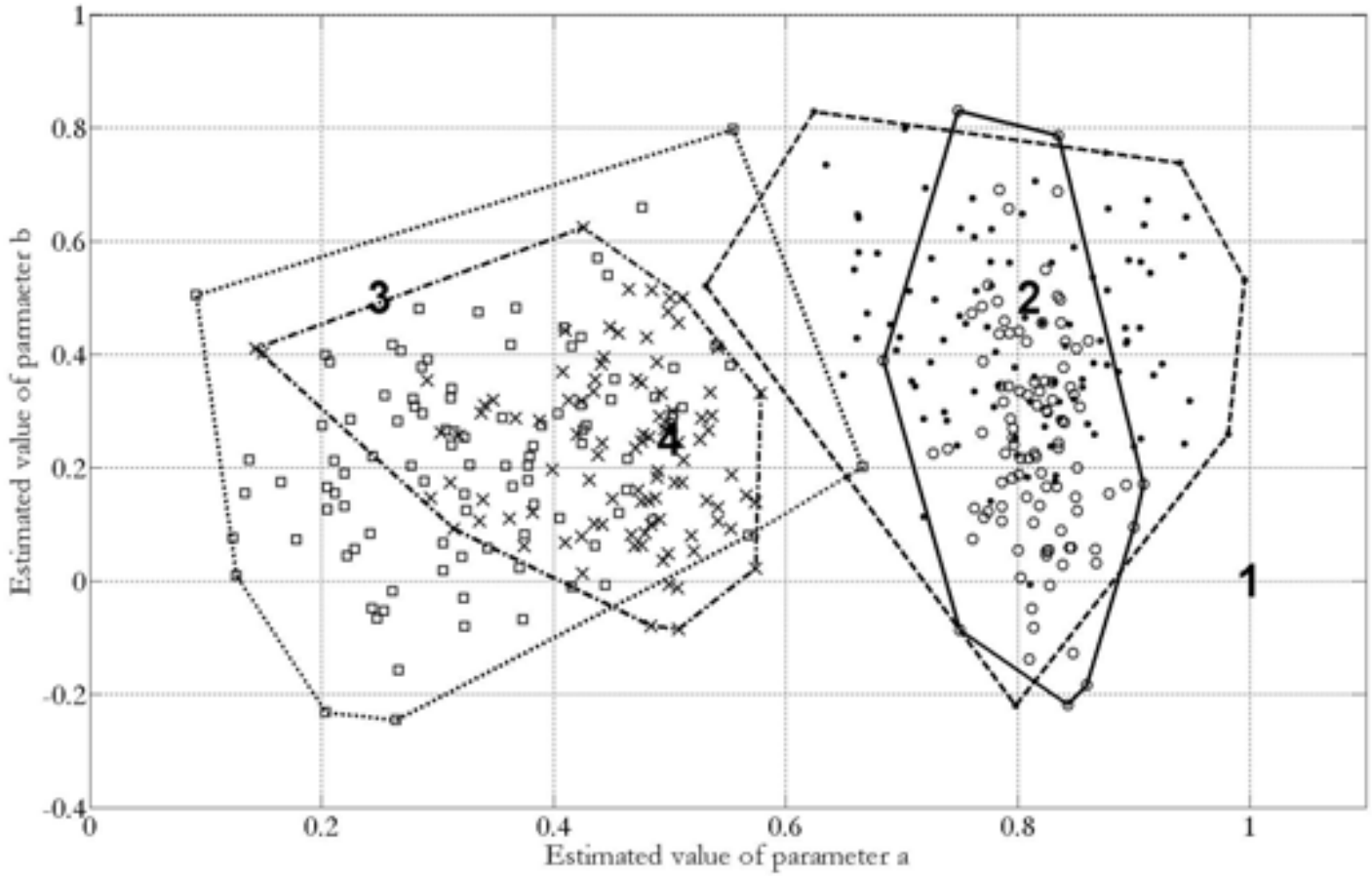
- Distributions D1-D4 intended to explore different extremal dependence structures
- Sample size 1000 used

Dependence structures

- Conditional dependence:
  - $(Y|X = x) = f(x)$
- Asymptotic dependence:
  - $Pr((X, Y) \in A + d) = e^{-d} Pr((X, Y) \in A)$
  - All conventional multiple extreme value methods start here
- Asymptotic independence:
  - $Pr(X, Y) \in A + d) = e^{-\frac{d}{\eta}} Pr((X, Y) \in A)$, $\eta \in (0, 1)$
- Asymptotic conditional dependence:
  - $(Y|X = x) = f(x)$ for large x

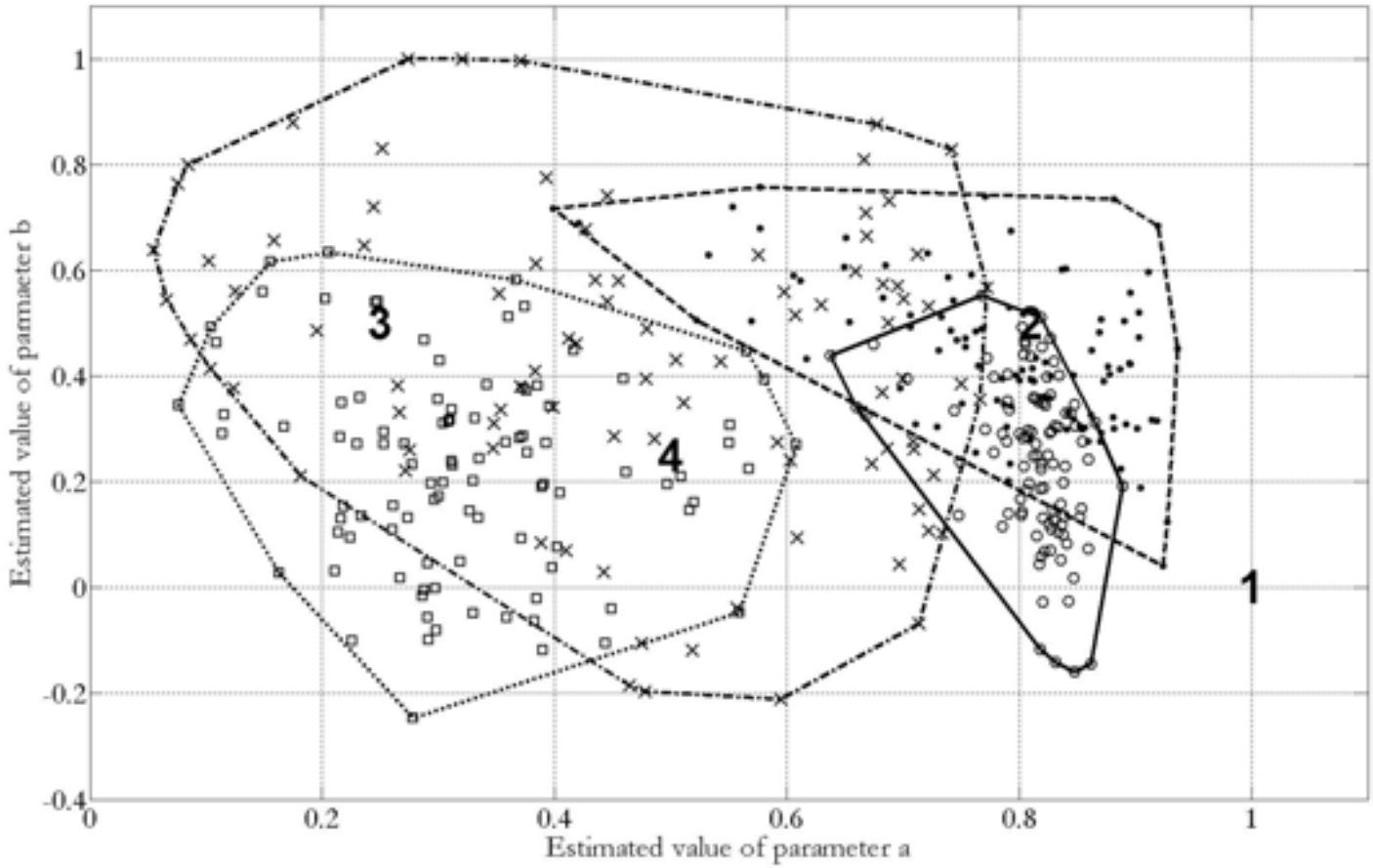
- Conditional model accommodates above

Evaluation: multiple samples, known marginals



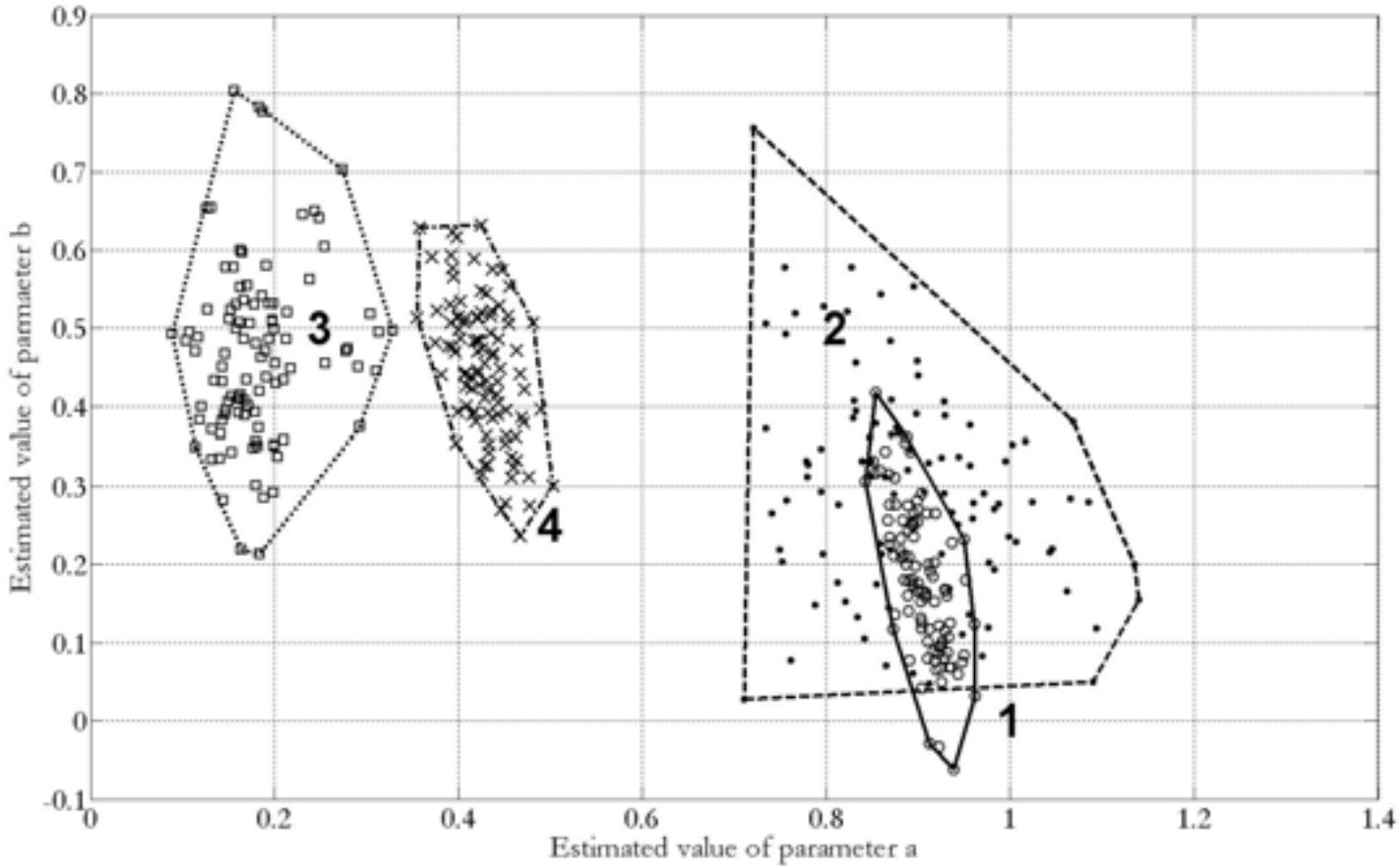
(Multiple sample realisations) Variation in parameter estimates $\hat{a}$ and $\hat{b}$ during simulation, assuming marginals known. Convex hulls enclose pairs corresponding to distribution D1 (circles), D2(dots), D3(squares) and D4(crosses). Location of true values indicated using numbers 1-4

Evaluation: multiple samples, estimated marginals



(Multiple sample realisations) Variation in parameter estimates $\hat{a}$ and $\hat{b}$ during simulation, with marginal estimation. Convex hulls enclose pairs corresponding to distribution D1 (circles), D2(dots), D3(squares) and D4(crosses). Location of true values indicated using numbers 1-4

Bootstrap interval estimates for single sample



(Single sample realisations) Bootstrap convex hulls for conditional parameters *a* and *b* corresponding to typical realisations from distributions D1-D4. Convex hulls enclose pairs corresponding to distribution D1 (circles), D2(dots), D3(squares) and D4(crosses). Location of true values indicated using numbers 1-4

Performance evaluation

Bias and uncertainty for extreme quantiles and model parameters

(a) Conditional model

Data \ Bias	q(0.25)	q(0.50)	q(0.75)
D1: EVEL	-0.33 (-0.39,-0.24)	-0.23 (-0.27,-0.18)	-0.15 (-0.19,-0.10)
D2: TNrm(ρ=0.9)	0.04 (-0.06, 0.09)	-0.00 (-0.06, 0.07)	-0.01 (-0.08, 0.05)
D3: TNrm(ρ=0.5)	0.18 (-0.08, 0.47)	0.13 (-0.18, 0.37)	0.04 (-0.21, 0.24)
D4: Conditional	-0.09 (-0.30, 0.03)	-0.09 (-0.22, 0.03)	-0.08 (-0.18, 0.04)

(b) Haver model

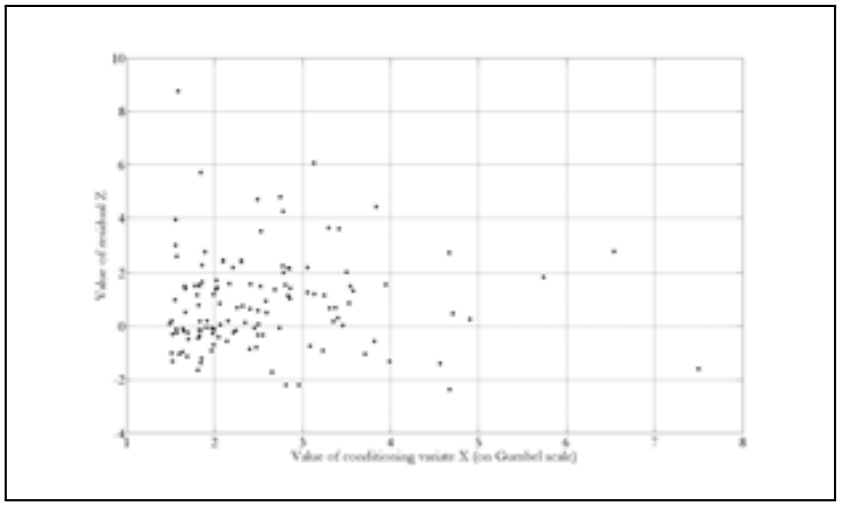
Data \ Bias	q(0.25)	q(0.50)	q(0.75)
D1: EVEL	-1.27 (-1.40,-1.11)	-0.97 (-1.04,-0.85)	-0.60 (-0.67,-0.52)
D2: TNrm(ρ=0.9)	-0.84 (-0.95,-0.74)	-0.71 (-0.79,-0.63)	-0.53 (-0.62,-0.45)
D3: TNrm(ρ=0.5)	-0.43 (-0.54,-0.34)	-0.52 (-0.63,-0.39)	-0.52 (-0.69,-0.38)
D4: Conditional	-1.00 (-1.15,-0.85)	-0.92 (-1.07,-0.79)	-0.73 (-0.88,-0.61)

(c) Conditional parameter estimates

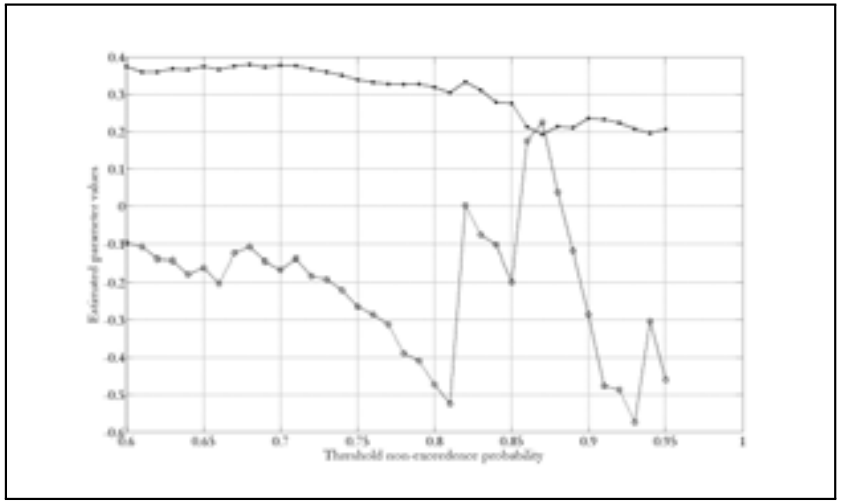
Data \ Bias of estimate	a	b
D1: EVEL (a=1.00,b=0.00)	-0.18 (-0.21,-0.16)	0.24 (0.12, 0.38)
D2: TNrm(ρ=0.9) (a=0.81,b=0.50)	-0.00 (-0.07, 0.06)	-0.07 (-0.18, 0.06)
D3: TNrm(ρ=0.5) (a=0.25,b=0.50)	0.07 (0.00, 0.17)	-0.29 (-0.42,-0.18)
D4: Conditional (a=0.50,b=0.25)	-0.02 (-0.08, 0.00)	-0.01 (-0.14, 0.08)

Estimation of extreme quantile when marginal models assumed known: bias in estimates for Y_{10} from (a) the conditional model and (b) the Haver model; (c) estimated bias of parameter estimates for conditional model; true values for parameters given with data description in first column).

Application: Validating model assumptions

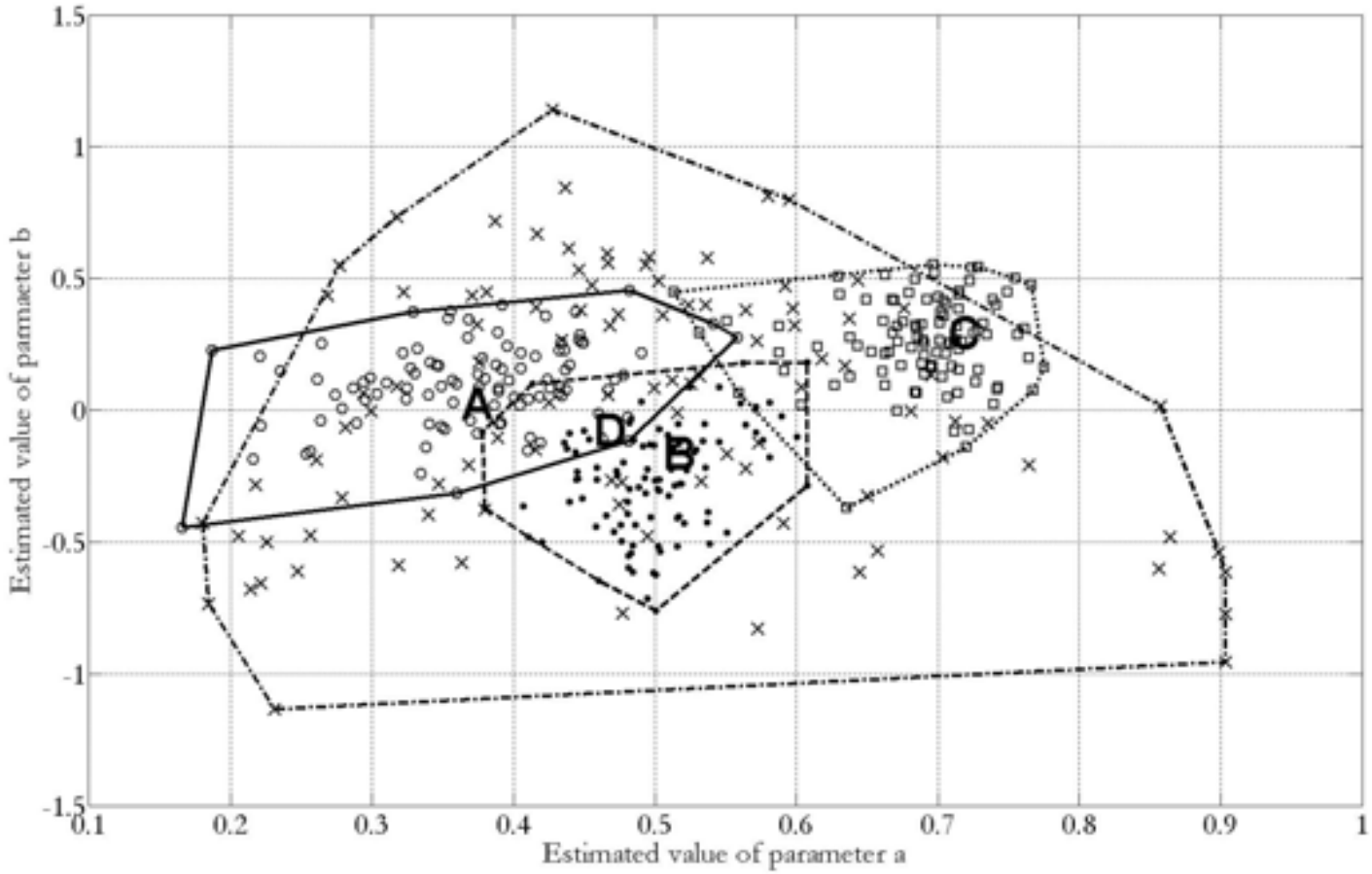


Plot of residuals $\{\hat{z}_i\}_{i=1}^n$ against values $\{x_i\}_{i=1}^n$ of conditioning variate, for the measured NNS sample. We hope for no structure in the scatter plot



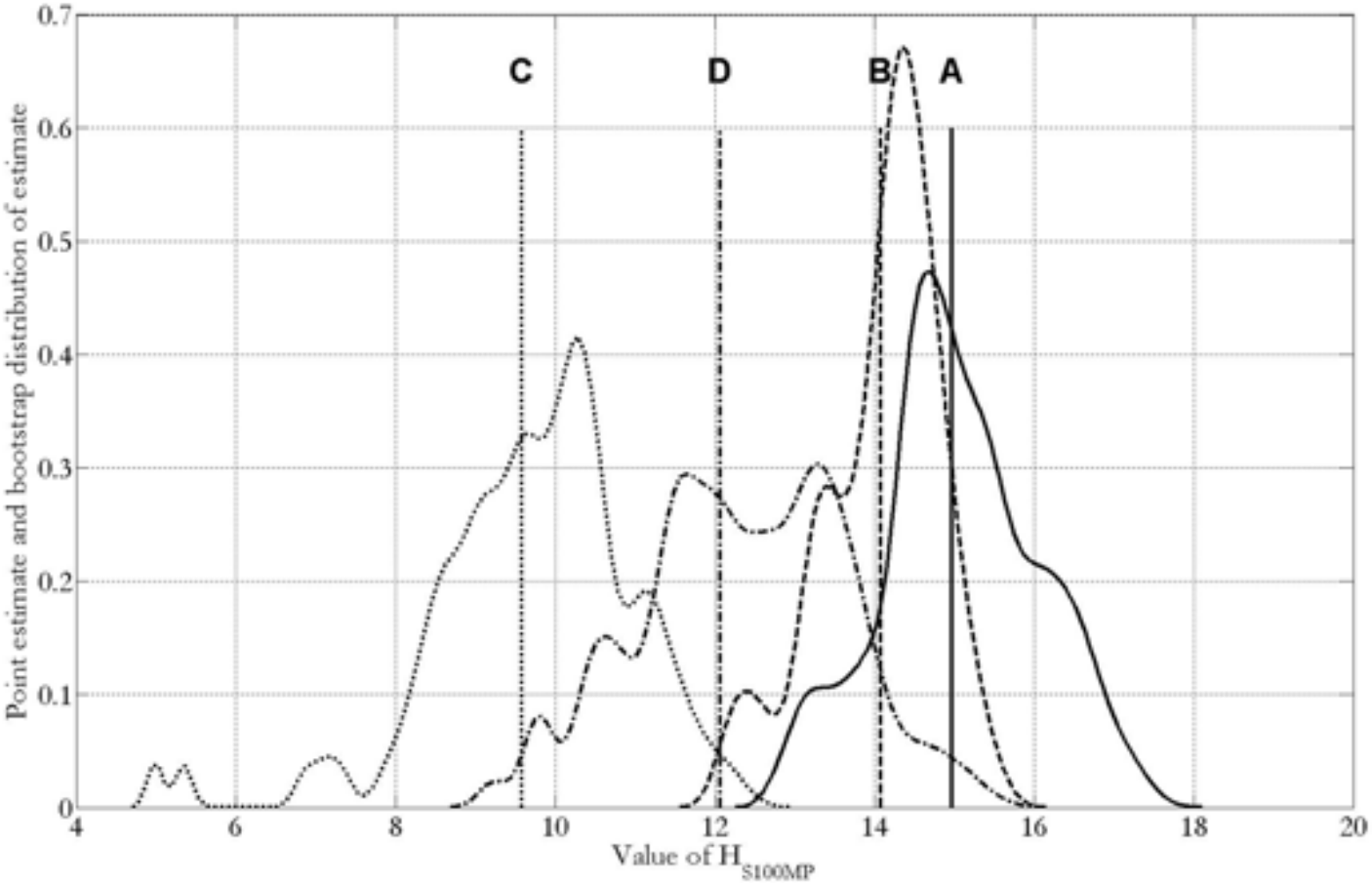
Variation of conditional model parameter estimates $\hat{a}$ (dots) and $\hat{b}$ (circles) with threshold for conditional modelling, for the measured NNS sample

Comparing model parameter estimates



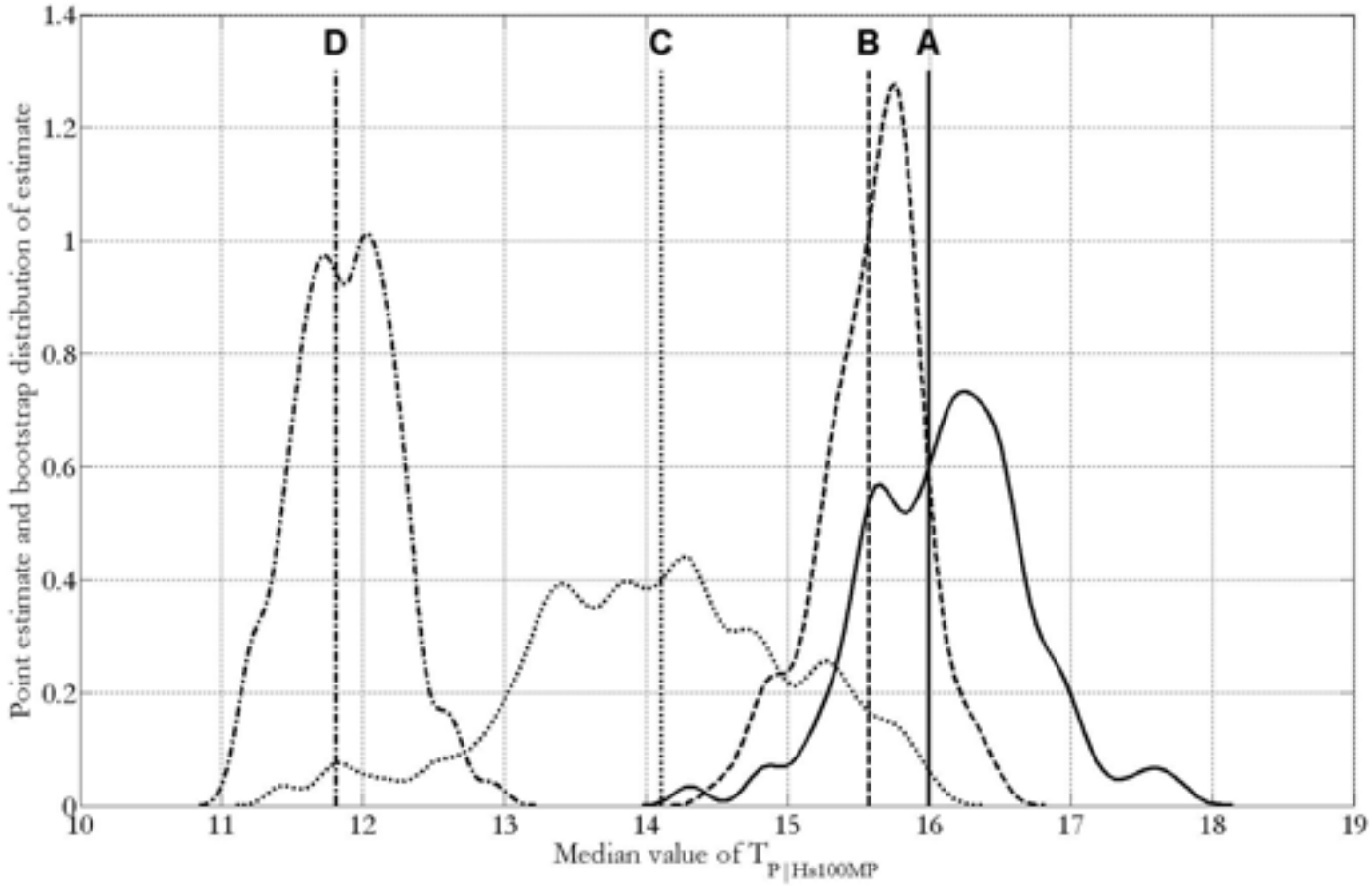
Point and bootstrap convex hull estimates for conditional model parameters *a* and *b* for the four applications. Point estimates (A-D) and bootstrap convex hulls shown corresponding to the measured NNS sample (A, circles), the NNS hindcast sample (B, dots), the measured GoM sample (C, squares), and the NWS hindcast (D, crosses).

Extreme quantile of H_{S100MP}



Point and kernel density estimates for the bootstrap distribution of extreme quantile H_{S100MP} for each of the four applications: measured NNS (A, solid), hindcast NNS (B, dashed), measured GoM (C, dotted) and hindcast NWS (D, dashed-dotted).

Associated extreme quantile of $T_{P|H_{S100MP}}$



Point and kernel density estimates for the bootstrap distribution of extreme quantile $T_{P|H_{S100MP}}$ for each of the four applications: measured NNS (A, solid), hindcast NNS (B, dashed), measured GoM (C, dotted) and hindcast NWS (D, dashed-dotted).

Main findings

Joint modelling of wave climate parameters based on solid statistics

- Works well in application to simulated, measured and hindcast data
- Threshold selection a particular issue
- Demonstrate stability of inferences with respect to tuning parameters
- Evaluate uncertainty of modelling procedure
- Key assumption: limit representation for $Z = \frac{(Y - a(x))}{b(x)}$, namely $Pr(Z < z | X = x) \rightarrow G(z)$ as $x \rightarrow \infty$
- Set of estimated residuals $\{\hat{z}_i\}_{i=1}^{n}$ may be inadequate to characterise the distribution G of Z
- Idea: Represent Haver model as special case of conditional approach

Conditional model:

- is applicable in much higher dimensions
- is useful not just for joint estimation of wave climate parameters
- avoids need for dimensionality reduction (e.g. EOF)
- avoids assumptions on structure variables (e.g. "response-based")

Need to consider:

- accommodation of covariate effects
- variable transformation for marginal modelling
- partitioning according to driving physical processes

Thanks for listening.

philip.jonathan@shell.com
jan.flynn@shell.com
kevin.ewans@shell.com

S. Haver. Wave climate off northern Norway. *Applied Ocean Research*, 7: 85–92, 1985.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497, 2004.