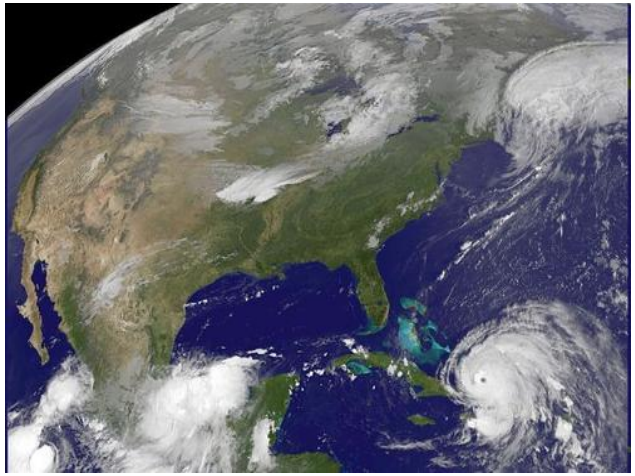
Extreme Event Analysis

Predicting the Unpredictable

Ross Towe
ross.towe@jbatrust.org

Slides prepared in collaboration with Jonathan Tawn

Extremes



Introduction

- Maxima modelling
- Threshold exceedance modelling
- Resources
- Variety of applications

Return levels

- To protect against coastal flooding a level: (e.g. sea wall height) needs to be selected which has a sufficiently small chance of being exceeded
- A T-year return level is a value, which is exceeded once on average every T-years
- Coastal flood designs use 50-1000 year return levels
- Problem: typically we have only 10-50 years of data

Return levels

- Hourly values $X_1, X_2, \ldots, X_n$
- Maxima $M_n = \max(X_1, X_2, \ldots, X_n)$
- If $n = 8766$ then M_n is the annual maximum
- T year return level is v_T with

$$\Pr(M_{8766} > v_T) = \frac{1}{T}$$

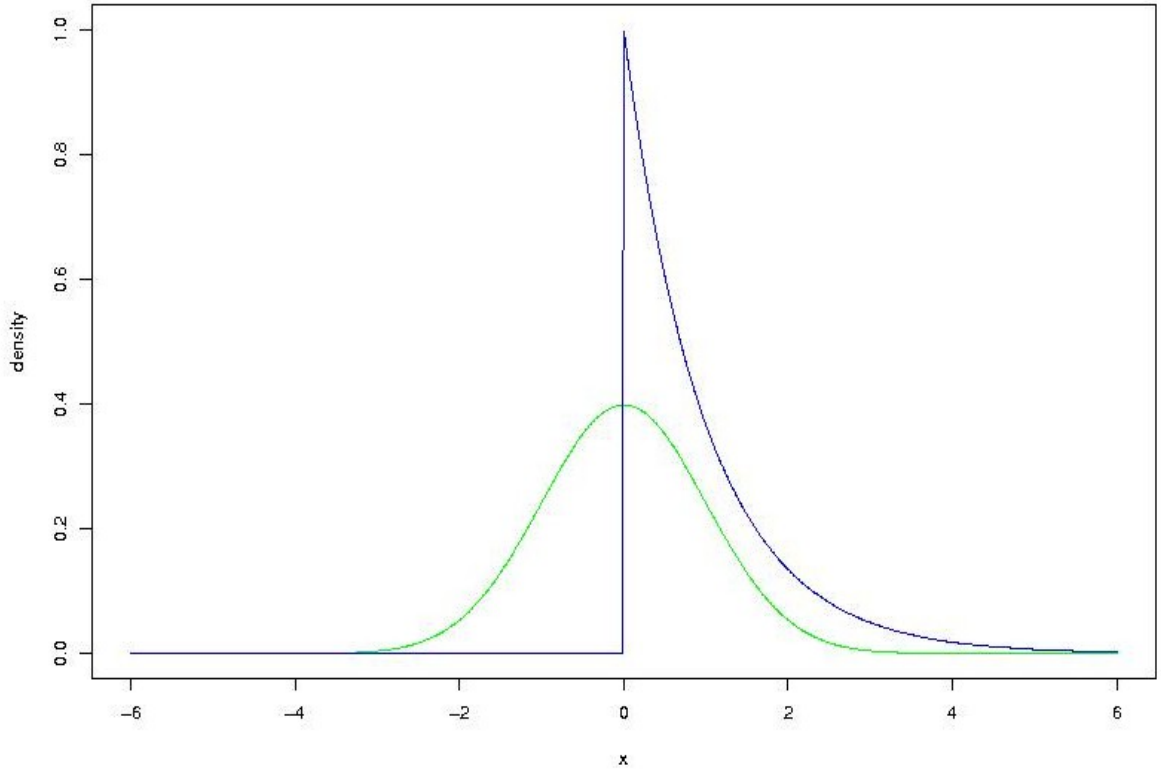
- So we need to find the probability distribution of M_{8766}.

Which distribution for M_n should we use?

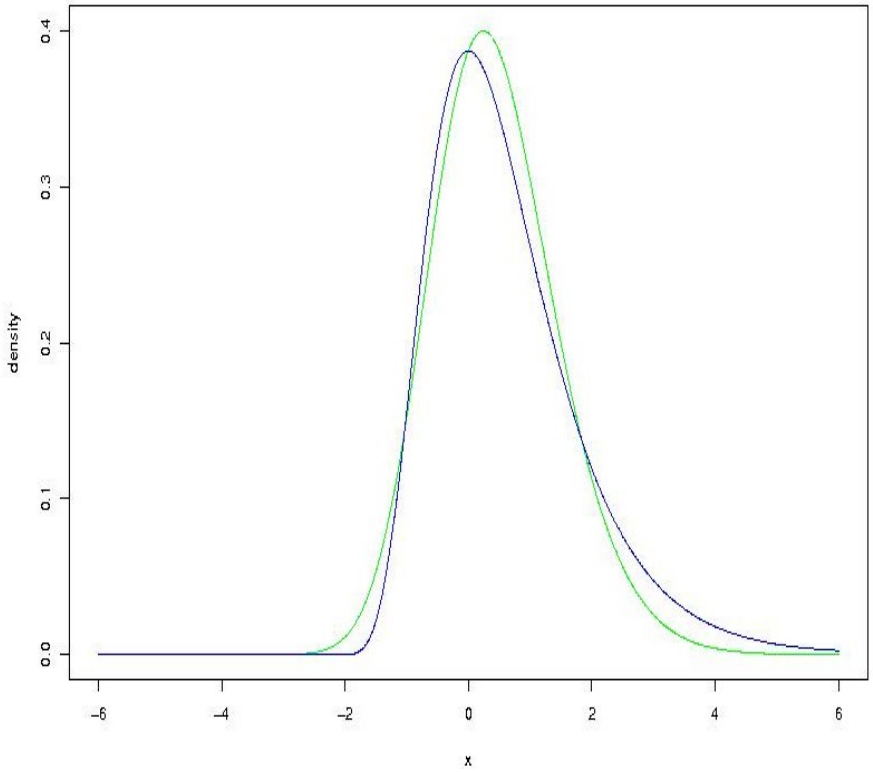
- Consider asymptotic distribution of M_n as $n \to \infty$
- Location and scale transform M_n

Illustration of convergence

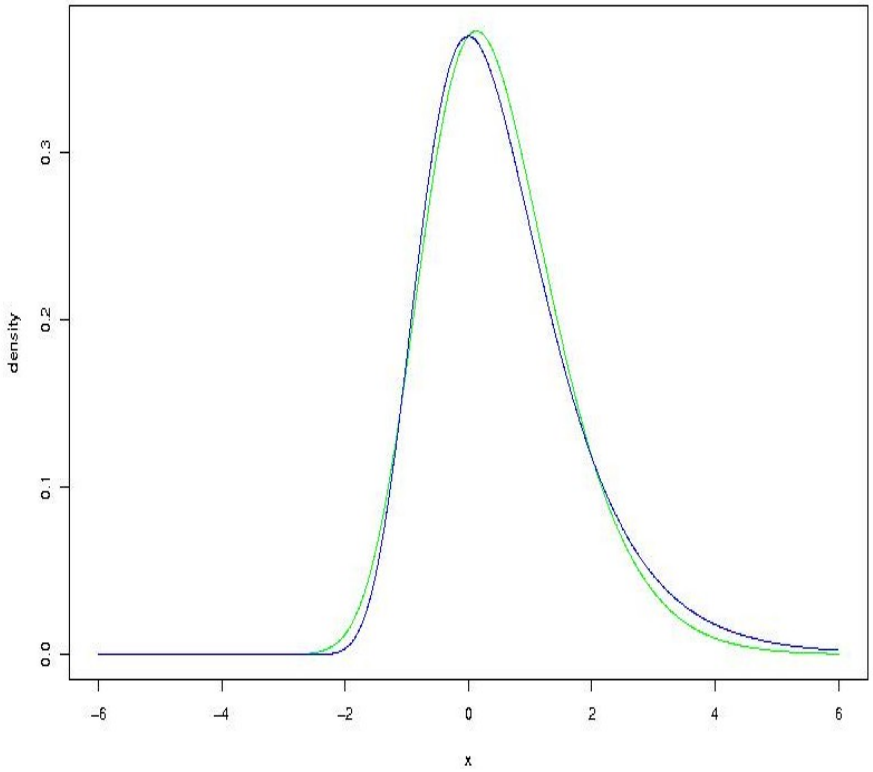
Two different initial distributions for X: (i) *Normal*, (ii) *Exponential*
Independent and identically distributed



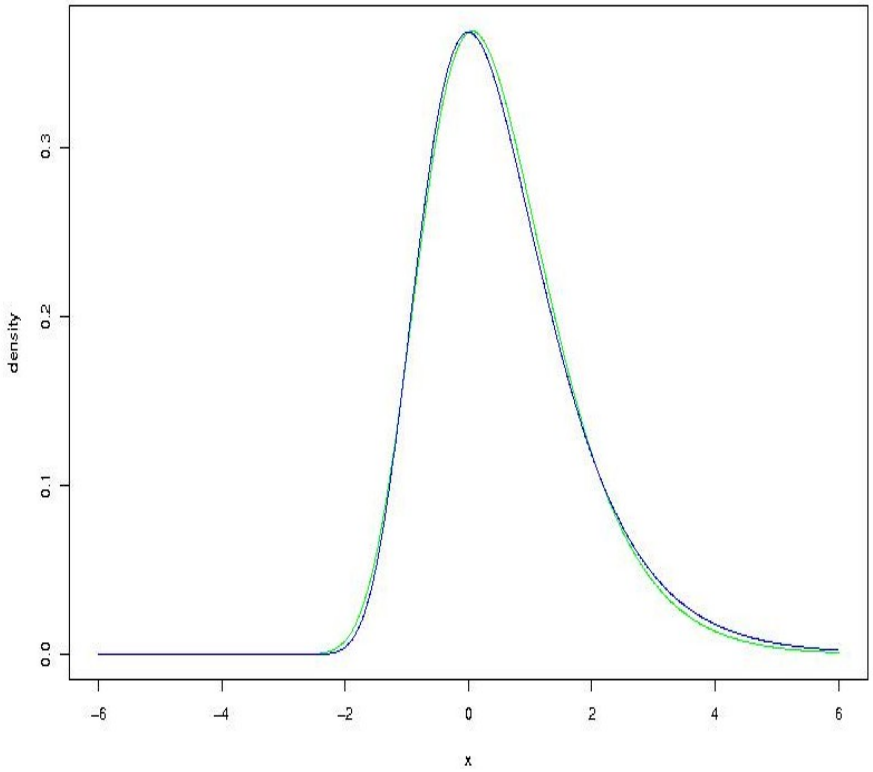
Distribution of M_{10}



Distribution of M_{100}



Distribution of M_{8766}



Extremal Types Theorem

Says

we get the same limit distribution for the maximum whatever the original distribution

- This provides a theoretical basis for extrapolation and estimation of return levels
- What is this distribution?

Generalised Extreme Value Distribution (GEV)

GEV is the limit distribution for maxima

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}$$

Parameters

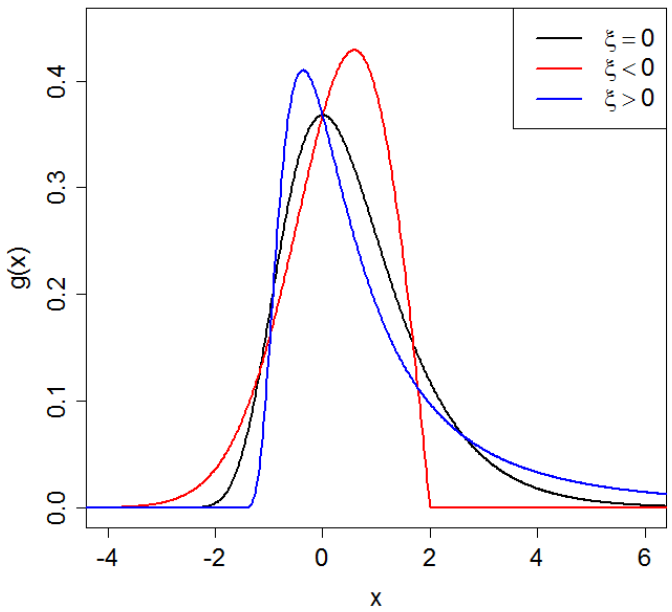
- μ location - not mean
- $\sigma > 0$ scale - not standard deviation
- ξ shape
- $[y]_+ = \max(y, 0)$ determines the support of the distribution
- Return level v_T solves

$$1 - G(v_T) = \frac{1}{T}$$

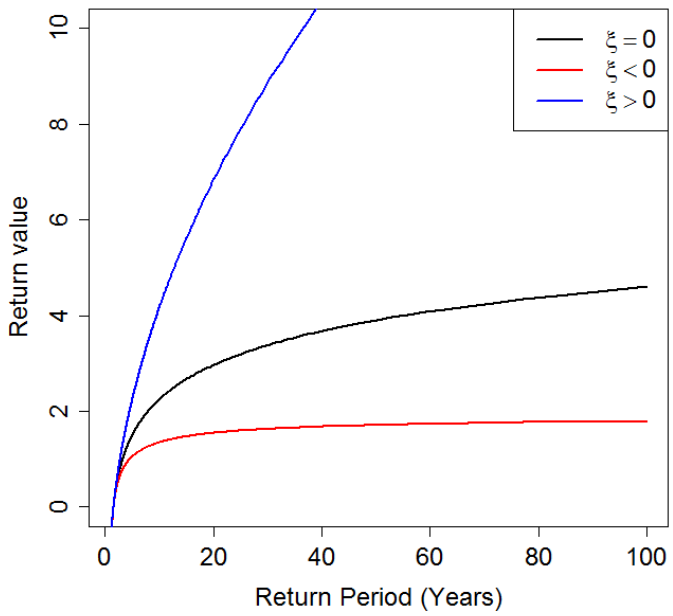
Shape parameter ξ

Fix $\mu = 0$, $\sigma = 1$ and vary ξ

Densities

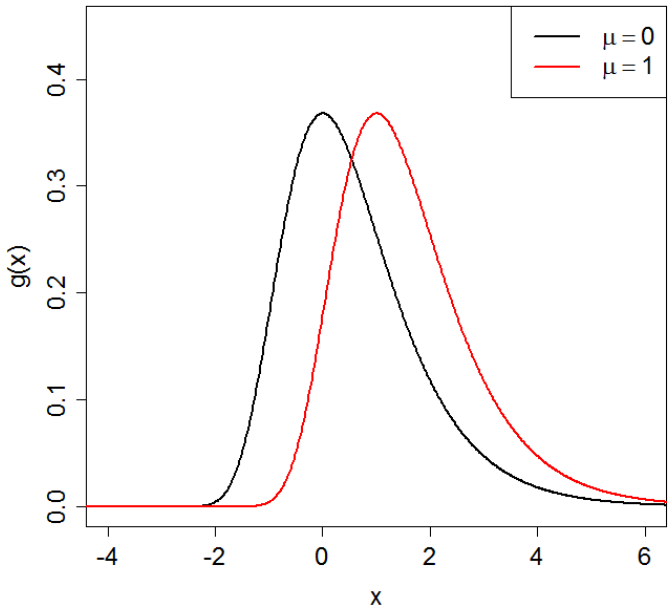


Return values

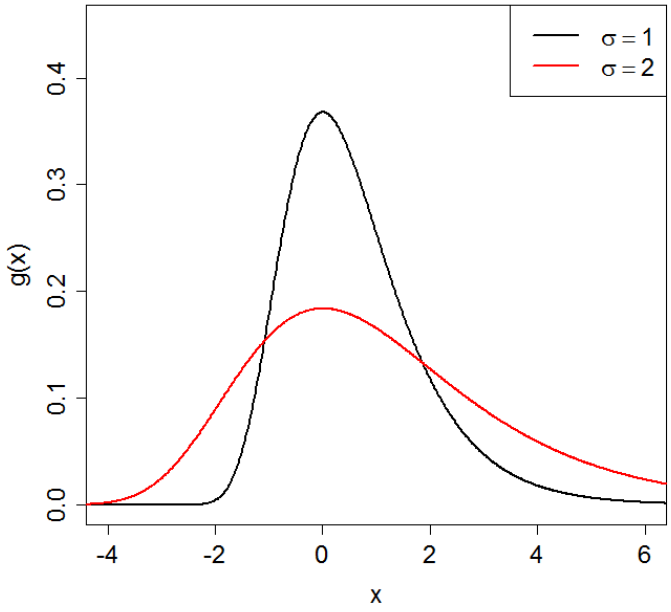


Vary μ and σ

$\mu \uparrow$



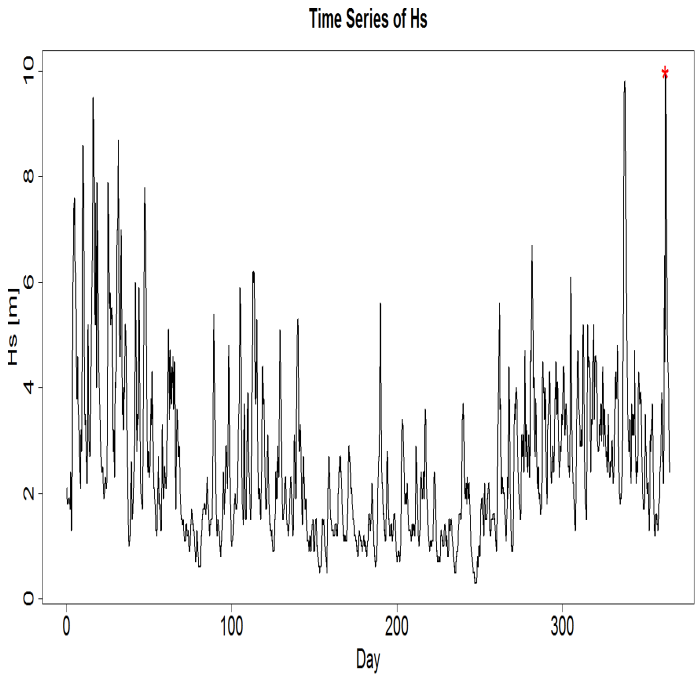
$\sigma \uparrow$



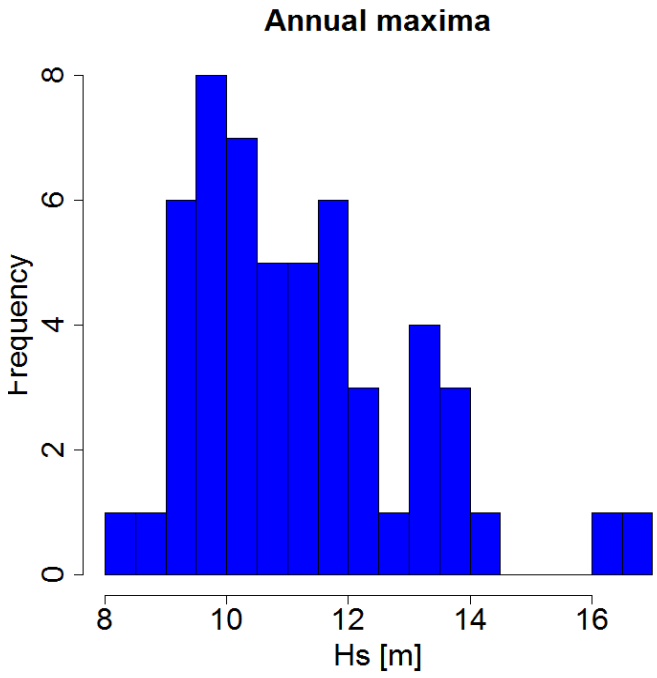
Annual Maxima

Wave Heights from 53 years of data

Time Series



Histogram of Annual maxima



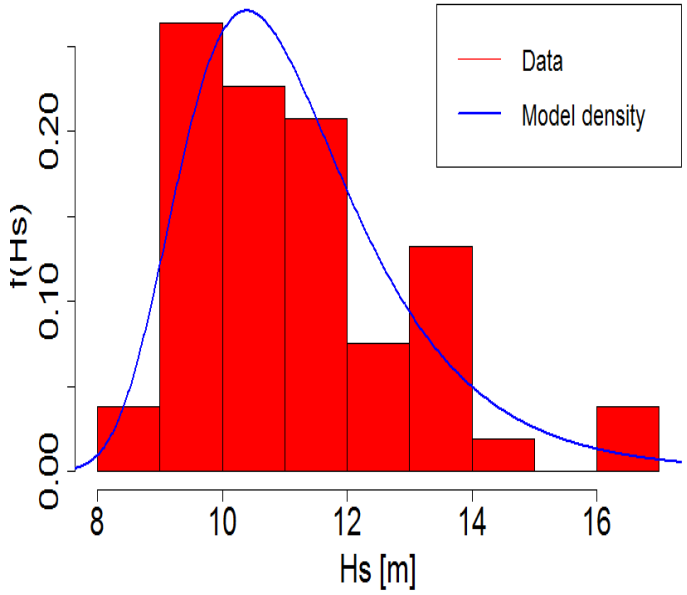
Annual Maxima

Parameter estimation

Maximise the likelihood of the GEV distribution

$$L(.) \propto \prod_{i=1}^{n} \frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi-1} \exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

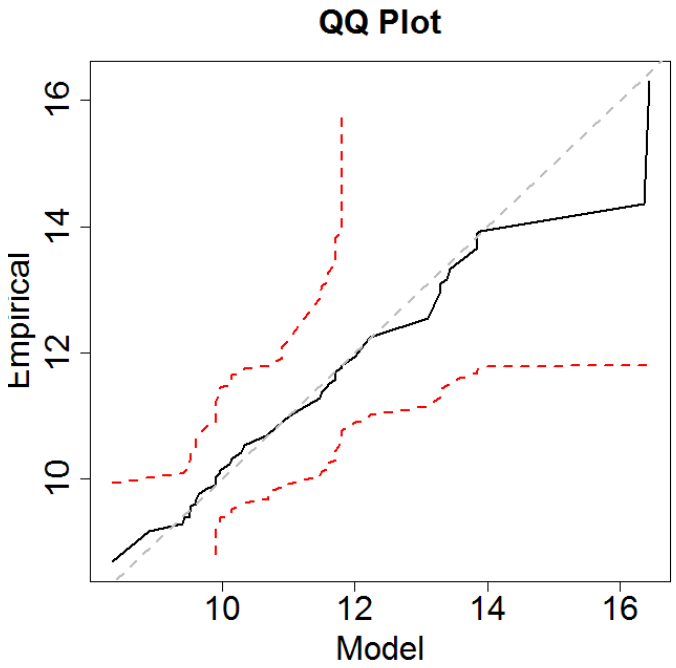
Parameter	MLE	SE
μ	10.41	0.21
σ	1.35	0.16
ξ	0.05	0.11



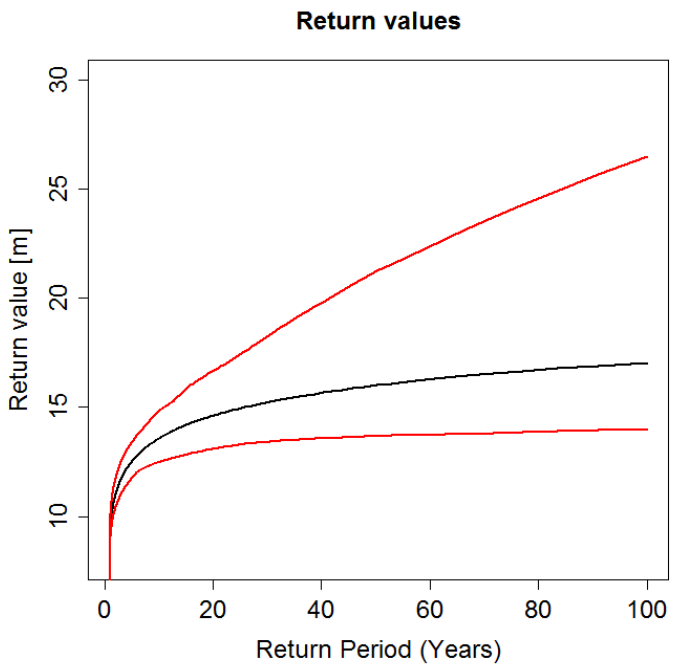
Annual Maxima

Model evaluation

QQ Plot



Return values



Environmental processes

Typical values of ξ

$\hat{\xi} > 0$:

- rainfall
- river flow
- economic damage

$\hat{\xi} < 0$:

- temperature
- wind speeds
- sea levels
- waves

$\hat{\xi} \neq 0$ but $\xi = 0$

- normal distribution
- gamma distribution

How is M_n distributed if Xs are not IID?

Not independent

- if weak long-range dependence then GEV still appropriate
- dependence changes limit from $G(x)$ to $[G(x)]^\theta$
- $0 \leq \theta \leq 1$ extremal index
- θ^{-1} mean duration of extreme events
- θ *linked* to terms like:

$$\lim_{x\uparrow} \quad \Pr(X_2 > x | X_1 > x)$$

How is M_n distributed if Xs are not IID?

Not identically distributed

- no fully general results
- if change of distribution in year is small the GEV still appropriate
- models of the form

$$GEV(\mu_t, \sigma_t, \xi_t)$$

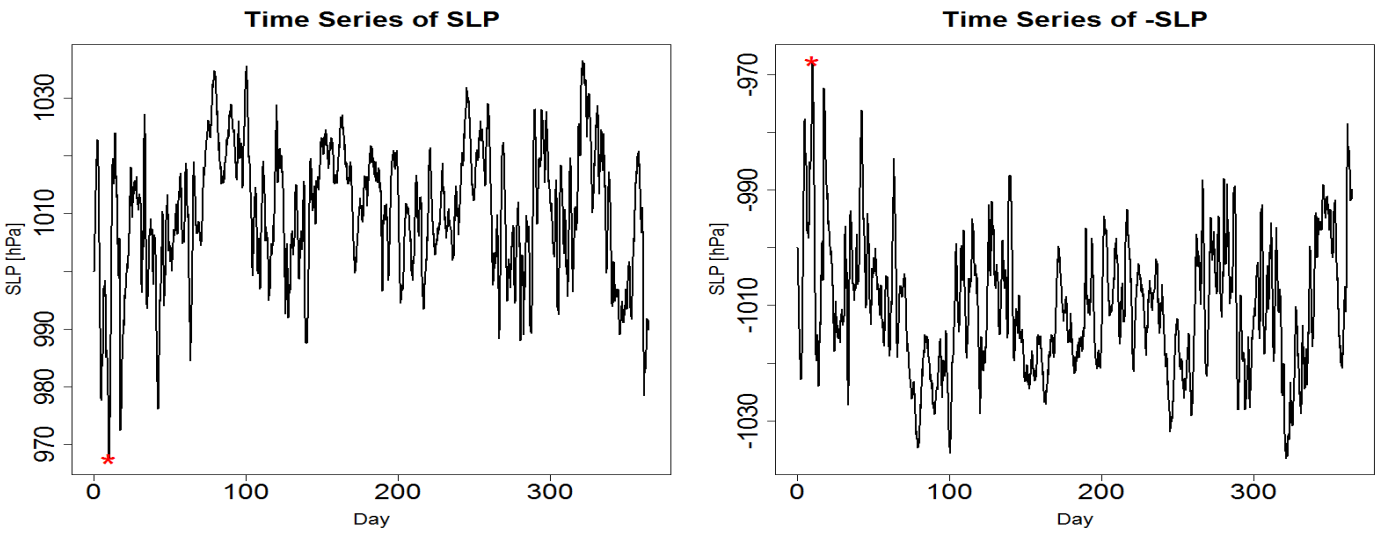
- here t is year or a function of year (e.g. yearly NAO)
- μ_t etc is a function of the covariate t

What if minima are the extremes of interest?

Suppose interest is in

$$\min(X_1, \ldots, X_n)$$

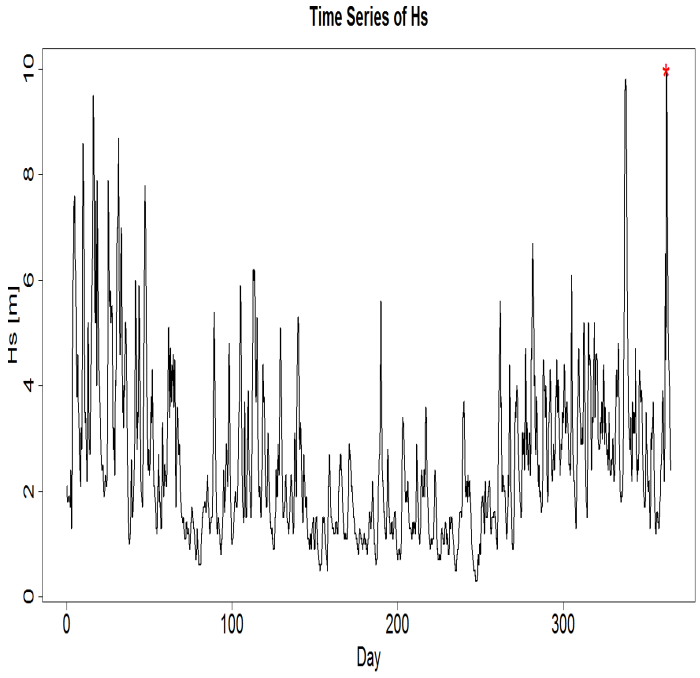
Note $$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$$



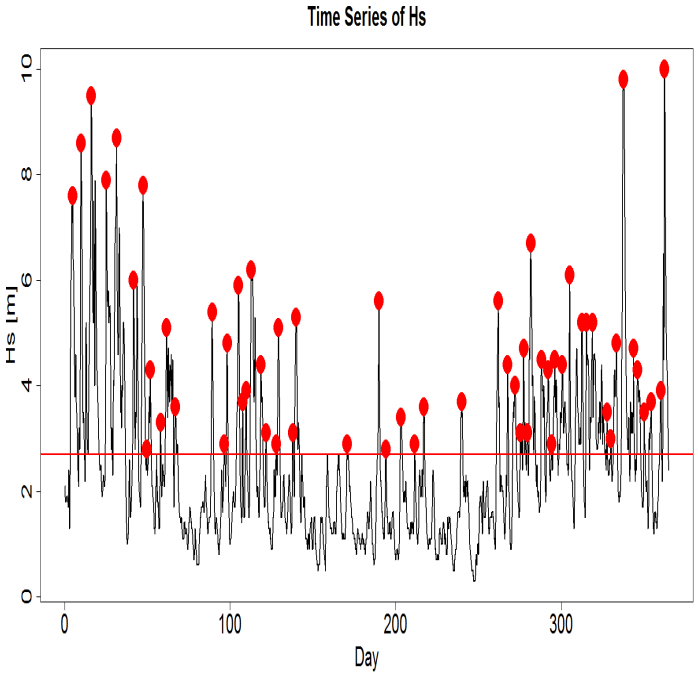
So can use methods for maxima to study minima.

Is maxima analysis an efficient use of data?

Maxima



Exceedances



Threshold Exceedances

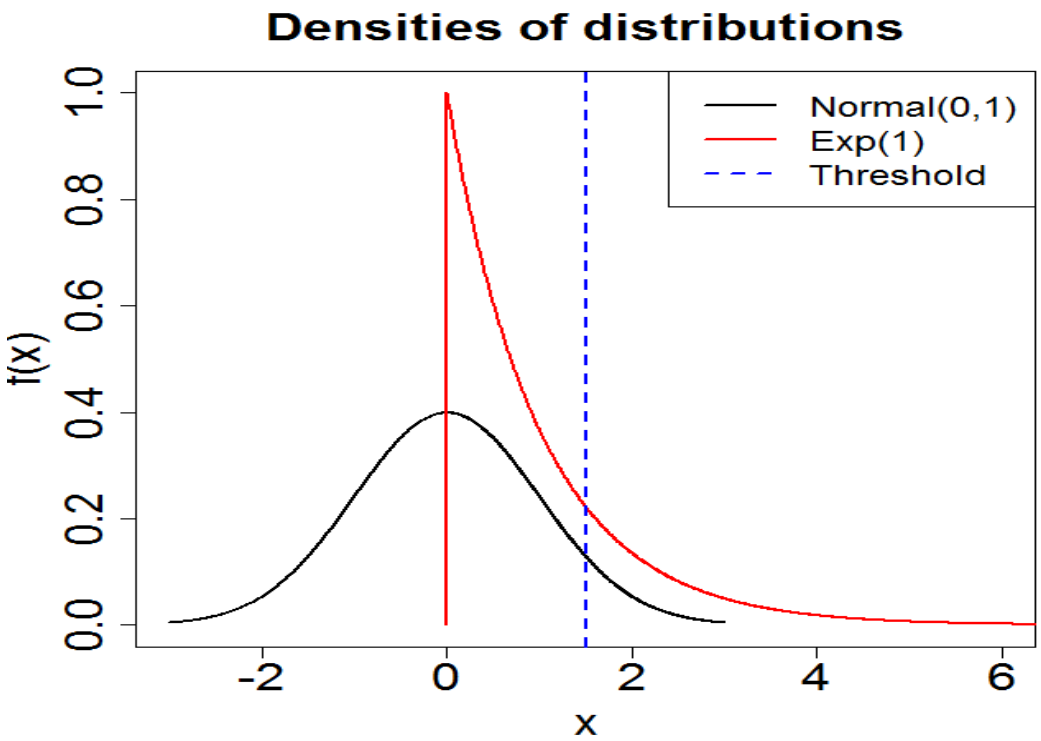
Suggests imposing a threshold u to define extremes and using

- distribution of the excesses over the threshold, $H_u(x)$
- the rate of exceedance of the threshold, λ_u
- Needs declustering: as dependence select maximum exceedance from each independent cluster of exceedances

Gives a description of the tail of the distribution function of original values

Exceedances

Convergence



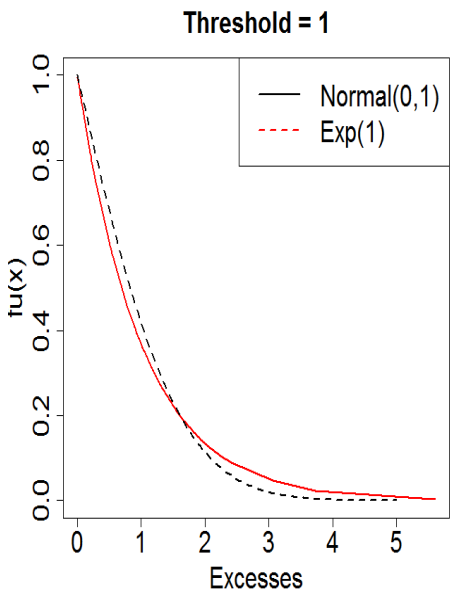
Consider points above a given threshold u

Convergence of excess distribution

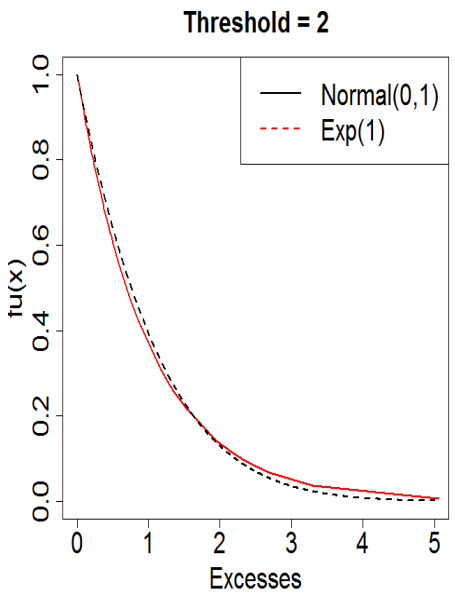
Threshold exceedances: scaled

$$a(u)(X - u) \mid X \geq u$$

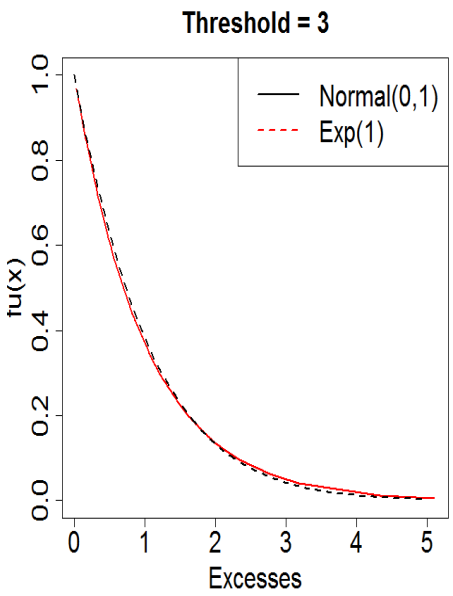
u = 1m



u = 2m



u = 3m



Generalised Pareto Distribution (GP)

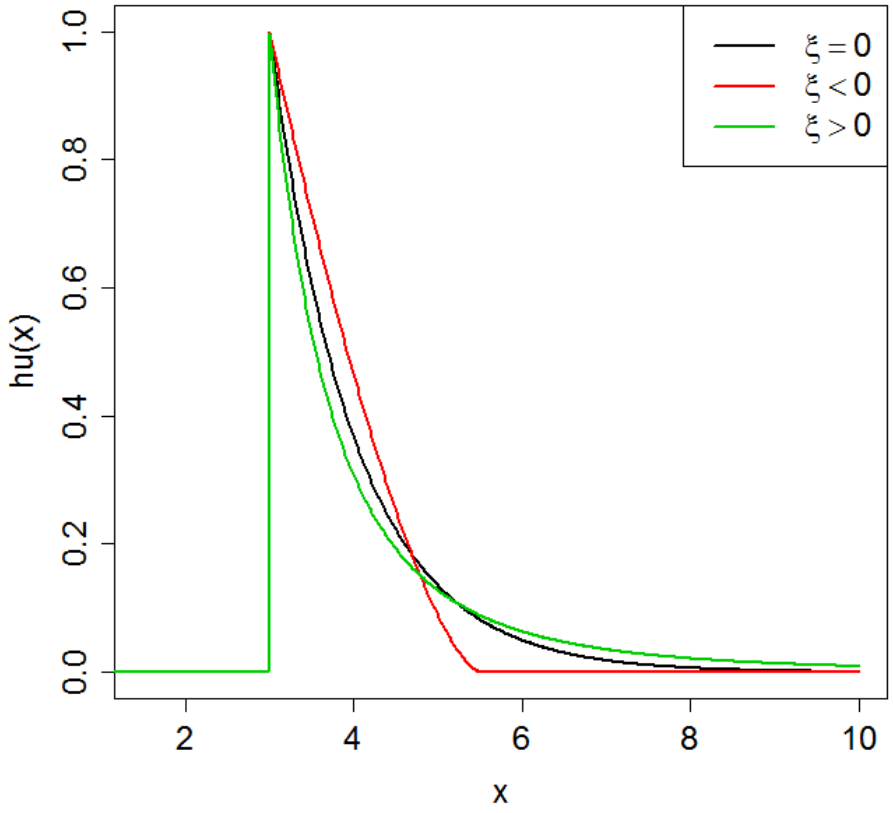
Result like the Extremal Types Theorem says that GP is the limit distribution for exceedances of threshold u

$$H_u(x) = 1 - \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \quad \text{for } x > u$$

- $\sigma_u > 0$ scale parameter that changes with threshold u
- ξ shape (same as for GEV)
- $\xi = 0$ corresponds to the Exponential(σ_u)
- $[y]_+ = \max(y, 0)$ determines the support of the distribution

Generalised Pareto Distribution (GP)

Fix $u = 3$, $\sigma_u = 1$ and vary ξ:



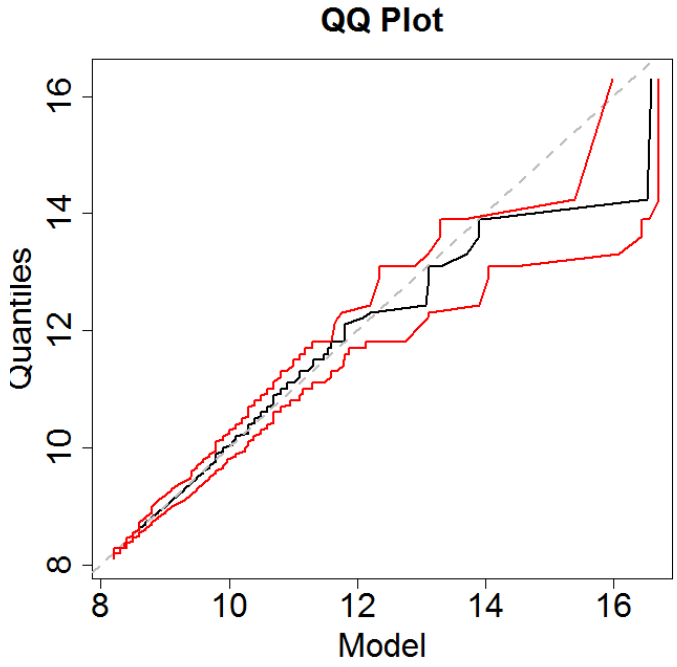
Statistical modelling

Threshold: $u = 8m$

Maximise the likelihood of the GPD distribution

$$L(.) \propto \prod_{i=1}^{n} \frac{1}{\sigma_u}\left[1+\xi\left(\frac{x_i-u}{\sigma_u}\right)\right]_+^{-1/\xi-1}$$

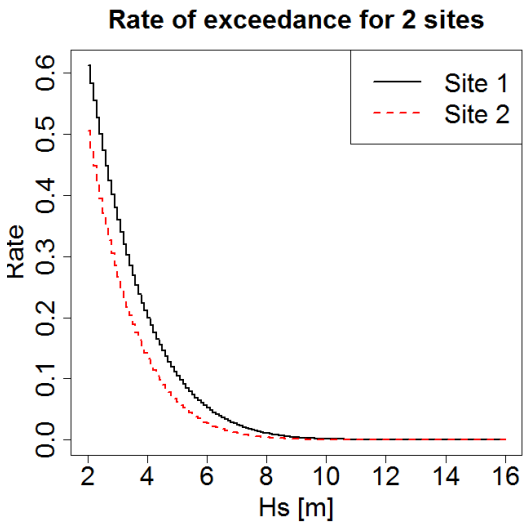
Parameter	MLE	SE
σ_u	1.56	0.12
ξ	-0.06	0.05



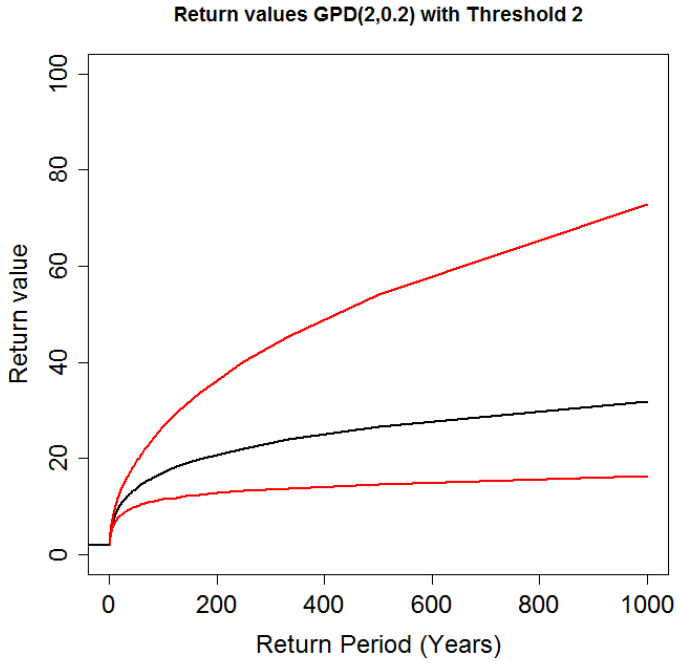
Model evaluation

3 hourly wave data

Rate of exceedance: λ_u



Return values:



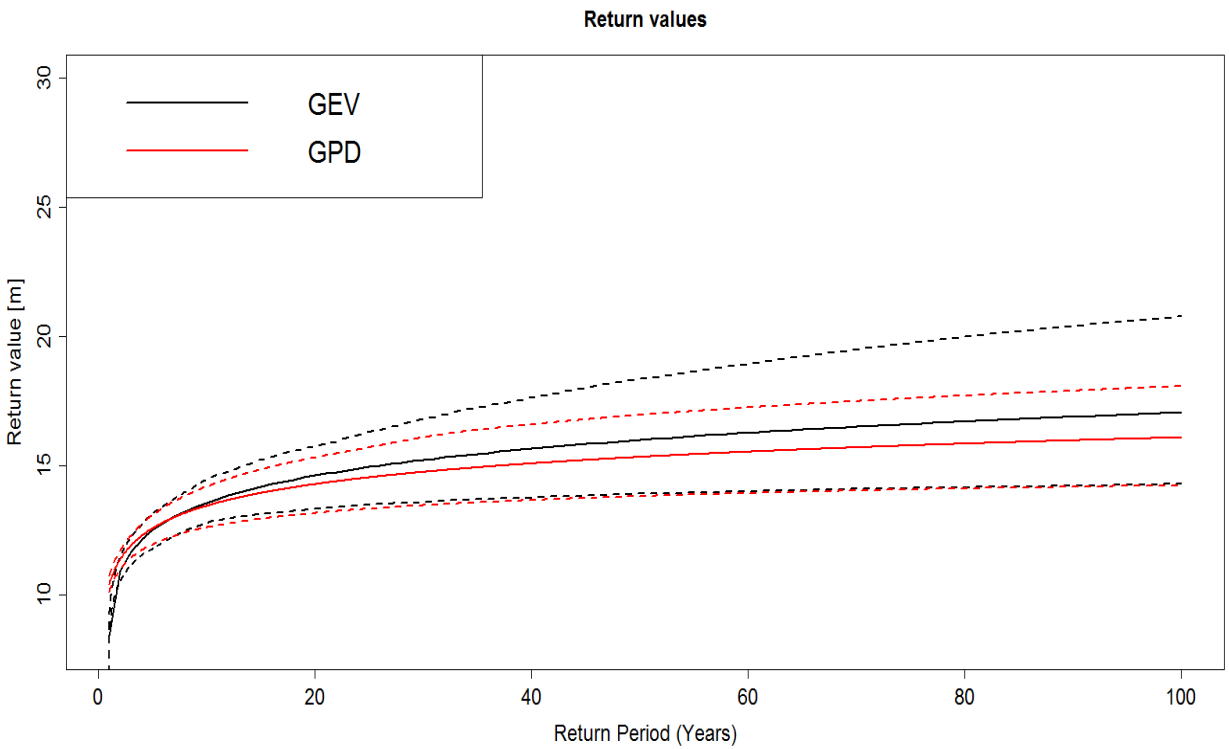
For site 1 with $u = 2m$:

$$rate = \begin{cases} 0.61 & \text{per obs.} \\ 1781 & \text{per year.} \end{cases}$$

Statistical Modelling

GEV vs GPD

Modelling excesses produces a larger data set, therefore more information about the extremes.



Leads to lower return values and greater certainty in estimates of return values.

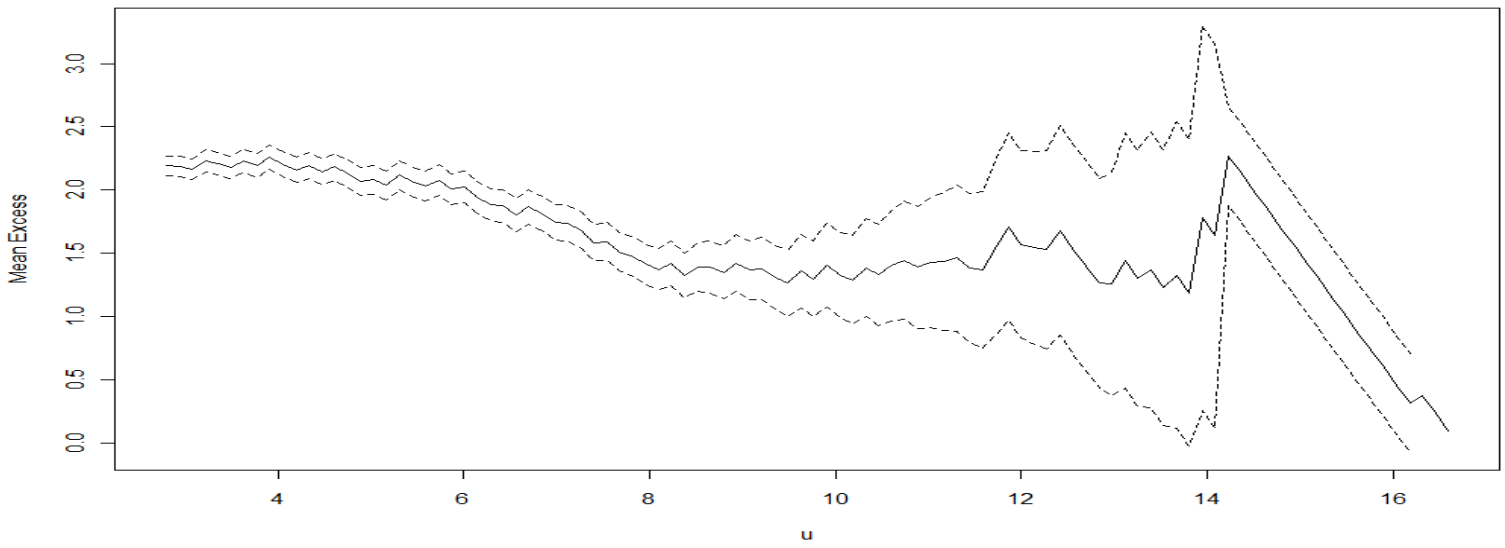
Choice of threshold

Mean Residual Life plot

If GP above u, then for v > u

$$\mathbb{E}[X - v \mid X > v] = \frac{\sigma_u + \xi(v-u)}{1-\xi}$$

Should be linear in v if u is large enough



Choice of threshold

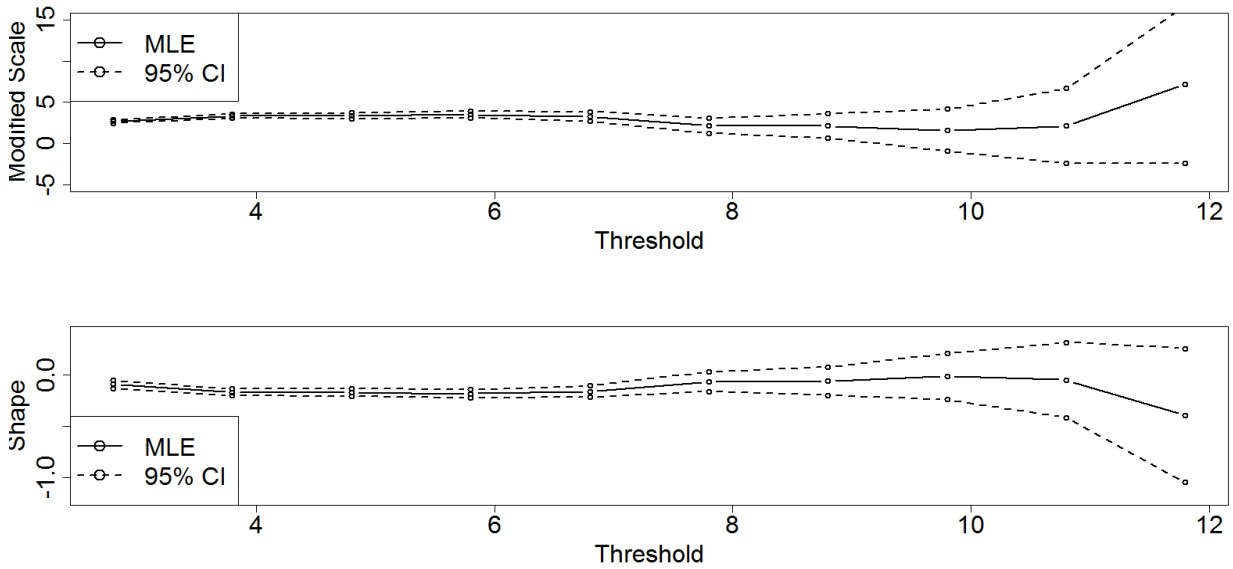
Parameter stability plots

$$X - u \mid X > u \;\sim\; GPD(\sigma_u, \xi) \text{ for some threshold } u$$

then for any higher threshold $v \geq u$

$$Y - v \mid Y > v \;\sim\; GPD(\sigma_u + \xi(v - u), \xi)$$

So check parameters are stable over threshold



Extensions to handle covariates

Incorporate covariates t into parameters

- $\sigma_u(t)$
- $\xi(t)$
- $\lambda_u(t)$
- Unlike for maxima distribution can change rapidly with covariates
- An extended theory for extremes using Poisson processes helps capture covariate changes in a location, scale, shape way - better for linear trend modelling

R Packages: library(ismev)

- In the R console type: install.packages(ismev)
- Select CRAN mirror UK (Bristol)

Block Maxima:

Function	Use
gev.fit	Fits the Generalised Extreme Value distribution
gev.diag	Determines the fit of the model

Threshold Exceedances:

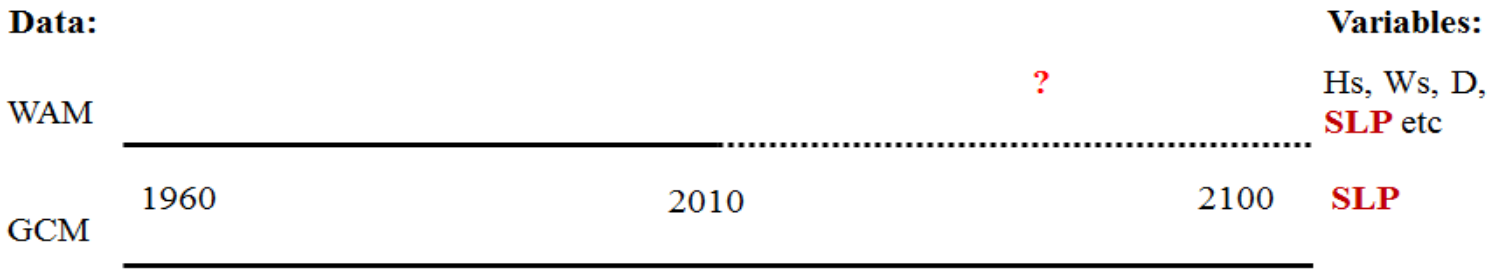
Function	Use
mrl.plot	threshold selection: mean residual life plot
gpd.fitrange	threshold selection: parameter stability plots
gpd.fit	Fits the Generalised Pareto distribution
gpd.diag	Determines the fit of the model

More Advanced Applications

- Predicting the distribution of future extreme wave heights
  - Surface Level Pressure and Wave Height
- Sea-level extremes
  - Tide and Surge
- The sinking of the MV Derbyshire
  - Wave impacts on a ship
- Assessing Spatial Flood Risk
  - River levels at different sites

Downscaling of Extreme Waves

Joint with: Shell



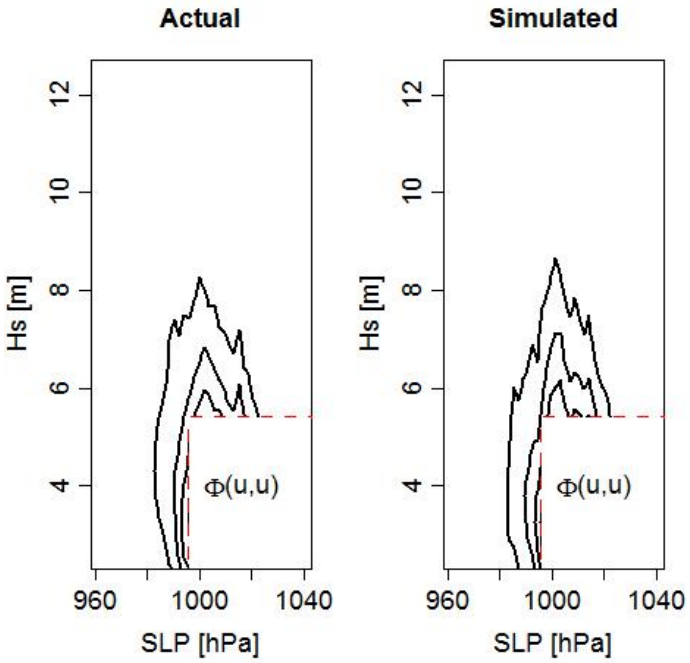
Want to determine the probability of observing a wave height above w in the future given what we know about SLP from the GCM:

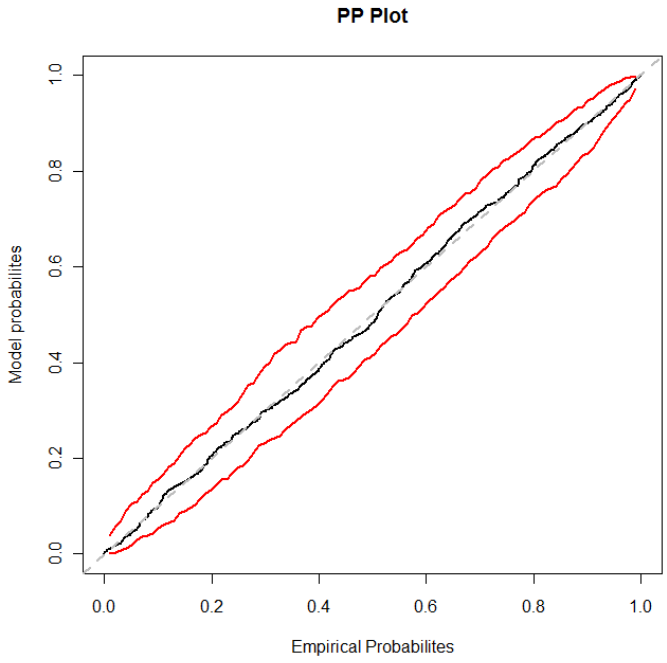
$$P(Hs > w) = \int P(Hs > w \mid SLP = s)\, f(s) ds$$

Downscaling of Extreme Waves

Tail model

- Fit a censored Bivariate Normal to the transformed data set
- Tail models: $Hs \sim GPD(\sigma_{u1}, \xi_1)$ and $SLP \sim GPD(\sigma_{u2}, \xi_2)$





Sea-level extremes

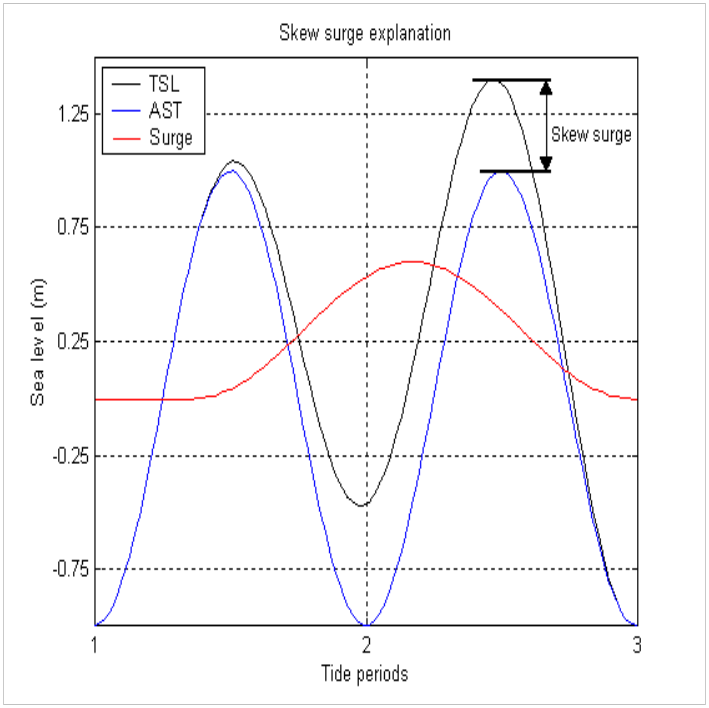
Joint with: National Oceanographic Centre, JBA Consulting, Haskoning

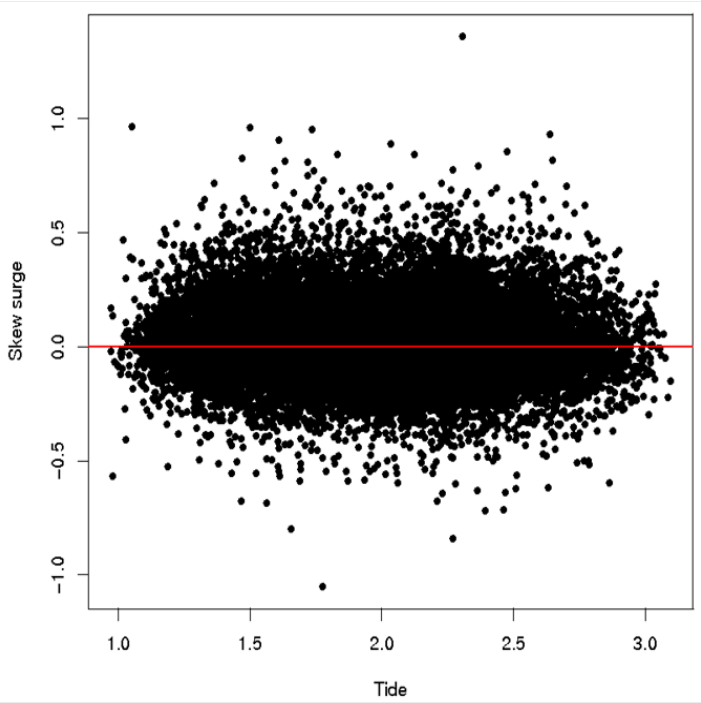
Sea-level = Tide + Surge

- Extreme value theory not appropriate directly for sea-levels but is fine for surges
- Complication: tides and surges interact
- Two approaches

1. model interaction by letting extreme surge parameters vary with tide
2. use skew surge

Interaction between Skew and Surge

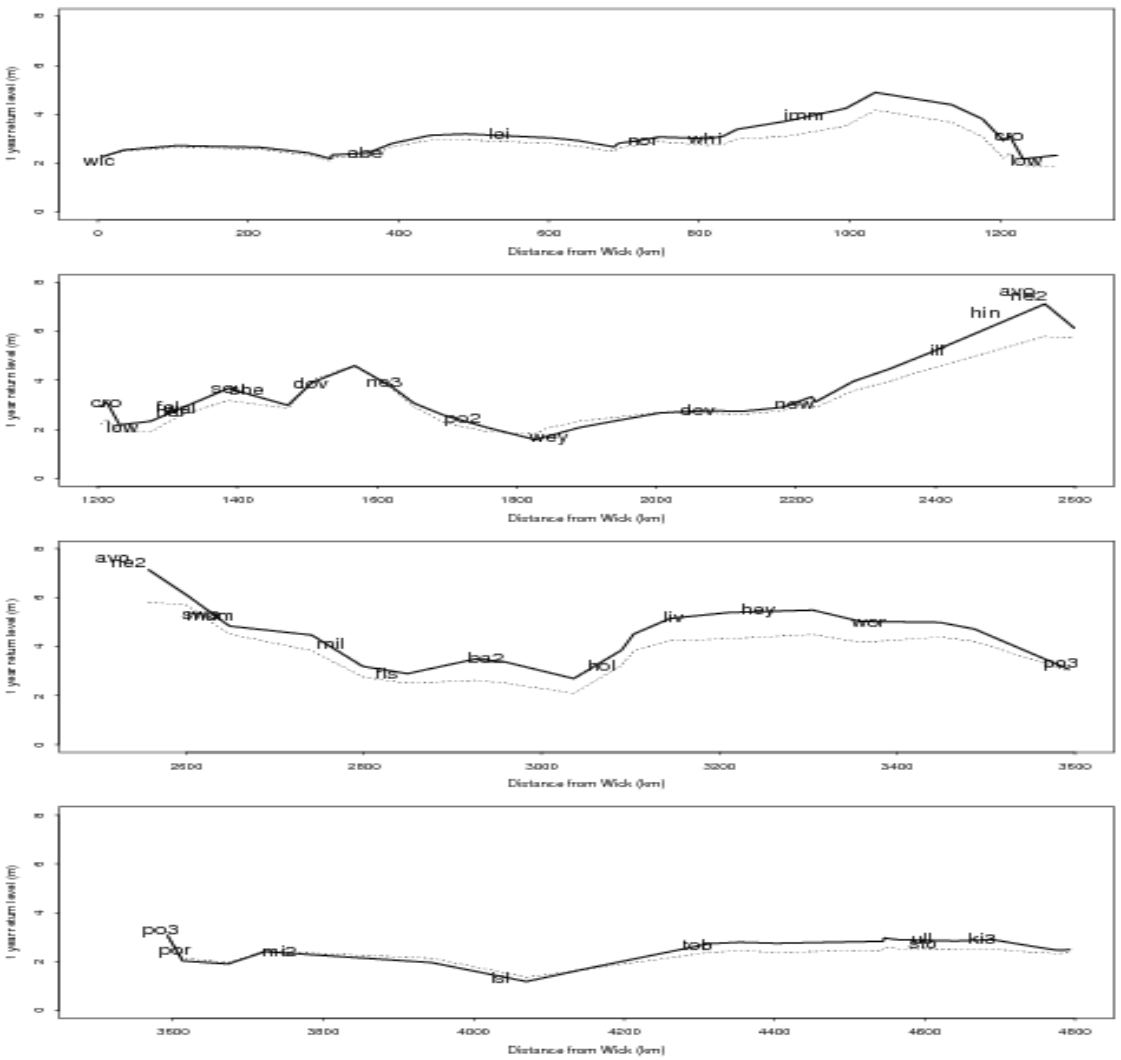




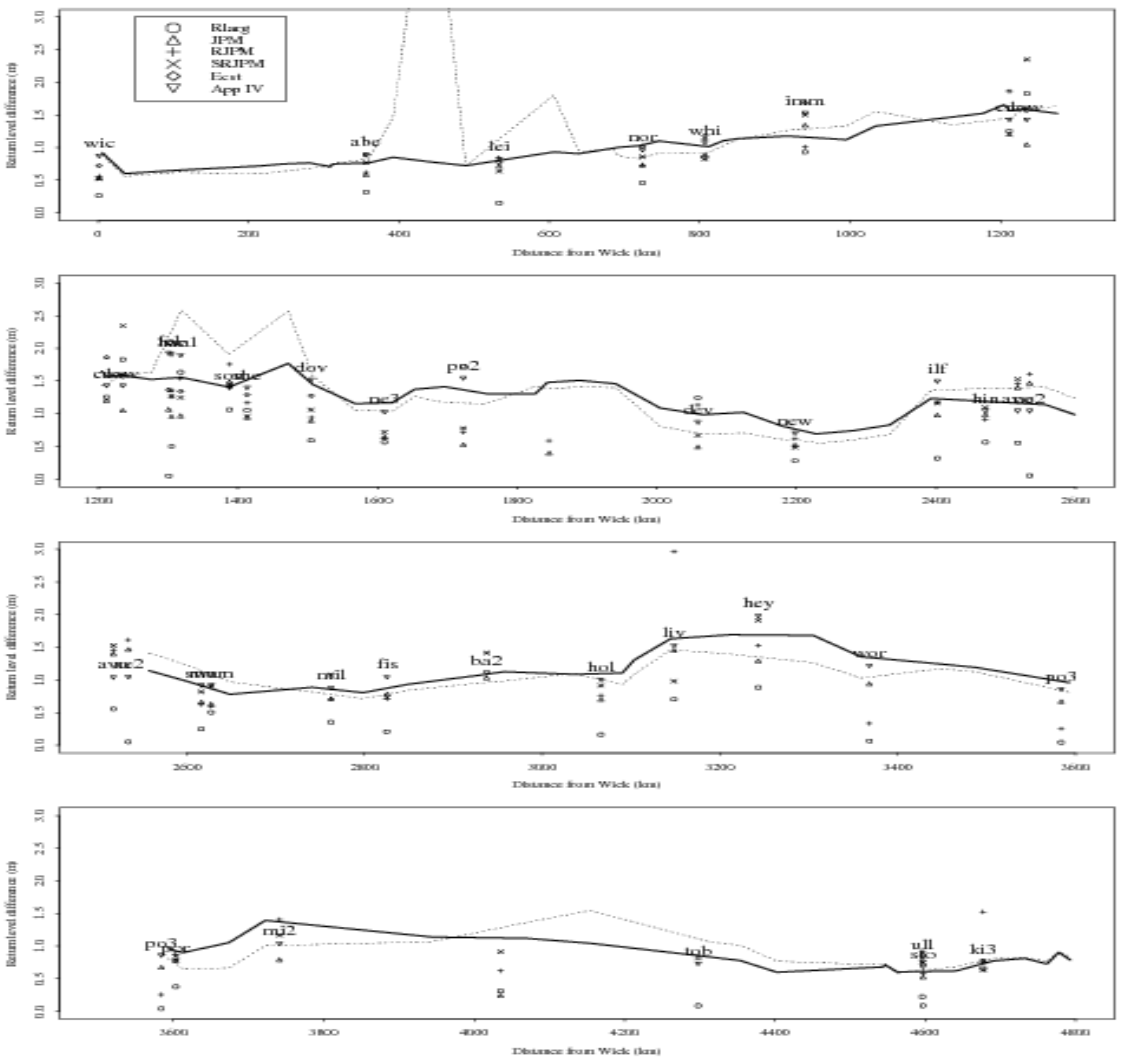
Joint Probability and Spatial Mapping

- Joint probability method combines extreme surge distribution and predicted tides
- Extreme surge distribution changes smoothly along coastline
- Tides predicted for any time and any place using numerical models (complex spatial variation)

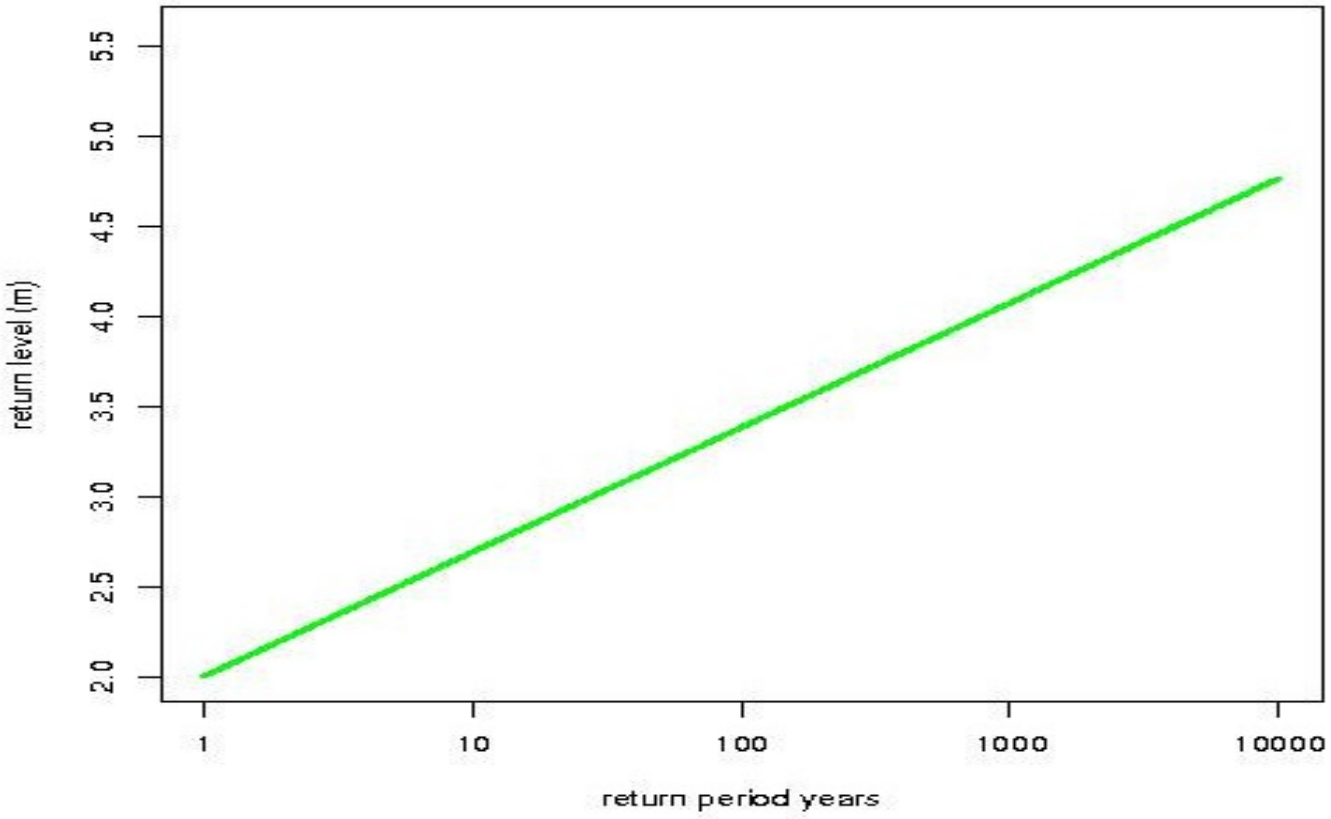
1 year return level map for UK



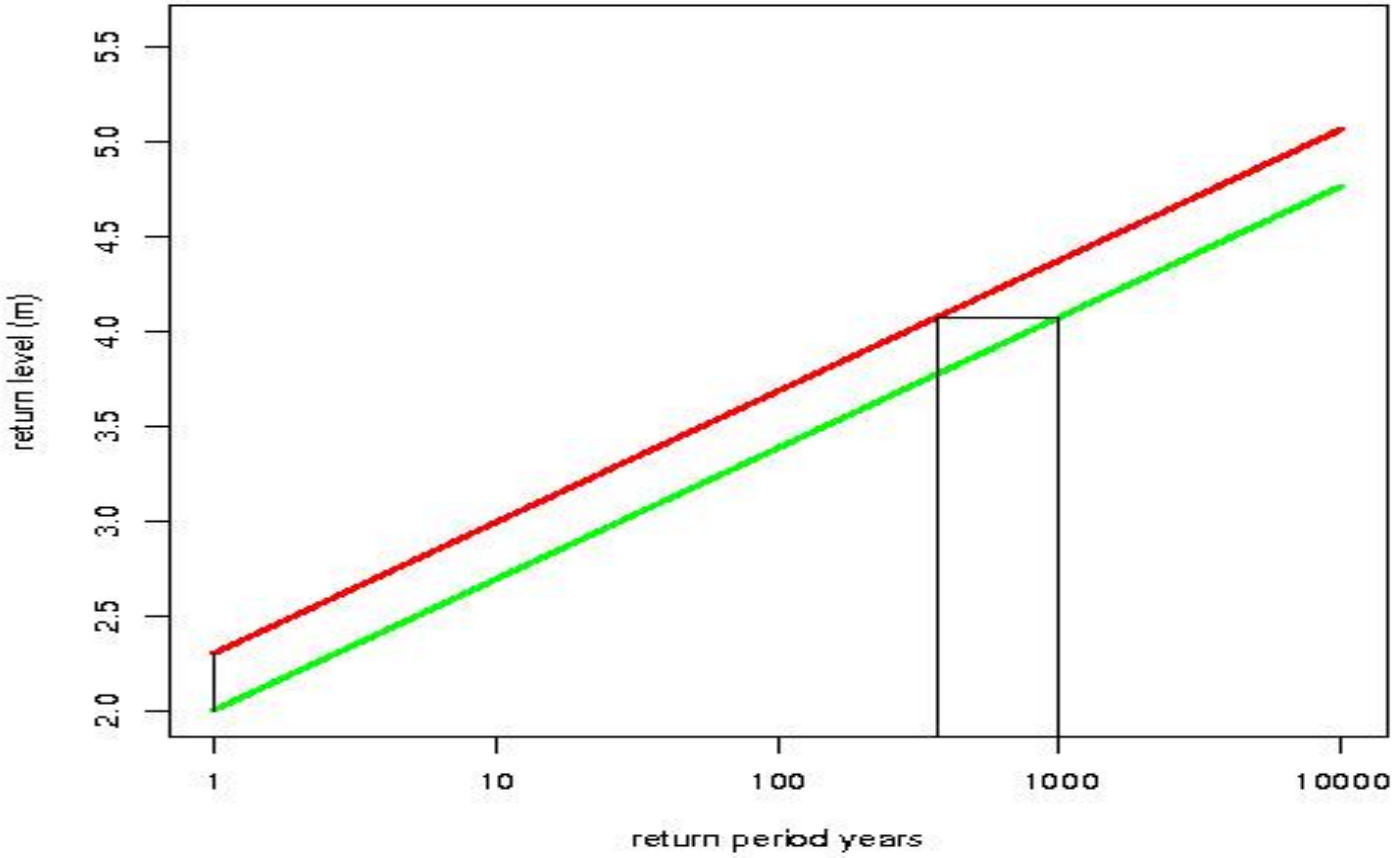
1000 year - 1 year return level map for UK



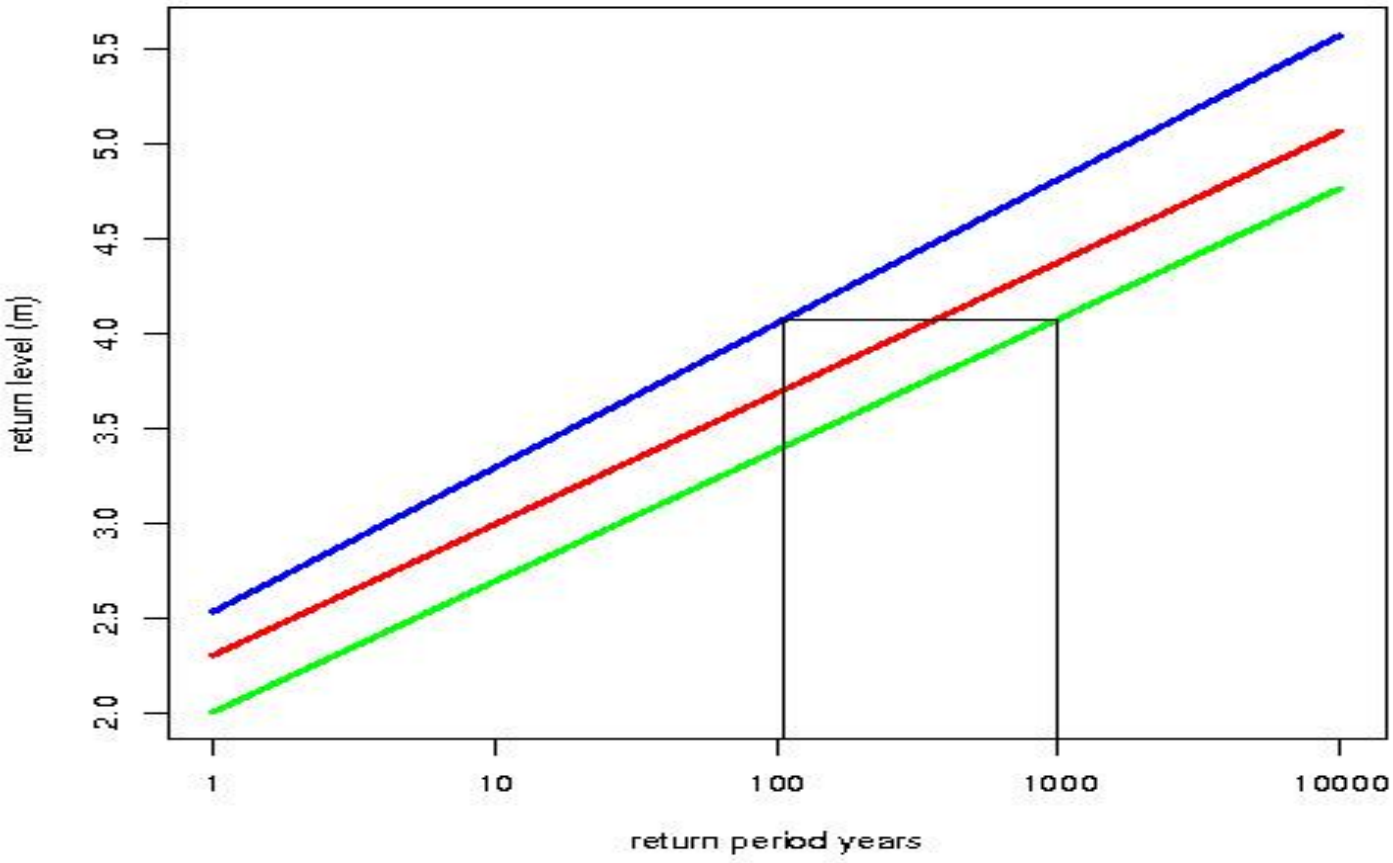
Impacts of climate change



Impacts of climate change



Impacts of climate change

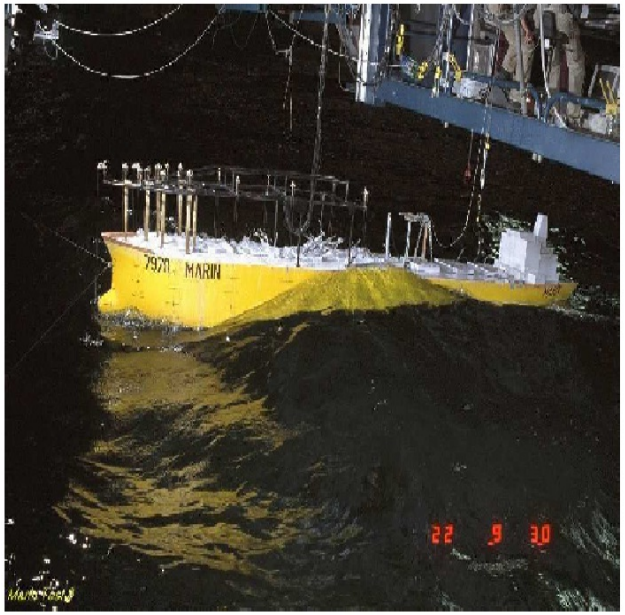


The sinking of the MV Derbyshire

- 1980- Sunk in Typhoon Orchid off Japan
- All 44 onboard died
- No mayday
- Largest UK shipping loss
- 1997 - Wreckage photographed
- 2000 - High Court Investigation

Background

- Design standard for front hatch cover is $42kPa$
- $P\,(\text{wave impact} > 42\text{kPa} \mid \text{how ship run})$
- Hindcast waves over typhoon
- Wave tank studies over conditions
- Model wave impacts over typhoon via linking extreme value parameters to conditions



Estimated Probabilities

Conditions		Model		
Ship	Waves	Distribution 1		
Intact	Hindcast	0.73		
Intact	10% higher	1.00		
Damaged Stores	Hindcast	0.90		
Damage Deep Tank	Hindcast	0.82		
Damaged Ballast Tank	Hindcast	1.00		

Estimated Probabilities

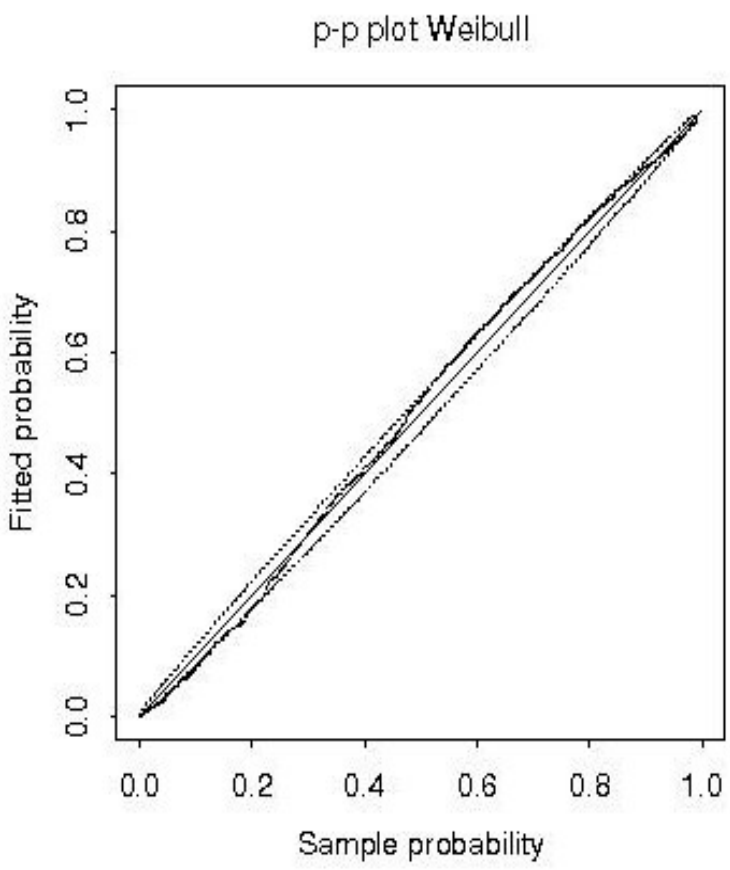
Conditions		Model		
Ship	Waves	Distribution 1	Distribution 2	
Intact	Hindcast	0.73	<0.01	
Intact	10% higher	1.00	0.12	
Damaged Stores	Hindcast	0.90	<0.01	
Damage Deep Tank	Hindcast	0.82	<0.01	
Damaged Ballast Tank	Hindcast	1.00	0.30	

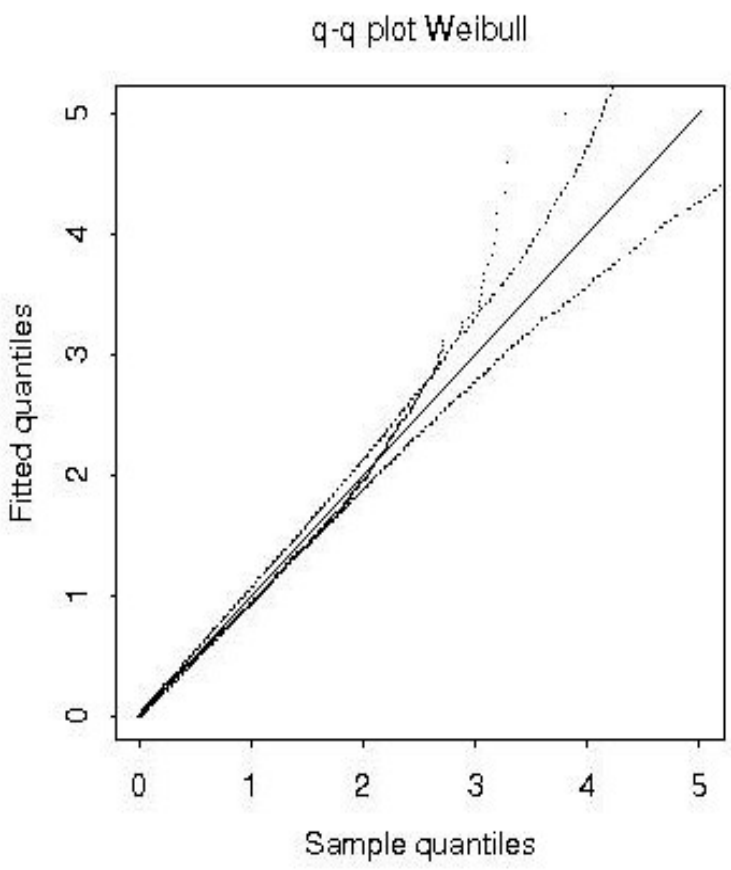
Estimated Probabilities and 95% Confidence Intervals

Conditions		Model		
Ship	Waves	Distribution 1	Distribution 2	Extreme Value
Intact	Hindcast	0.73	<0.01	0.00 (0.00,0.00)
Intact	10% higher	1.00	0.12	0.01 (0.00,0.29)
Damaged Stores	Hindcast	0.90	<0.01	0.00 (0.00,0.05)
Damage Deep Tank	Hindcast	0.82	<0.01	0.00 (0.00,0.00)
Damaged Ballast Tank	Hindcast	1.00	0.30	0.31 (0.08,0.73)

How do we know which is best?

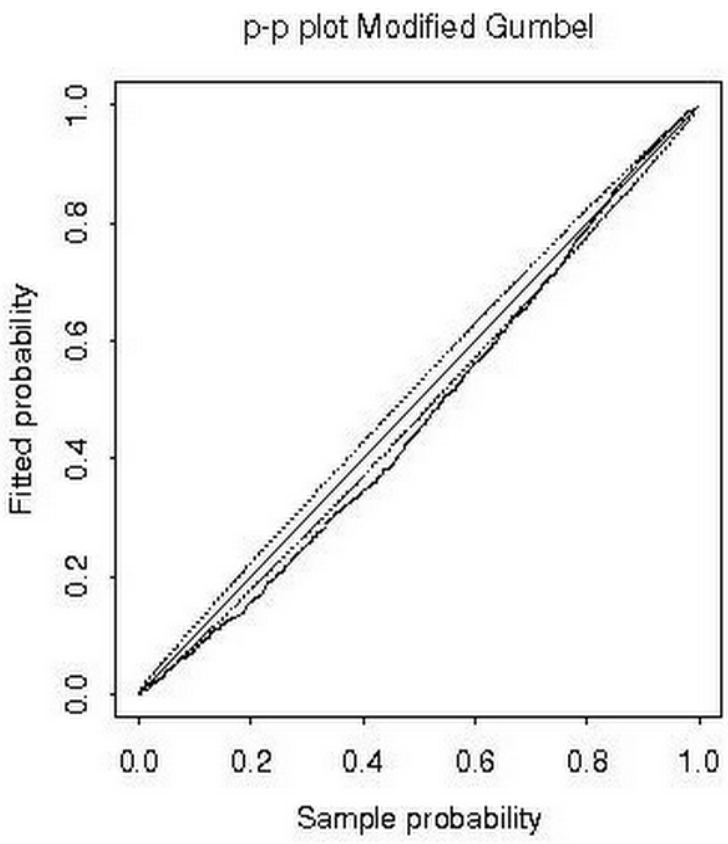
Distribution 1

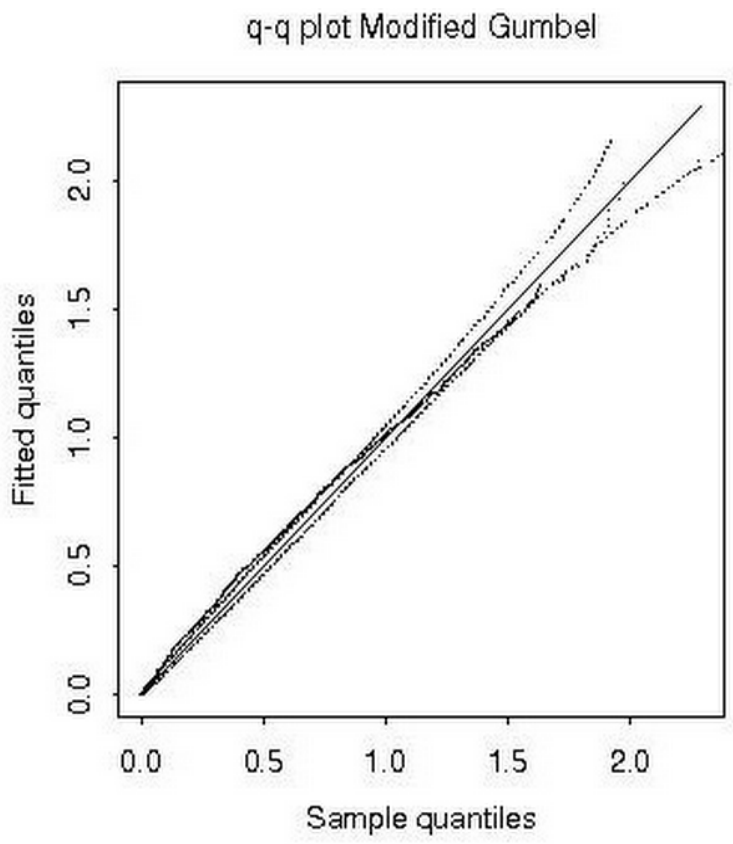




How do we know which is best?

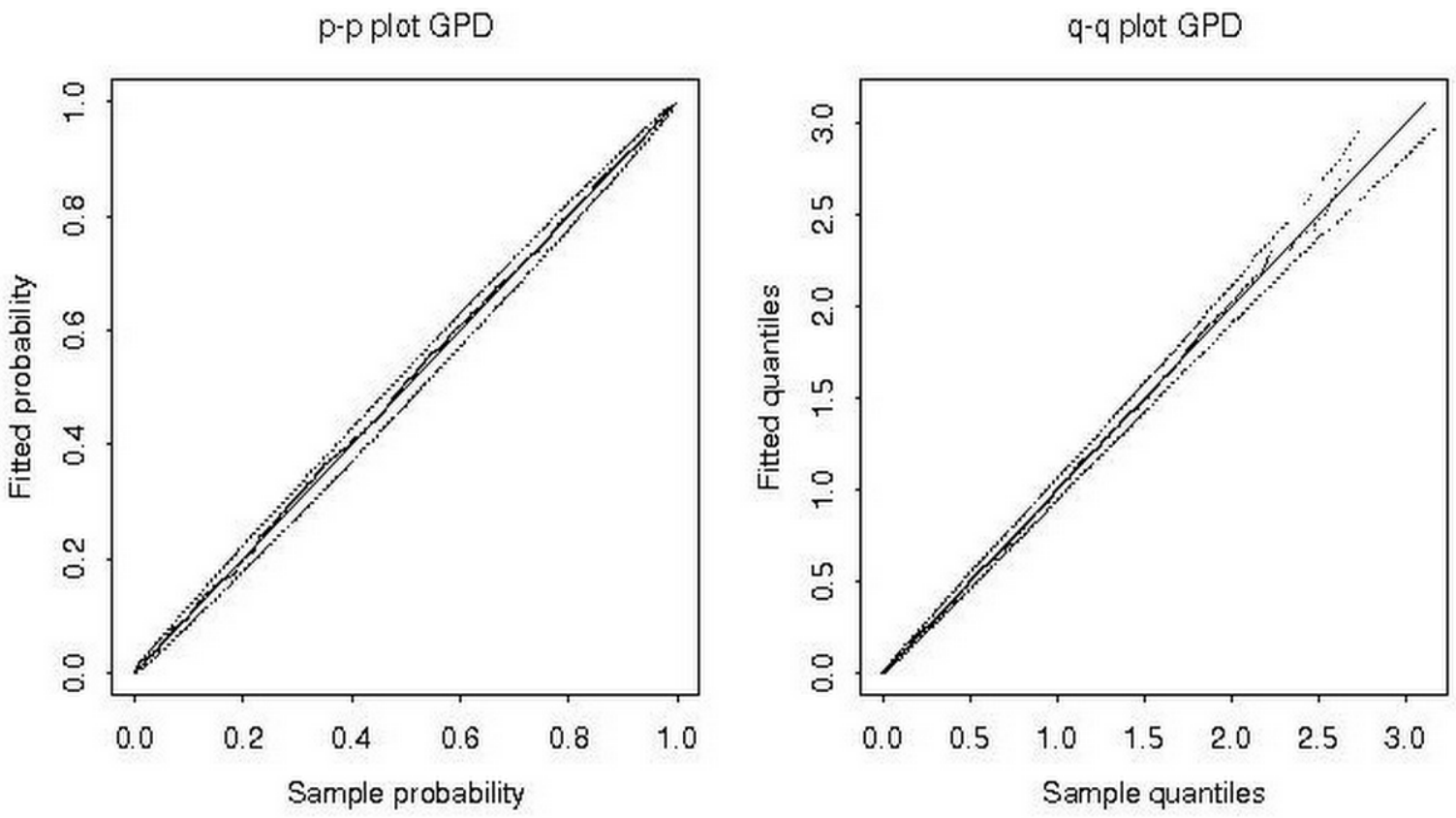
Distribution 2





How do we know which is best?

Distribution 3

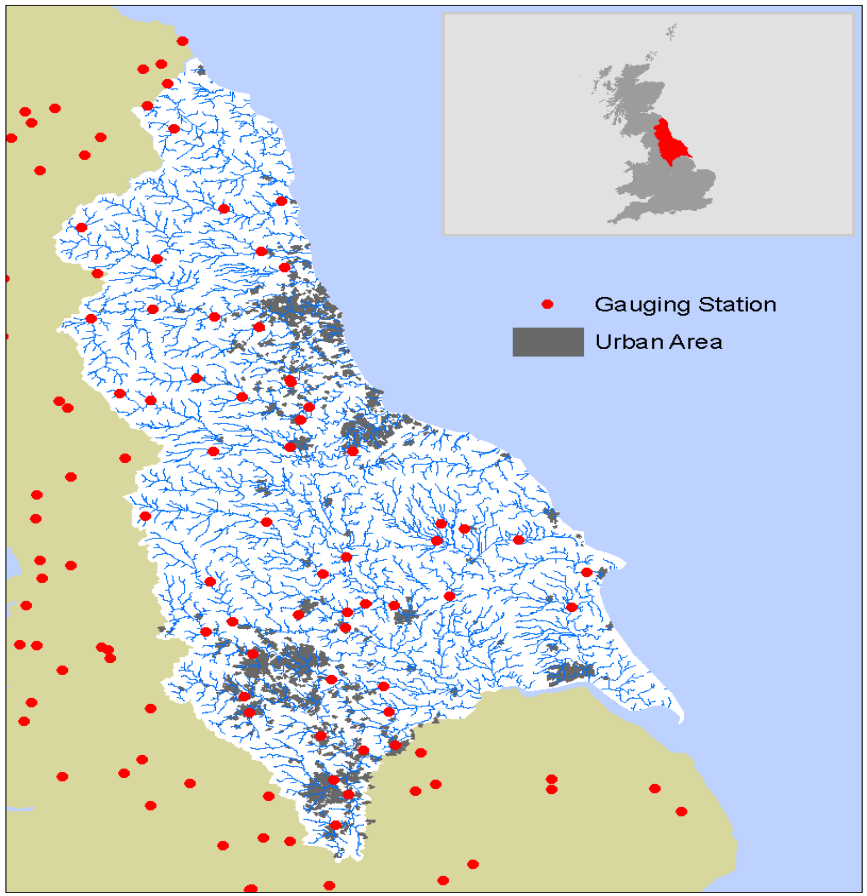


Spatial flood risk

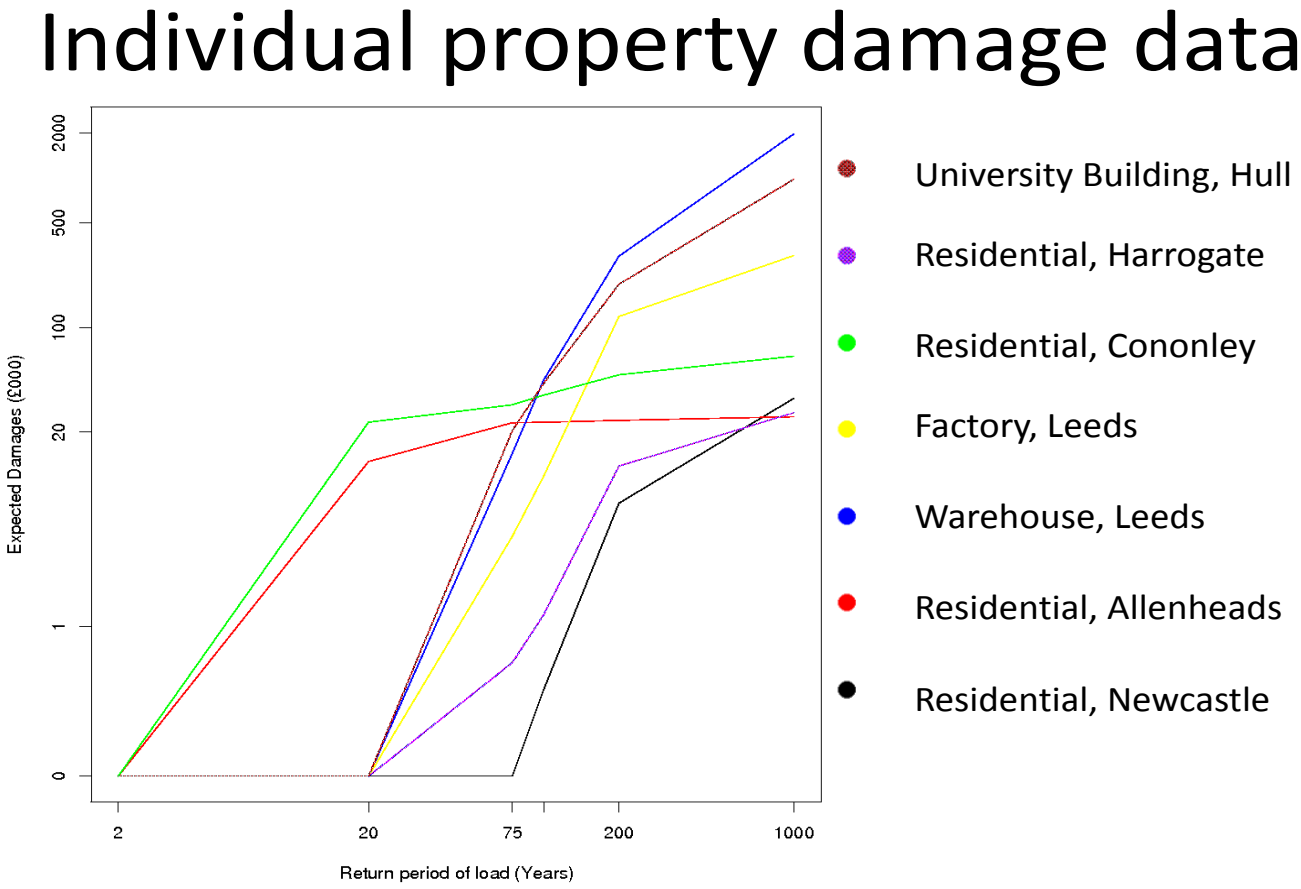
Joint work with Environment Agency and JBA Consulting

- Knowledge of the likelihood of extremes at a site is important for design of flood prevention schemes
- Spatial impact of an extreme event is important for insurance industry and emergency planning

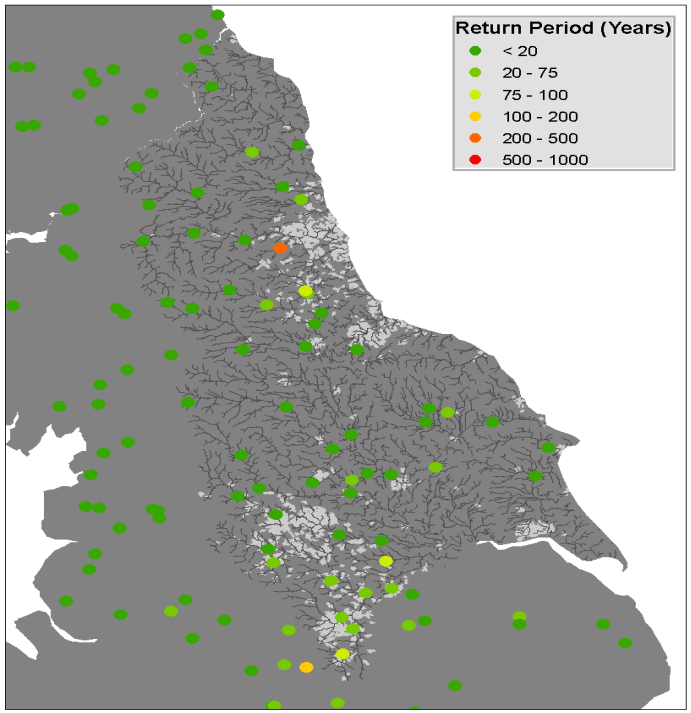
North East Region Fluvial Extremes

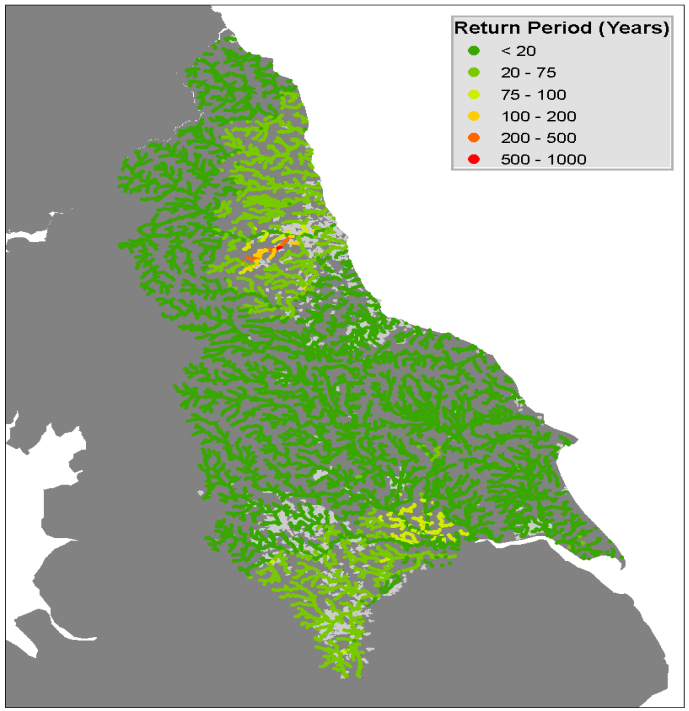


Individual Property Data

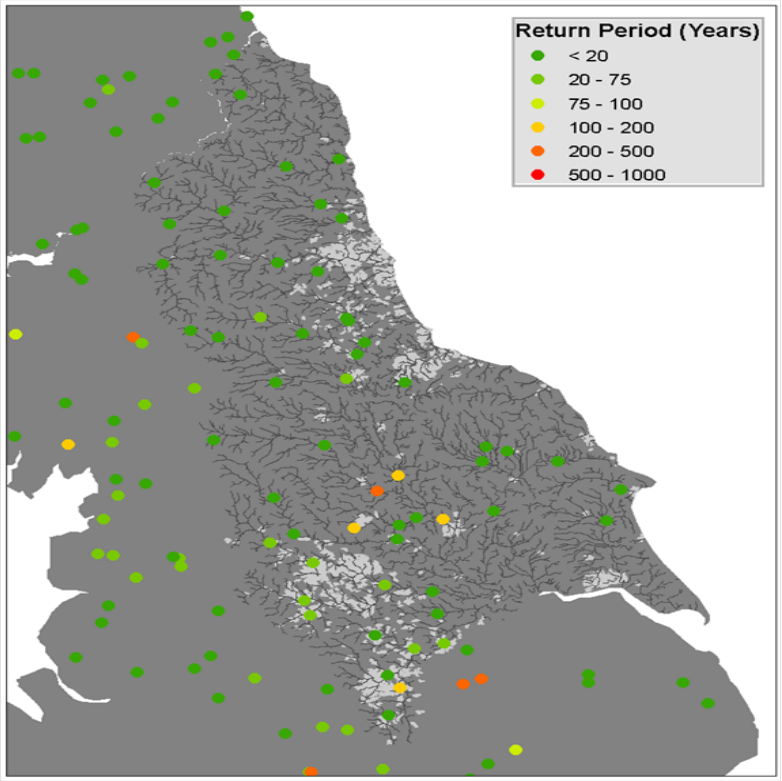


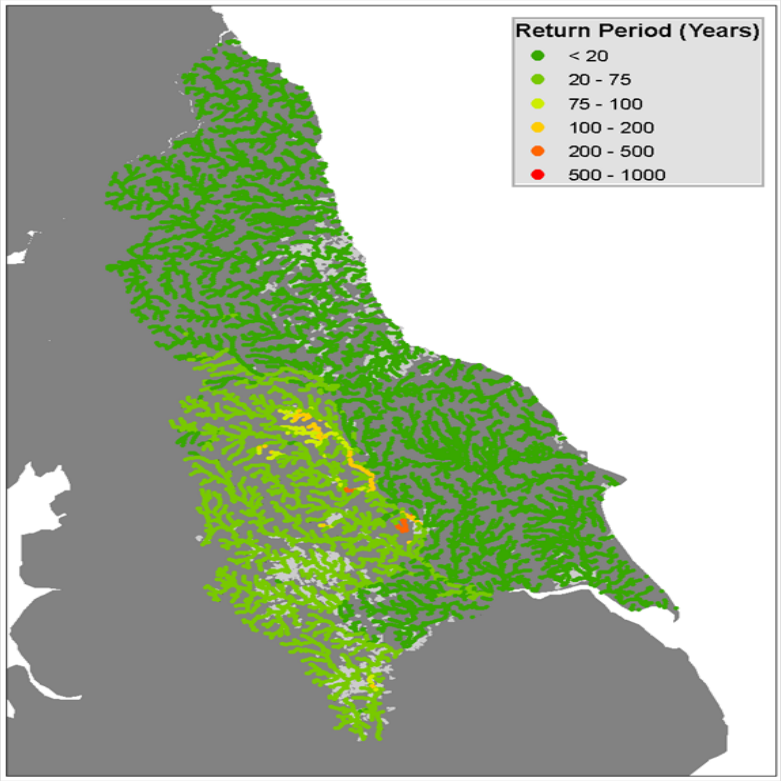
Worst Observed Loss Event: Oct 2000



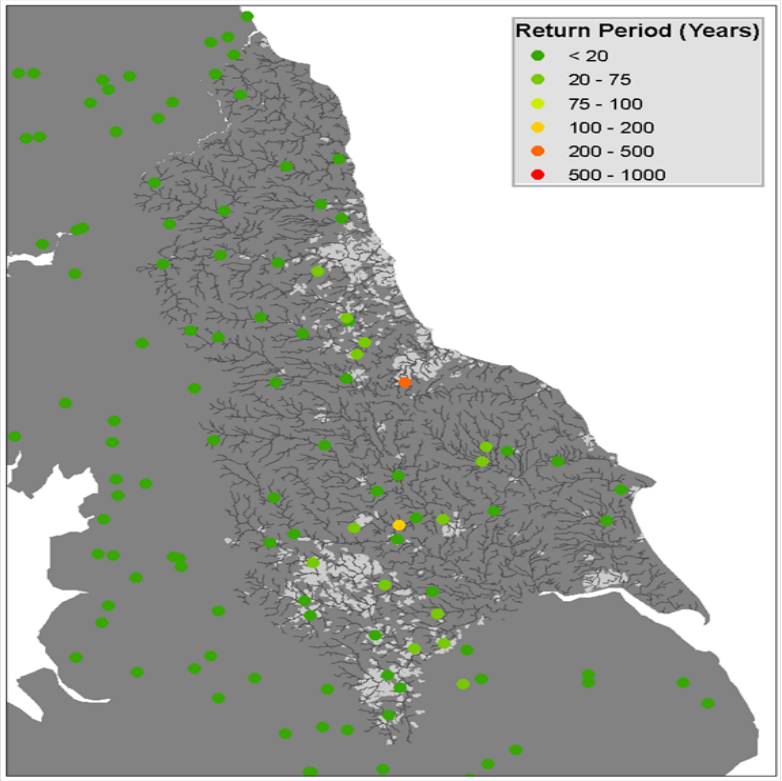


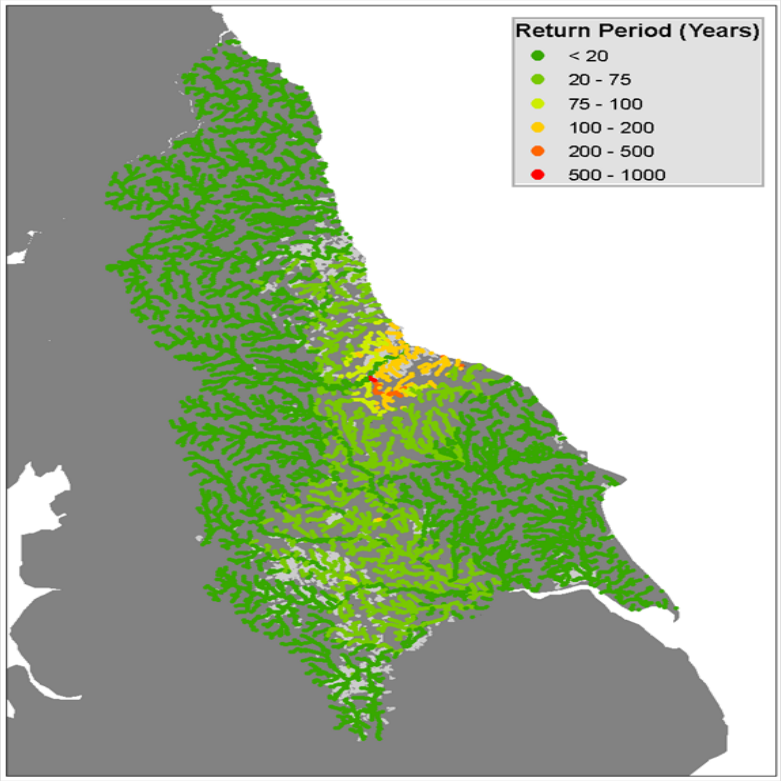
Different Simulated 100 Year Loss Events



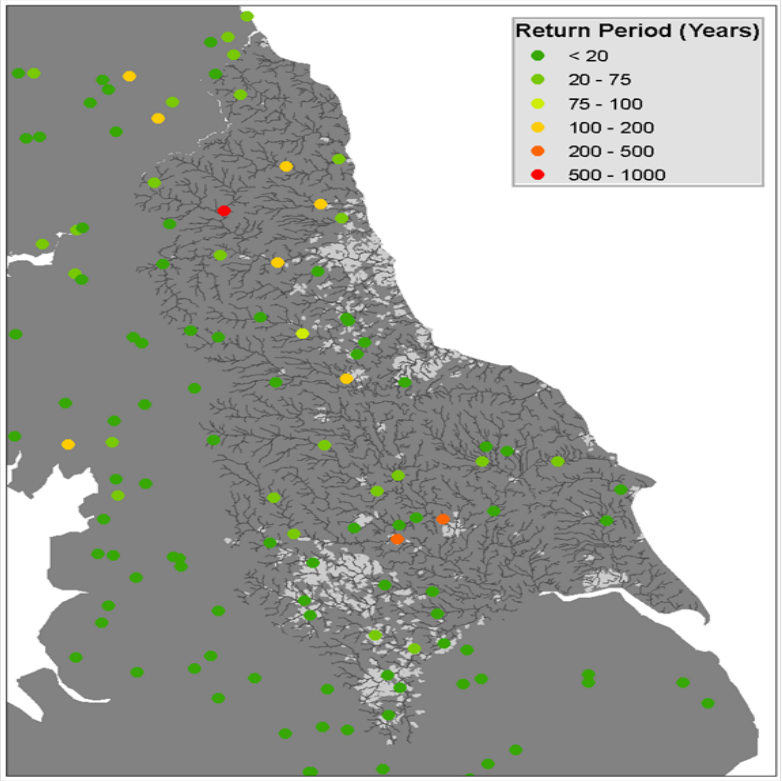


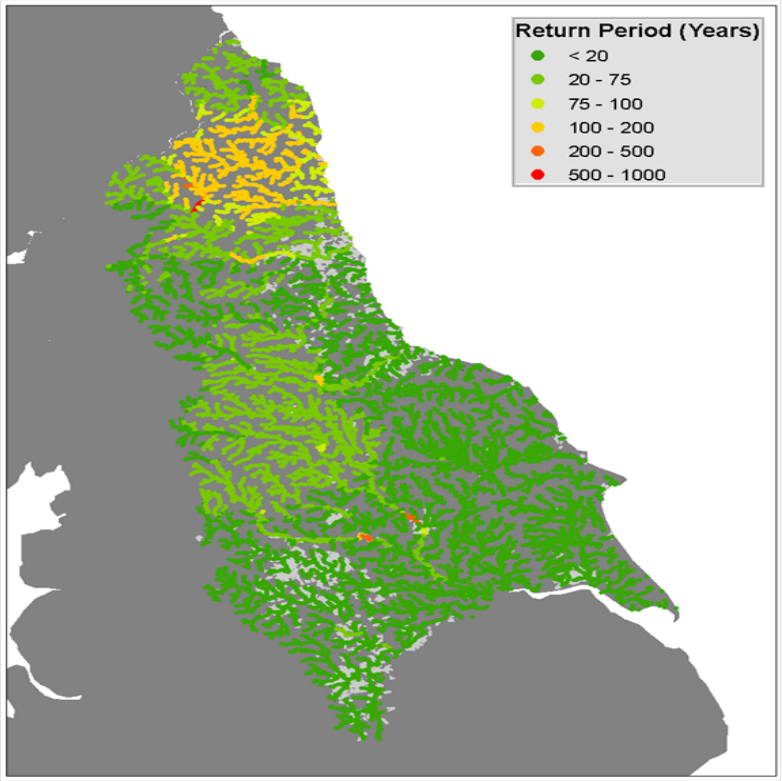
Different Simulated 100 Year Loss Events



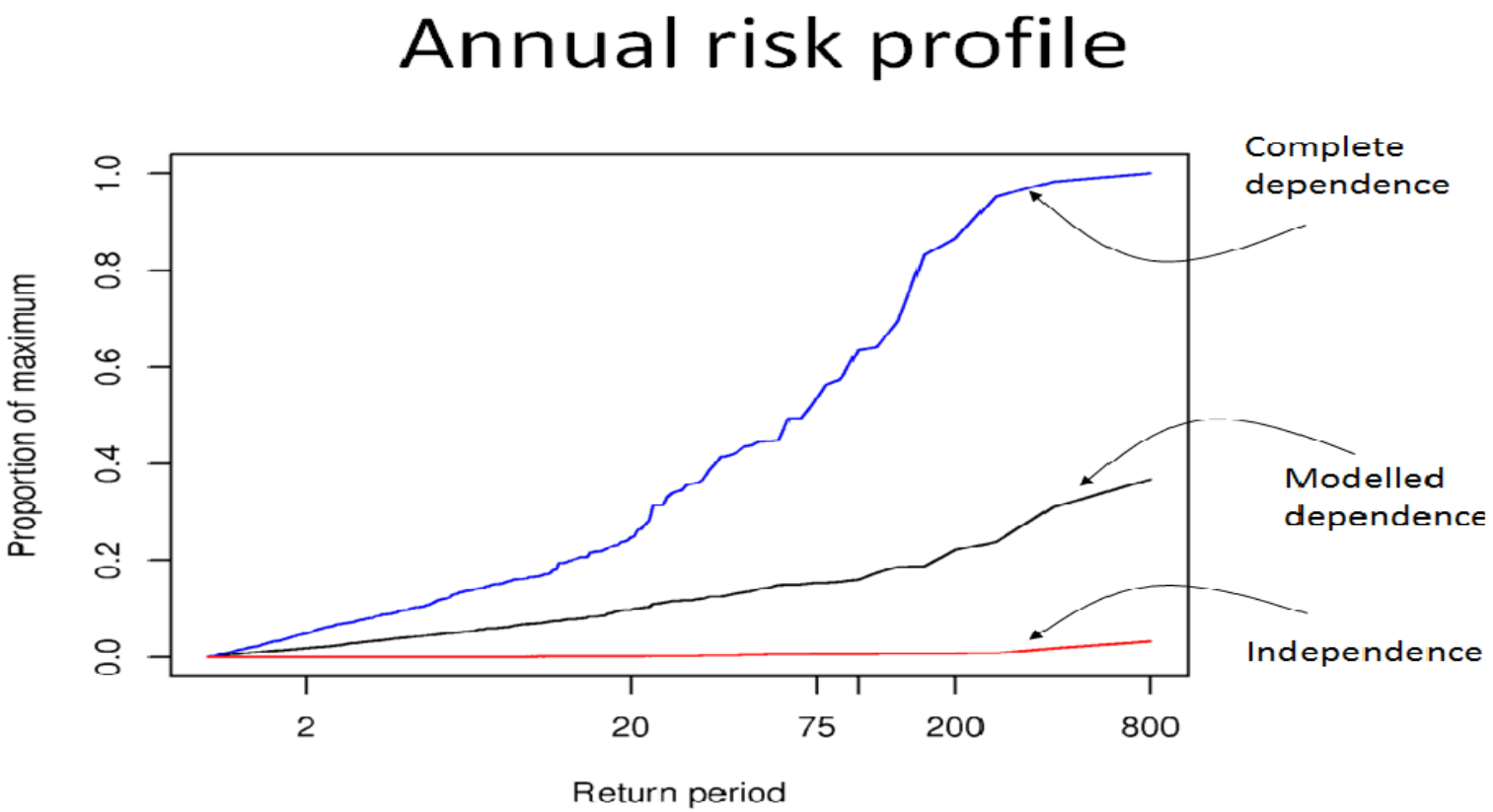


Different Simulated 100 Year Loss Events





Estimated Annual Loss Distribution



Summary

- Extreme value methods are based on a rigorous probability theory which based on asymptotic arguments
- The limit theory suggests models which may be of practical value
- Limited data means efficient use (pooling) of any information about extremes is highly beneficial
- Statistical methods only work in advanced applications if they are fully integrated with physical knowledge of the application
- Collaboration is the key