Uncertainties In Extreme Wave Height Estimates For Hurricane Dominated Regions

Philip Jonathan
Shell Research Limited

Kevin Ewans
Shell International Exploration and Production

Overview

- Background
- Motivating data
- Generalised Pareto modelling
- Effect of sample size and site averaging
- Bootstrapping to estimate parameter uncertainty
- The statistics of H_{S100}

Acknowledgement

- Shell International Exploration and Production
- Shell Research Ltd
- Key references

  - Elsignhorst C, Groeneboom P, Jonathan P, Smulders L and Taylor PH (1998) "Extreme value analysis of North Sea storm severity" J Offshore Mechanics Ocean Engineering 120 177-183.
  - Efron B and Tibshirani RJ (1993) "An introduction to the bootstrap" Chapman and Hall (New York).

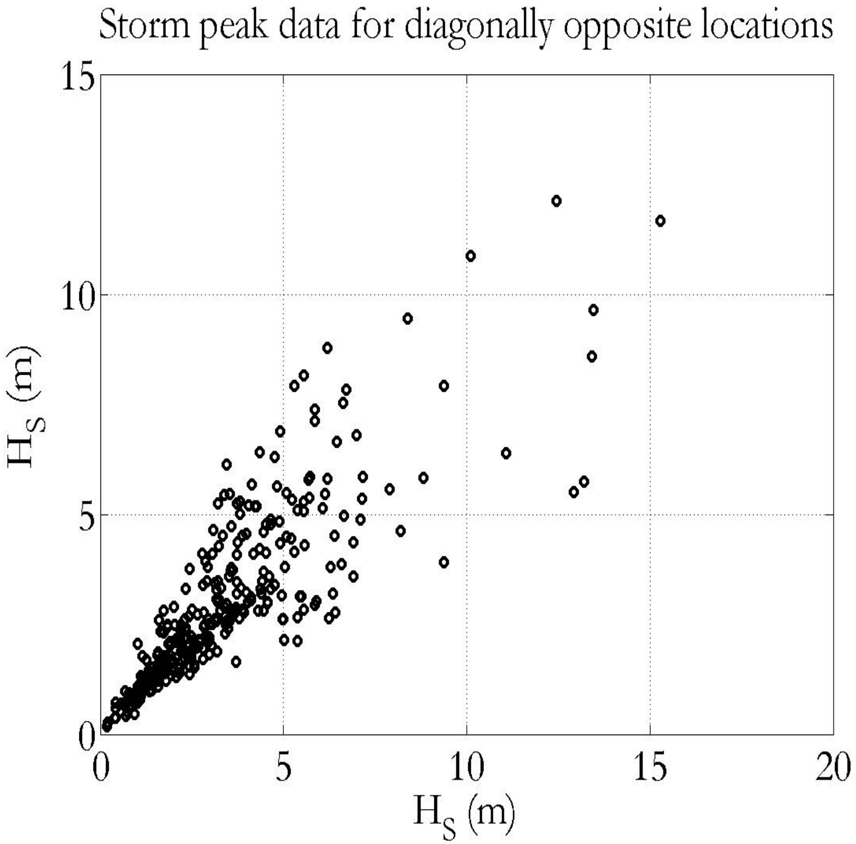
Estimating extremes using correlated data

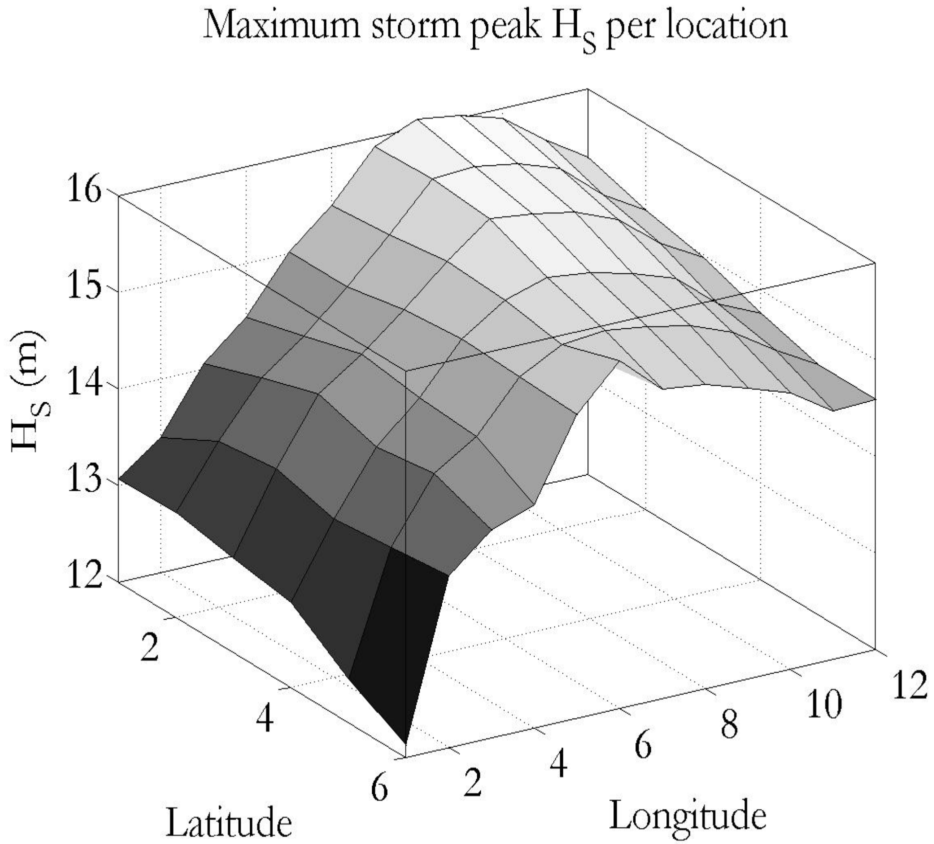
- Environmental design criteria inherently uncertain
  - Climate variability
  - Quantity and quality of data estimating design criteria
- GoM specifics
  - Hurricanes infrequent (c.f., e.g., extra-tropical storms in NNS)
  - Hurricanes relatively small scale (c.f. extra-tropical storms)
    - Hurricane track important influence on severity of sea state any location.
- Site averaging
  - In principle:
    - Increases sample size for modelling
    - Accounts for effects of random storm track
  - However:
    - Data from even quite largely separated locations are highly correlated.
    - Difficult to determine the reliability (or equivalently the degree of uncertainty) associated with design criteria derived from the site averaging approach.

The GOMOS data

- H_S values from GOMOS Gulf of Mexico hindcast Study (Oceanweather, 2005)
- September 1900 to September 2005 inclusive
- 30-minute sampling intervals
- 72 grid points arranged on a 6 x 12 rectangular lattice with spacing with 0.125° (14 km).
- For each storm period for each grid point, we isolated the storm peak significant wave height

The GOMOS data





Diagonally opposite locations are >150km apart

Storm peak H_S values for different locations are highly correlated

The model, estimates and uncertainties

- The cumulative density function: $F(x;\gamma,\sigma) = 1 - \left(1 + \frac{\gamma}{\sigma}(x-u)\right)^{-\frac{1}{\gamma}} \quad x > u, \sigma > 0$

  γ = shape / extreme value index, σ = scale, u = threshold (pre-specified)

- Maximum likelihood estimates $(\hat{\gamma},\hat{\sigma})$. Care that $(\hat{\gamma},\hat{\sigma})$ not materially affected by choice of u.
- Analytic asymptotic variances for $(\hat{\gamma},\hat{\sigma})$: $\sigma^2{}_{a\hat{\gamma}} = \frac{(1+\hat{\gamma})^2}{n}$ $\quad \sigma^2{}_{a\hat{\sigma}} = \frac{2\hat{\sigma}^2(1+\hat{\gamma})}{n}$
- Most probable 100 year significant wave height: $H_{S100yrMP} = \frac{\hat{\sigma}}{\hat{\gamma}}\left(p^{-\hat{\gamma}} - 1\right) + u$

  $p = \frac{P}{100n}$, P = period of data, n = number of data.

- Analytic asymptotic variance of $H_{S100yrMP}$:

$$\sigma^2{}_{H100yrMP} = \frac{K^2(1+\hat{\gamma})^2 + 2K\hat{\sigma}\left(\frac{1+\hat{\gamma}}{\hat{\gamma}}\right)\left(p^{-\hat{\gamma}}-1\right) + 2\hat{\sigma}^2\left(\frac{1+\hat{\gamma}}{\hat{\gamma}^2}\right)\left(p^{-\hat{\gamma}}-1\right)^2}{n}, \quad K = \frac{\hat{\sigma}}{\hat{\gamma}^2}\left(p^{-\hat{\gamma}} - 1\right) + \frac{\hat{\sigma}}{\hat{\gamma}}p^{-\hat{\gamma}}\log_e p.$$

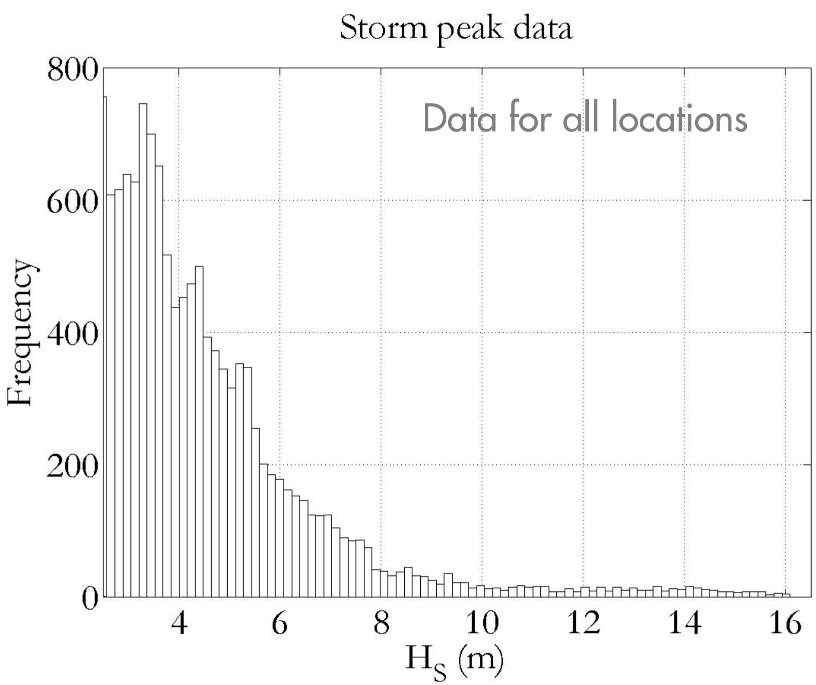
Care needed with choice of threshold, u
Asymptotics used to studentise during bootstrapping

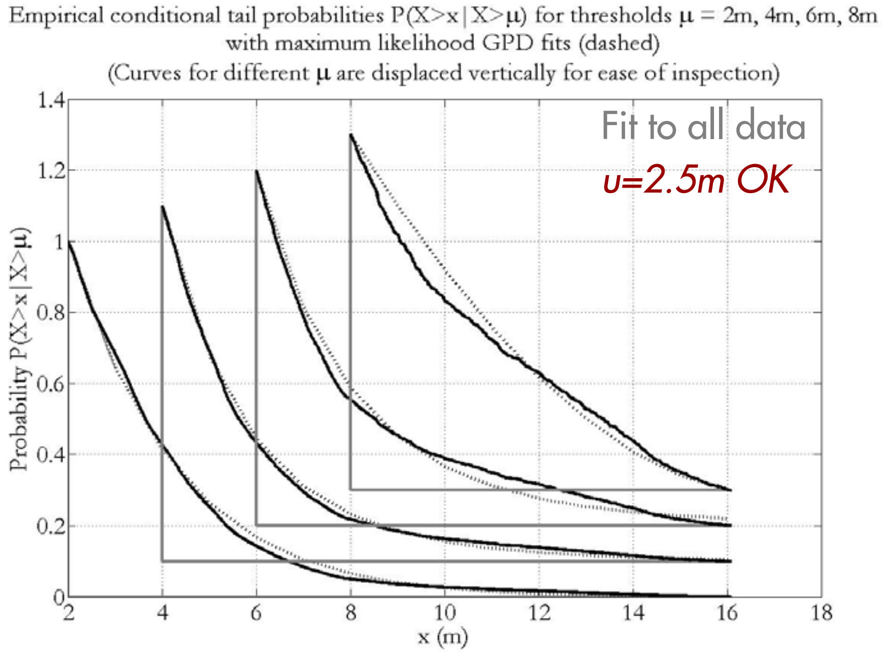
Extreme value estimates for GOMOS

Method	$\hat{q}$ (m)	$\sigma_{a\hat{q}}$ (m)	$\hat{\gamma}$	$\sigma_{a\hat{\gamma}}$	$\hat{\sigma}$	$\sigma_{a\hat{\sigma}}$
Mean of individual estimates per grid location	13.16	1.42	-0.024	0.08	2.22	0.24
Single estimate based on all grid locations	13.19	0.17	-0.022	0.01	2.22	0.03

Table 1: Extreme value estimates and uncertainties. Note: q used as short hand for $H_{S100yrMP}$.

Estimates above agree. Uncertainties are very different. Reality is between these extremes. But where?



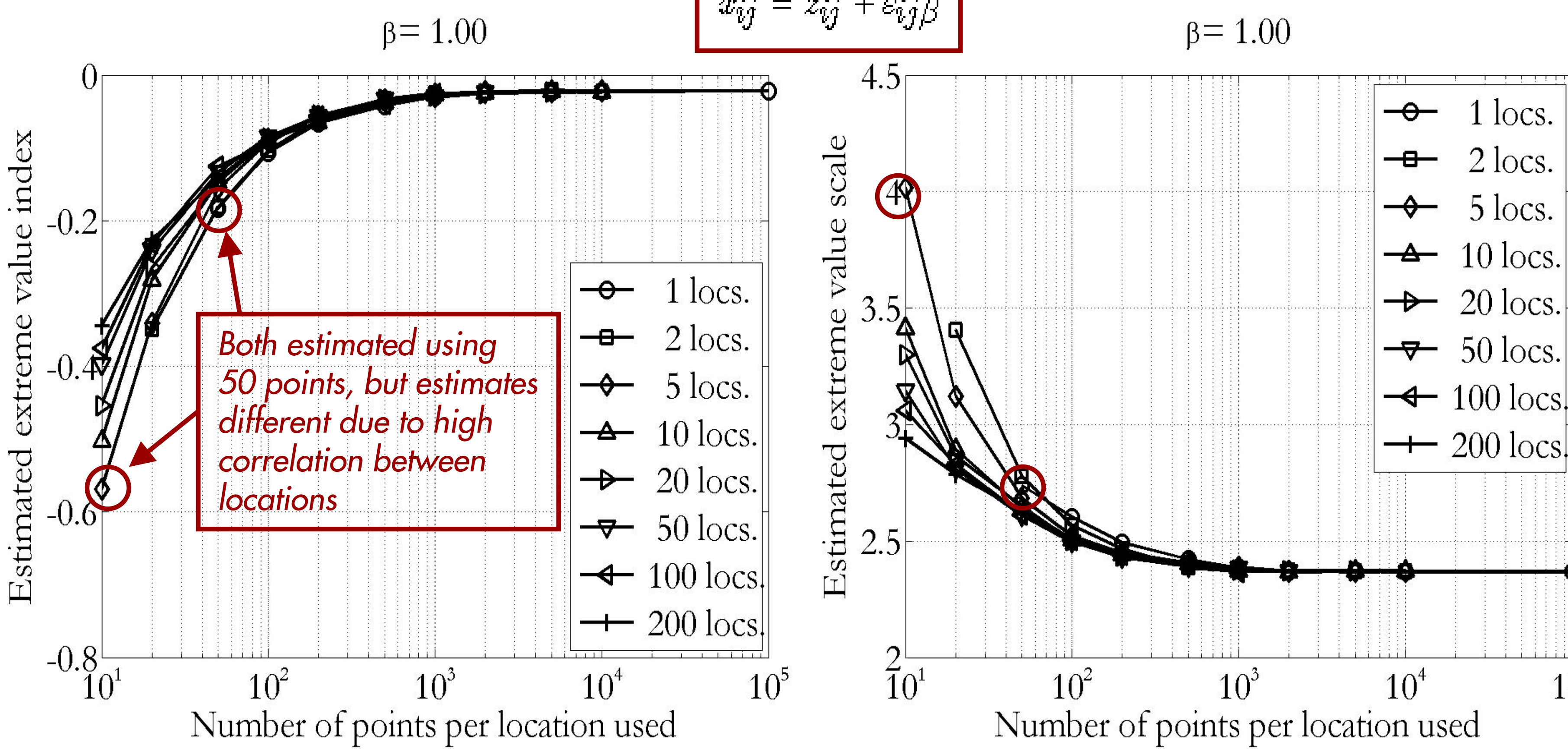


Effect of sample size and site averaging on extreme value estimates for correlated data

- Simulation study
- Simple model for inter-correlated storm peak data
- Draw random samples from GPD, perturbed by Gaussian noise
- Select GPD parameters to mimic the GoM
- Adjust inter-correlation by changing the standard deviation of the Gaussian noise perturbation.

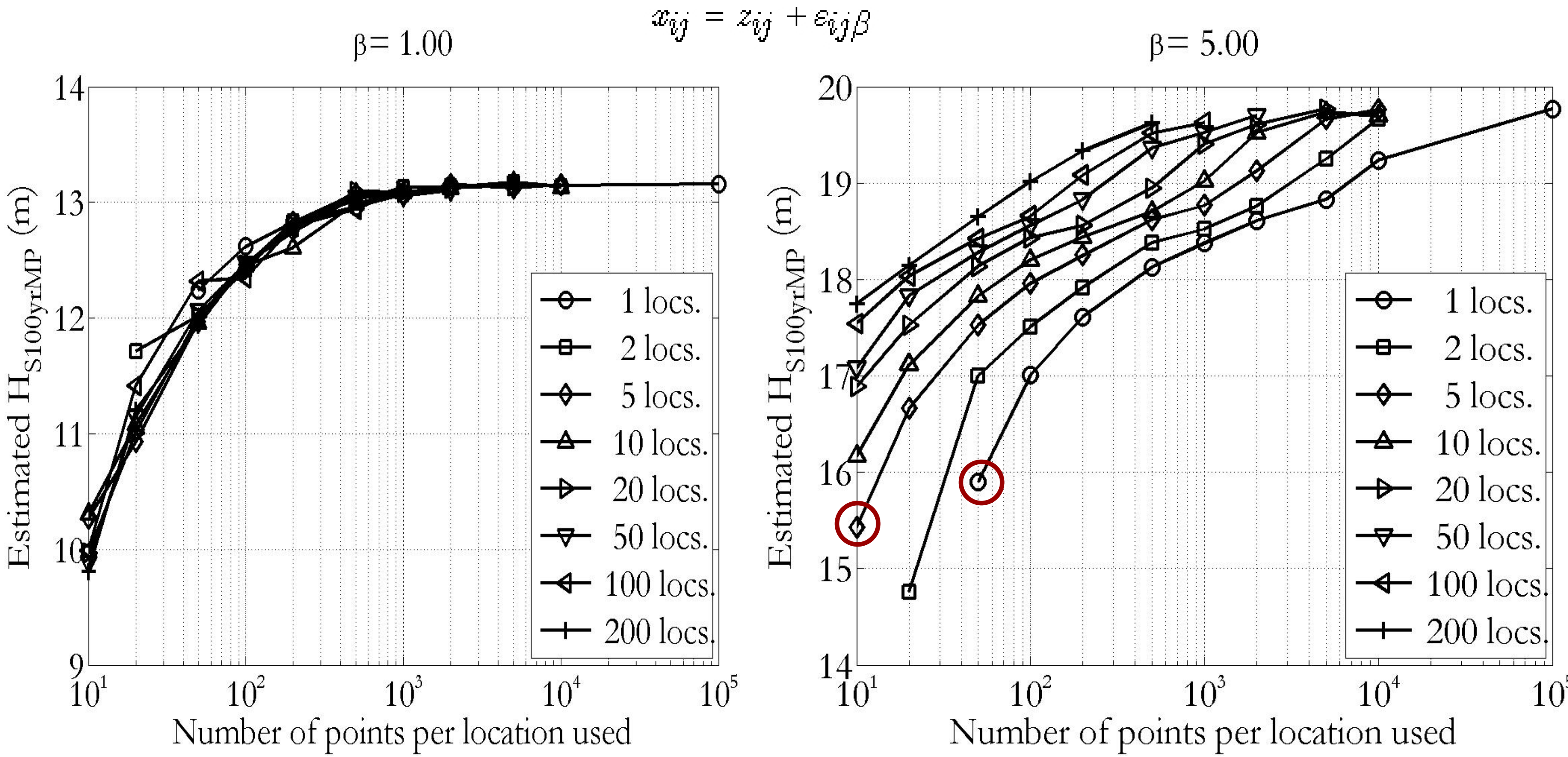
Effect of sample size and site averaging on extreme value estimates for correlated data

$$x_{ij} = z_{ij} + \varepsilon_{ij}\beta$$



beta=1 corresponds to high intercorrelation of locations

Effect of sample size and site averaging on extreme value estimates for correlated data



High intercorrelation

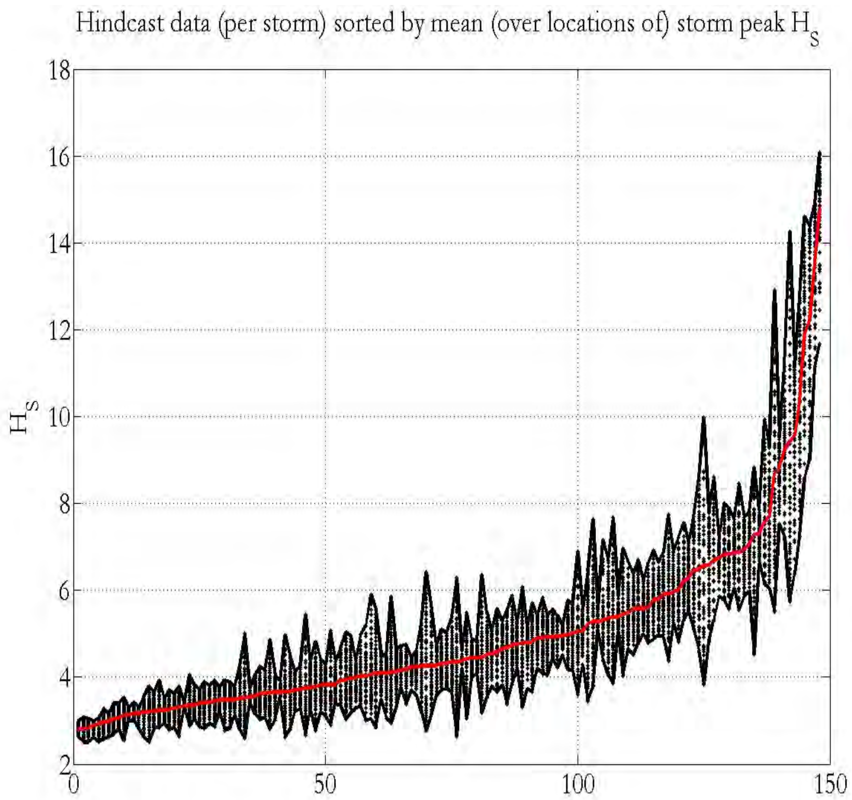
Low intercorrelation

Estimating uncertainty using bootstrapping

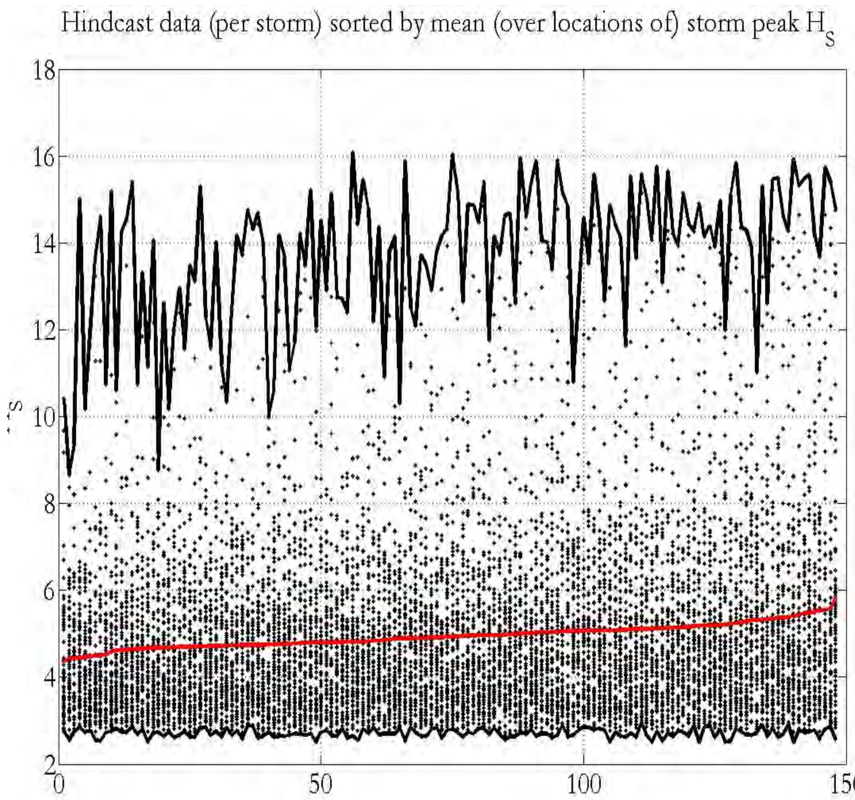
- Estimate uncertainties of extreme value parameters and quantile estimates directly
- Simple studentised bootstrapping (resampling)
  - Fit a GPD model to actual data (to obtain parameters P1)
  - Create bootstrap sample by *resampling storm-wise* from actual data with *replacement*
  - Refit GDP model to bootstrap sample (to obtain parameters P2)
  - Use variability of P2 around P1 to characterise the uncertainty of P1 with respect to the true (unknown) underlying parameters P0
  - Standardise variability of parameters with respect to their asymptotic standard errors (ie, we *studentise*)
- Can be generally for an arbitrary dependence structure

GOMOS dependency structure

Actual storm data



Randomly selected groups of 72 values



Hindcast storm peak H_S data is highly spatially correlated
Number of effectively "independent" data somewhere between 148 and148 x 72, probably nearer 148!

Simple studentised bootstrap – exploring dependency

(Scaled) 95% limits for parameter uncertainties						
Data	$\theta=\hat{q}$		$\theta=\hat{y}$		$\theta=\hat{\sigma}$	
	$-c_\theta^+$	$-c_\theta^-$	$-c_\theta^+$	$-c_\theta^-$	$-c_\theta^+$	$-c_\theta^-$
No dependence	-1.8	2.1	-1.8	2.1	-1.9	1.9
Perfect dependence	-10.6	39.9	-13.8	28.5	-17.9	18.7
GOMOS resample	-10.3	25.2	-8.2	14.0	-11.0	11.6

Coverage for 95% interval estimates (numbers of left- and right-hand exceedences)						
Data	$\theta=\hat{q}$		$\theta=\hat{y}$		$\theta=\hat{\sigma}$	
	LH	RH	LH	RH	LH	RH
No dependence	14/500	14/500	17/500	13/500	18/500	17/500
Perfect dependence	27/500	21/500	8/500	9/500	16/500	7/500
GOMOS resample	21/500	9/500	24/500	12/500	6/500	28/500

Point and interval estimates for GOMOS data			
Parameter	$\hat{q}$	$\hat{y}$	$\hat{\sigma}$
Point Estimate (all locations)	13.0	-0.098	2.68
Single location (mean asymptotic)	(10.7, 15.3)	(-0.246, 0.044)	(2.11, 3.27)
Simple studentised bootstrap	(11.5, 16.3)	(-0.164, 0.015)	(2.28, 3.07)
All locations (asymptotic)	(12.7, 13.3)	(-0.115, -0.081)	(2.61, 2.75)
All location parametric bootstrap	(12.8, 13.3)	(-0.114, -0.078)	(2.61, 2.75)

Studentised resampling scheme to estimate parameter uncertainty

Performance quantified in terms of coverage

Scaled limits consistent with intuition

Coverage reasonable

Interval estimates using simple studentised bootstrap are:

- *Narrower than single location estimate*
- *Much wider that those obtained using other "all location" estimates*
- *Reflects chance that next storm will be a Rita or Katrina*

The statistics of H_{S100}

- Extreme value analysis estimates *most probable* quantile value H_{S100MP} only
  - but H_{S100} is a stochastic quantity!
- Possible to estimate distributional properties (e.g 95% confidence intervals) for quantile estimates also
- Width and asymmetry of distribution of quantile estimates increases with increasing extreme value index
  - For Northern North Sea conditions, distribution is approximately symmetric (index is $\cong -0.3$)
  - For GoM, distribution is asymmetric (index is $\cong 0$)
  - *Larger values of H_{S100} more likely in GoM than in NNS for same value of H_{S100MP}*

The statistics of H_{S100}

- We know that $H_{S100yrMP} = \frac{\hat{\sigma}}{\hat{\gamma}}\left(p^{-\hat{\gamma}} - 1\right) + u$

  where $p = \frac{1}{n_{100}}$ and n_{100} is the expected number of storms in a 100 year period.

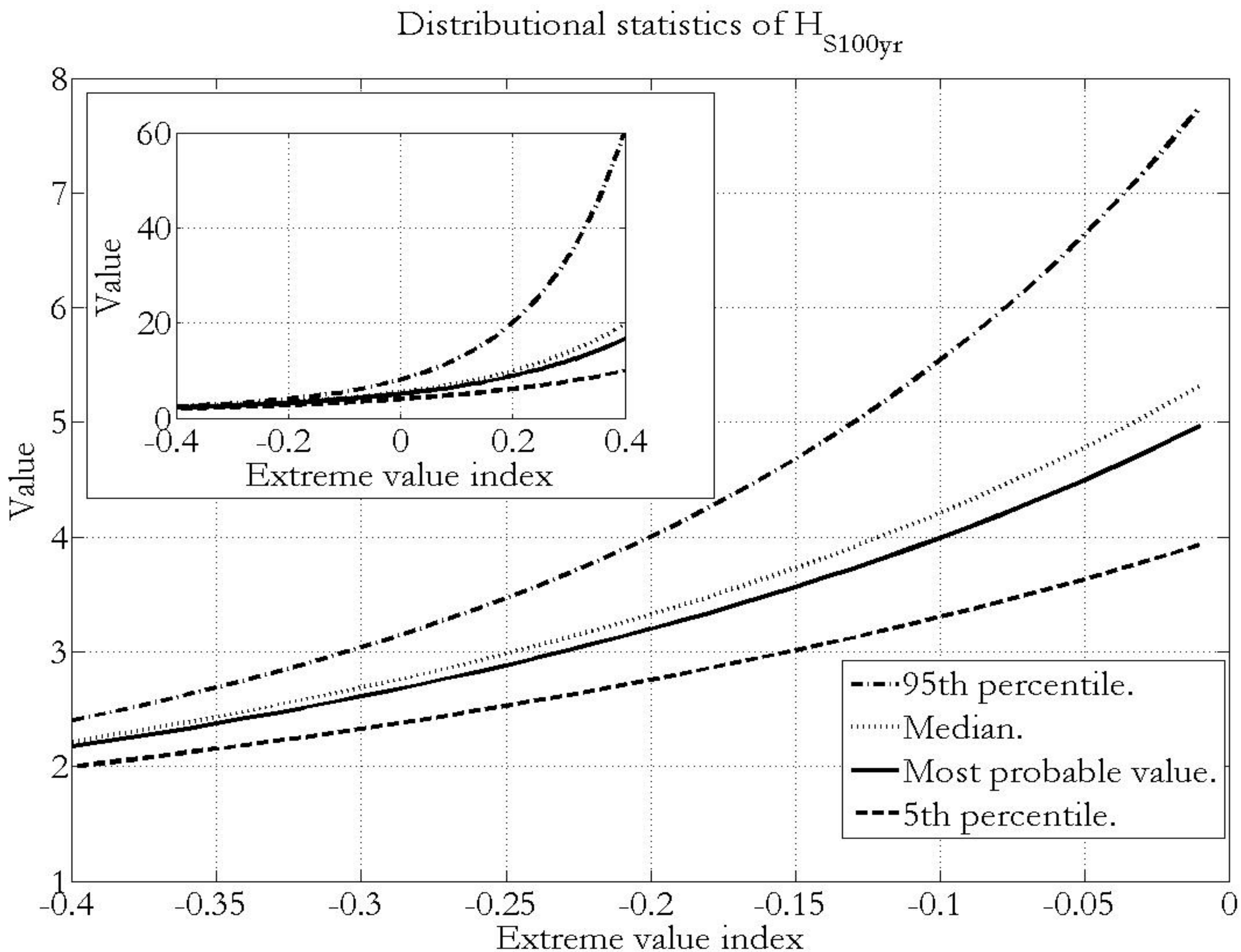
- Similarly we derive an expression for $H_{S100yr(1-q)}$, the value of the 100-year significant wave height exceeded with probability q in any 100-year period:

  $$H_{S100yr(1-q)} = \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(1-(1-q)^{1/n_{100}}\right)^{-\hat{\gamma}} - 1\right) + u$$

- Specifically, $H_{S100yr0.95}$ is the value of H_{S100yr} exceeded once in every 20 independent locations studied. For $H_{S100yr0.95}$, q =0.05.

- For current data, most probable value for H_{S100yr} is approximately 13m, whereas the value of $H_{S100yr0.95}$ is above 17m.

The statistics of H_{S100}



Asymmetry of H_{S100} distribution increases with increasing extreme value index

For given fixed H_{S100MP}, chance of a very large event increases as extreme value index increases

Assumes standard case $\hat{\sigma}=1$, $u=0$.

For other $\hat{\sigma}$ and u, multiply value here by $\hat{\sigma}$ then add u.

Conclusions

- Small samples give biased extreme value estimates
  - Using less than 100 points clearly causes underestimation of events such as H_{S100MP} for the GoM conditions considered here
  - Magnitudes of extreme value index and scale overestimated for small samples
  - Aside: Similar bias effects are observed for both GPD and Weibull fitting
- Site averaging is recommended for point estimation of parameters
  - Combining correlated data from different locations:
    - Reduces bias of estimates by increasing effective sample size
    - Can have (at worst) no effect, but can also be of little benefit
  - Site averaging provides less benefit for estimates of extreme quantiles than than for index and scale
- Estimation of parameter uncertainty when site-averaging is not trivial
  - The simple studentised bootstrap provides a straightforward and (demonstrably) reliable method applicable to arbitrary dependency structures
- Estimates of the most probable value of extreme quantiles should be used with care for structural design purposes
  - Larger values of H_{S100} more likely in GoM than in NNS for same value of H_{S100MP}

Thank You

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