Australian Centre of Excellence for Mathematical and Statistical Frontiers in Big Data, Big Models, New Insights

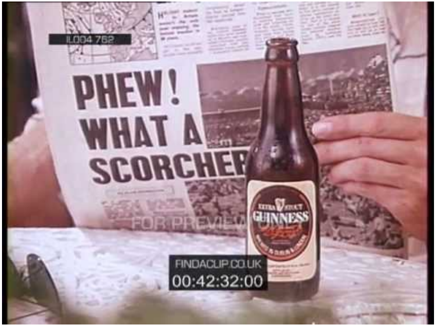
Phew what a scorcher!
A short introduction to modelling extremes

Scott A. Sisson, Boris Beranger, Seth Westra, Tom Whitaker

UNSW Sydney & ACEMS

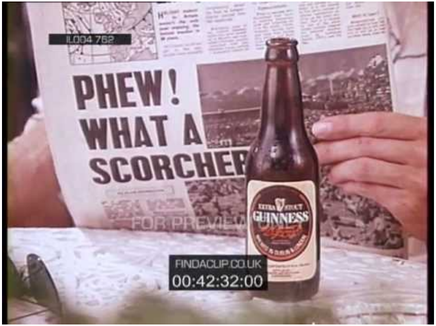
Bayes on the Beach 2017, Nov 13-15 2017, Gold Coast

Talk Outline



1. Univariate extremes
   - Block maxima results
   - Block maxima practice
   - Threshold models
2. Multivariate extremes
3. Spatial extremes
4. Extras
   - Stereological extremes (ABC)
   - Big data extremes (SDA)

Talk Outline



1. Univariate extremes
   - Block maxima results
   - Block maxima practice
   - Threshold models
2. Multivariate extremes
3. Spatial extremes
4. Extras
   - Stereological extremes (ABC)
   - Big data extremes (SDA)

Univariate extremes – block maxima

Central limit theorem:

- Suppose $X_1, X_2, \ldots \sim F(\mu, \sigma^2)$ (i.i.d.)
- $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$
- Then as $n \to \infty$ the distribution of $\bar{X}_n$ is $N(\mu, \frac{\sigma^2}{n})$
- Useful for modelling means of processes
- Asymptotic distribution depends on μ and σ^2 but not on form of F.

Distribution of block maxima

- Suppose $X_1, X_2, \ldots \sim F$ (i.i.d.)
- $M_n = \max\{X_1 + \ldots + X_n\}$
- What is the distribution of M_n?
  - For finite n?
  - For $n \to \infty$?
- Useful for modelling extreme of processes

Univariate extremes – block maxima

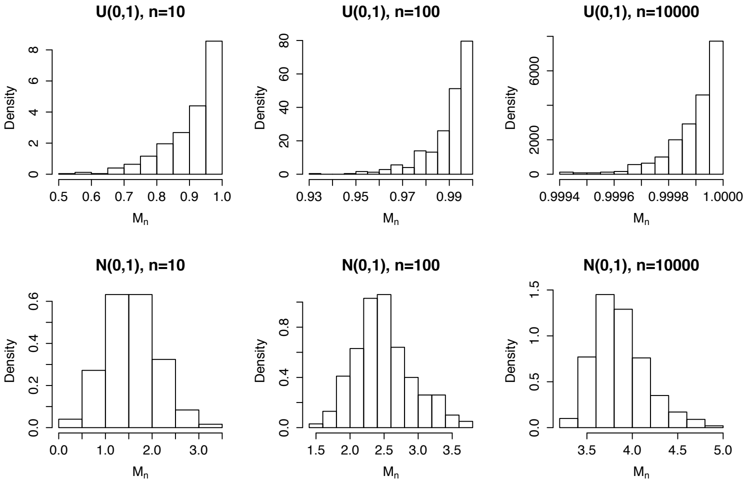
What is the distribution of M_n – for finite n? (1st year exercise!)

$$\begin{aligned} F(M_n) = Pr(M_n \leq x) &= Pr(\max\{X_1 + \ldots + X_n\} \leq x) \\ &= Pr(X_1 \leq x, \ldots, X_n \leq x) \\ &= Pr(X_1 \leq x) \ldots Pr(X_n \leq x) \\ &= Pr(X \leq x)^n = F^n(x). \end{aligned}$$

- So if we know F, we can model M_n.
- But what if F is unknown?
- For finite n we are stuck.
- But CLT works for unknown F as $n \to \infty$
- Can we do something similar for M_n as $n \to \infty$? (Answer = yes!)

Univariate extremes – block maxima

- Set $x_0 = \sup\{x : F(x) < 1\}$
- Then $Pr(M_n \leq x_0) = F^n(x_0) \to 0$ as $n \to \infty$
- I.e. distribution of M_n degenerates to point mass at x_0



Univariate extremes – block maxima

Central Limit Theorem

- Consider limiting distribution of

$$\frac{\bar{X} - b_n}{a_n} \to N(0,1) \qquad (X_i \sim F(\mu, \sigma^2))$$

- Limit distribution exists if: $b_n = \mu$ and $a_n = \sigma/\sqrt{n}$.

Extremes

- Consider limiting distribution of rescaled sample maxima

$$M_n^* = \frac{M_n - b_n}{a_n} \to \text{? as } n \to \infty.$$

- i.e. require non-degenerate distribution G and sequences $a_n > 0, b_n$ such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = Pr(M_n \leq a_n x + b_n) = F^n(a_n x + b_n) \to G(x)$$

converges in distribution to G as $n \to \infty$.

- If it converges then F is in the domain of attraction of G, $F \in D(G)$.

Univariate extremes – block maxima

Extremal Types Theorem

If sequences $a_n > 0$ and b_n exist such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \qquad \text{as } n \to \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

1. Gumbel:
$$G(z) = \exp\{-\exp[-(\frac{z-b}{a})]\} \quad -\infty \leq z \leq \infty$$

2. Fréchet:
$$G(z) = \begin{cases} 0 & z \leq b \\ \exp\{-(\frac{z-b}{a})^{-\alpha}\} & z > b \end{cases}$$

3. Weibull:
$$G(z) = \begin{cases} \exp\{-[-(\frac{z-b}{a})^{\alpha}]\} & z < b \\ 1 & z \geq b \end{cases}$$

The precise limiting distribution depends on tail behaviour of F.

Univariate extremes – block maxima

Generalised extreme value (GEV) distribution

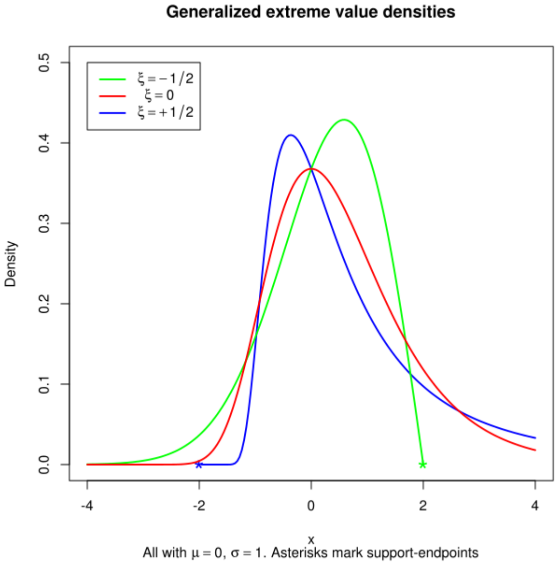
$$G(z) = \exp\left\{ -\left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}$$

where $(a)_+ = \max(0, a)$.

- μ, σ, ξ are location, scale, shape parameters
- ξ determines sub-family:
  - $\xi \to 0$: Gumbel
  - $\xi > 0$: Fréchet
  - $\xi < 0$: Weibull

GEV is practically useful for modelling (see later).

Univariate extremes – block maxima



Outline proof of Extremal Types Theorem is fairly straightforward ⇒

Univariate extremes – block maxima

Definition: Max-stability

A distribution G is max-stable if, for $n = 2, 3, \ldots$, there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z).$$

- I.e. The maximum of samples drawn from a max-stable distribution come from the same distribution, but with changed location and scale parameters.
- Compare: mean-stable distributions (e.g. Gaussian).

Theorem

A distribution is max-stable iff it is a GEV distribution.

- Easy to check that Gumbel, Fréchet, Weibull are all max-stable.
- Converse is harder (but true).

Univariate extremes – block maxima

Outline Proof of Extremal Types Theorem:

- Suppose that for large n
$$Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \approx G(z).$$
- Hence for any integer $k \in \{1, 2, \ldots\}$, since nk is large
$$Pr\left(\frac{M_{nk} - b_{nk}}{a_{nk}} \leq z\right) \approx G(z). \tag{1}$$
- M_{nk} is max of k variables, each having the same distribution as M_n.
-
$$Pr\left(\frac{M_{nk} - b_n}{a_n} \leq z\right) = \left[Pr\left(\frac{M_n - b_n}{a_n} \leq z\right)\right]^k. \tag{2}$$
- From (1) and (2) respectively we have
$$Pr(M_{nk} \leq z) \approx G\left(\frac{z - b_{nk}}{a_{nk}}\right) \text{ and } Pr(M_{nk} \leq z) \approx G^k\left(\frac{z - b_n}{a_n}\right).$$
- G and G^k are identical, apart from location/scale parameters.
Hence G is max-stable, and therefore GEV.

Univariate extremes – block maxima

Simple examples:

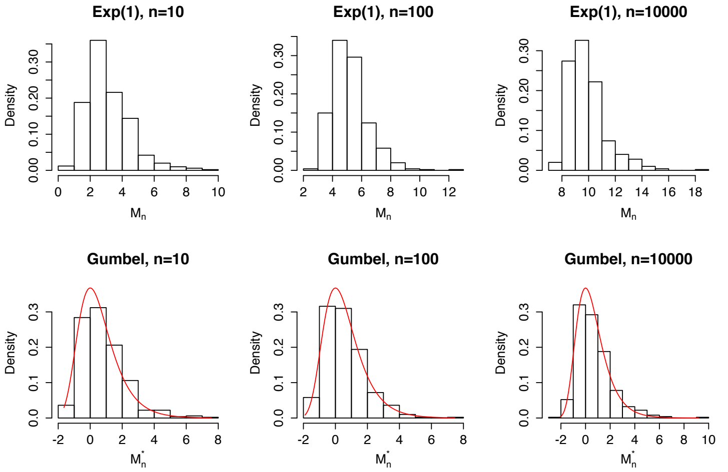
#1 Exponential(1): $F(x) = 1 - e^{-x}$ for $x > 0$.

- Let $a_n = 1$ and $b_n = \log n$.
- Then

$$
\begin{aligned}
Pr\left(\frac{M_n - b_n}{a_n} \le z\right) &= F^n(z + \log n) \\
&= (1 - e^{-(z+\log n)})^n \\
&= (1 - e^{-z}/n)^n \\
&\to \exp(-e^{-z}) \qquad \text{as } n \to \infty.
\end{aligned}
$$

which is Gumbel, with $\xi = 0$ in the GEV family.

Univariate extremes – block maxima



Univariate extremes – block maxima

Simple examples:

#2 Fréchet(1): $F(x) = \exp(-1/x)$ for $x > 0$.

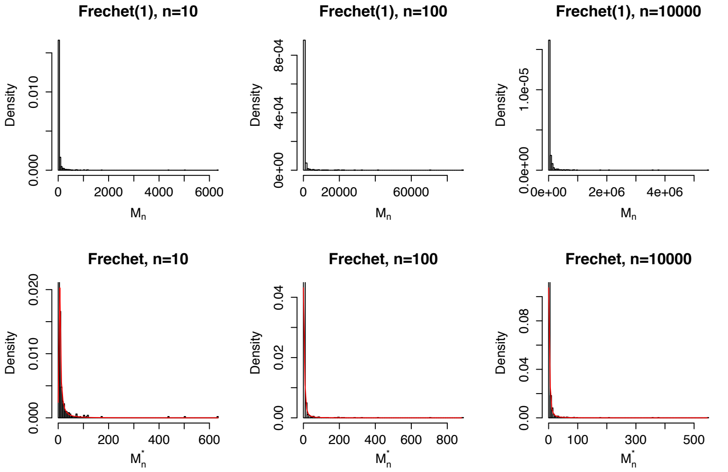
- Let $a_n = n$ and $b_n = 0$.
- Then

$$
\begin{aligned}
Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) &= F^n(nz) \\
&= [\exp(-1/(nz))]^n \\
&= \exp(-1/z) \qquad \text{for any } n.
\end{aligned}
$$

which is again standard Fréchet ($\xi = 1$ in the GEV family).

Clearly demonstrates the max-stable property of Fréchet distributions!

Univariate extremes – block maxima



- Spot the heavy tail!

Univariate extremes – block maxima

Simple examples:

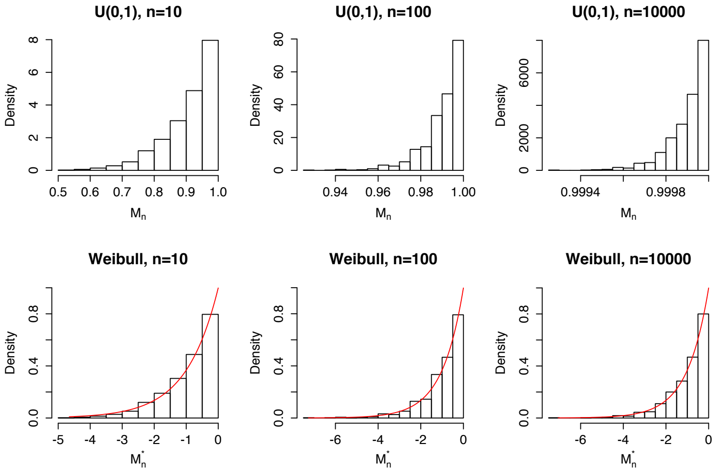
#3 Uniform(0,1): $F(x) = x$ for $0 \leq x \leq 1$.

- Let $a_n = 1/n$ and $b_n = 1$.
- Hence $(M_n - b_n)/a_n \leq 0$
- Then

$$
\begin{aligned}
Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) &= F^n(z/n + 1) \qquad \text{(for } z < 0\text{)} \\
&= (1 + \frac{z}{n})^n \\
&\to e^z \qquad \text{as } n \to \infty.
\end{aligned}
$$

which is Weibull (finite upper end point), with $\xi = -1$ in the GEV family.

Univariate extremes – block maxima



Univariate extremes – block maxima

Comments:

- Choice of a_n and b_n is not unique.
- Different choices (that lead to non-degenerate limits) are always in the GEV family with the same ξ parameter (it is just a location/scale shift).
- But with different location and scale parameters.

New Questions:

- How do we know the domain of attraction for a given F?
- How can we find a_n and b_n?

Answers:

- Could fill books on this (and people have!)
- I'll give outline sketch for a particular simple class.

Univariate extremes – block maxima

- Distribution of sample maxima depends on tail behaviour of F.
- Consider a particular form of tail behaviour:

Von Mises function

A distribution F with right end point x_0 is a Von Mises function if there exists a $z_0 < x_0$ such that

$$1 - F(x) = c \exp\left\{ - \int_{z_0}^{x} \frac{1}{f(u)} du \right\} \qquad \text{for } x > z_0$$

with $c > 0$. Here $f(u) > 0$ is an auxiliary function.

Result:

If F is a Von Mises function, then $F \in D(\text{Gumbel})$.

Also: $a_n = f(b_n)$ and $1 - F(b_n) = 1/n$.

Univariate extremes – block maxima

Helpful result

Suppose F is absolutely continuous with negative second derivative F'' for all $x \in (z_0, x_0)$. Then if

$$\lim_{x \to x_0} \frac{F''(x)(1 - F(x))}{(F'(x))^2} = -1$$

then F is a Von Mises function, and $f(x) = (1 - F(x))/F'(x)$.

This means that if the above holds:

- we know the limit distribution is Gumbel
- we know how to calculate the normalising constants.

Univariate extremes – block maxima

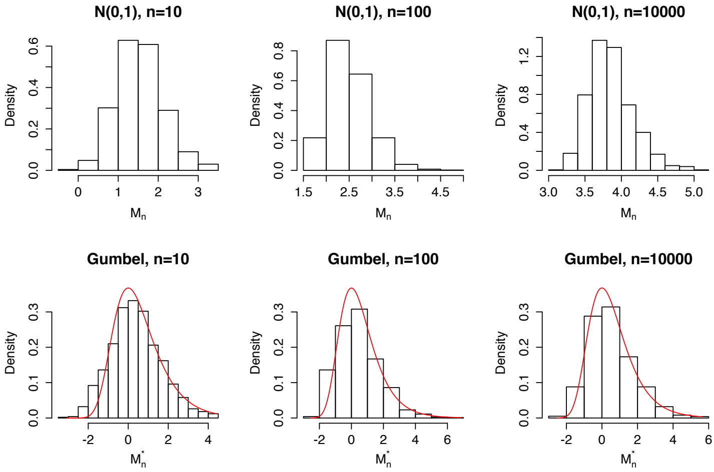
Example: $F(x) = \Phi(x)$ (Gaussian)

- $F'(x) = \phi(x)$
- $F''(x) = -x\phi(x)$
- $$\lim_{x\to\infty} \frac{(1-F(x))F''(x)}{(F'(x))^2} = \lim_{x\to\infty} \frac{-x^{-1}\phi(x)x\phi(x)}{(\phi(x))^2} = -1$$

  and so F is Von Mises and $F \in D(\text{Gumbel})$.
- $$f(x) = \frac{1-F(x)}{F'(x)} = \frac{1-\Phi(x)}{\phi(x)} \sim \frac{x^{-1}\phi(x)}{\phi(x)} = \frac{1}{x}$$

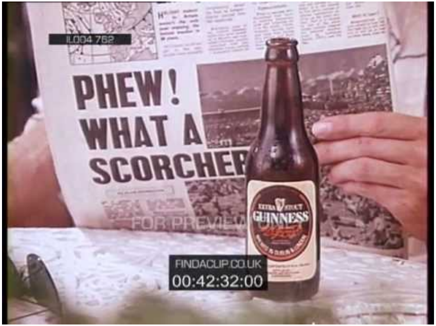
  for large x (Mill's ratio).
- Finally (and after a bit of work ...)
  - $a_n = (2\log n)^{-1/2}$
  - $b_n = (2\log n)^{1/2} - 1/2(\log\log n + \log 4\pi)/(2\log n)^{1/2}$

Univariate extremes – block maxima



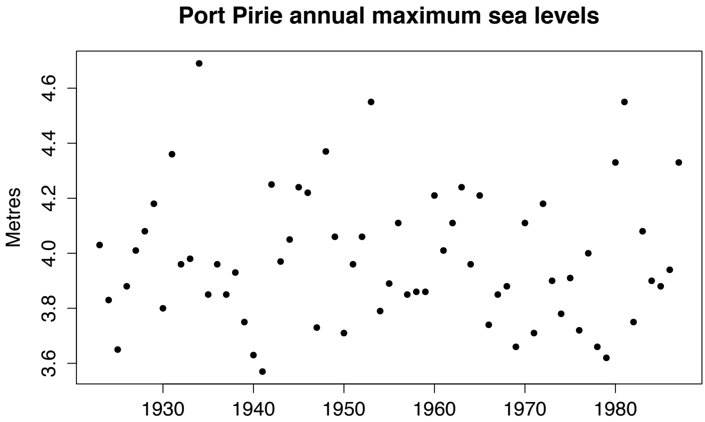
Normal convergence to Gumbel is quite slow.

Talk Outline



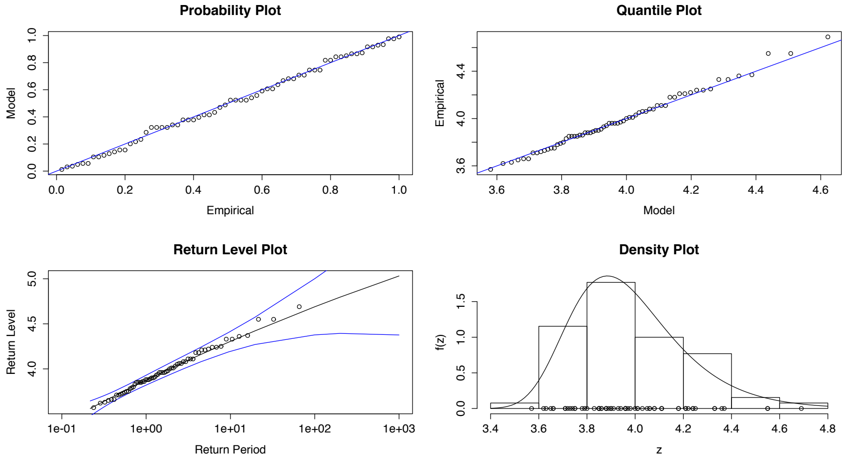
1. Univariate extremes
    - Block maxima results
    - Block maxima practice
    - Threshold models
2. Multivariate extremes
3. Spatial extremes
4. Extras
    - Stereological extremes (ABC)
    - Big data extremes (SDA)

Univariate extremes – block maxima



- Annual (daily) maximum sea-levels at Port Pirie (north of Adelaide), 1923-1987.
- Sea defences: what is the maximum (daily) sea-level we can expect to see in the next 100 years? 1000 years?
- Empirically $Pr(X \geq 4.69) = 1/65$. Need EVT to go beyond this.

Univariate extremes – block maxima



- (asymptotic) GEV fit seems reasonable.
- For predictive inference, use return levels.

Univariate extremes – block maxima

Return levels

- What process level can we expect to be exceeded, on average, once every e.g. 100 years? (E.g. 1 in 100 year storm.)

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

Procedure

- Estimate quantile z_p associated with tail probability p

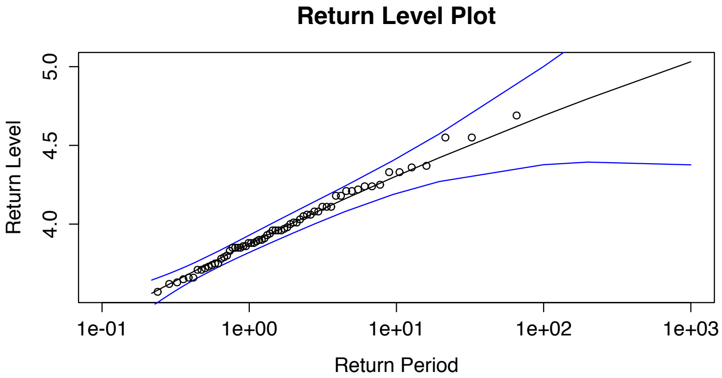
$$G(z_p) = 1 - p.$$

- E.g. For 100-year storm, $p = 0.01$, and

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}] & \text{for } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases}$$

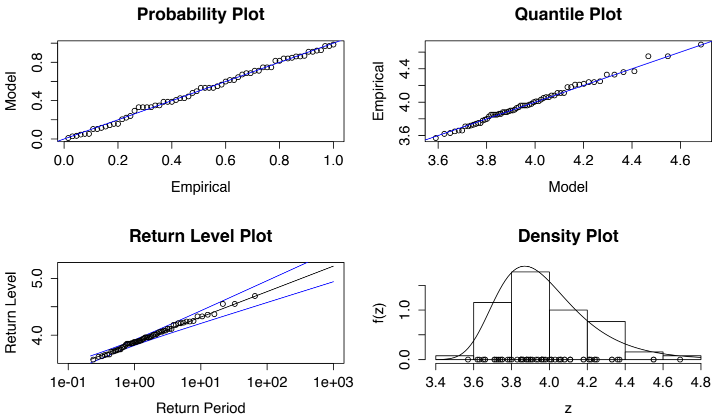
  - z_p is the return level associated with return period $1/p$.
  - E.g. Process can be expected to exceed z_p, on average, once every $1/p$ years.

Univariate extremes – block maxima



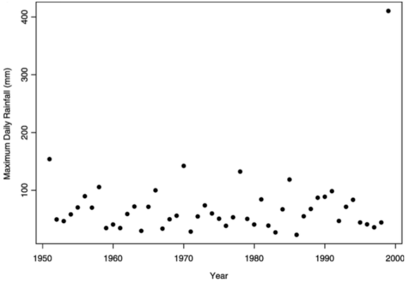
- E.g. 100-year return level is $\approx$ 4.65m.
- Return period of an $\approx$ 5m event is 1000 years.
- Note: approx. 95% CI for ξ is (-0.242, 0.142) suggesting $\xi = 0$ is a possibility.
- Would accept likelihood-ratio test for reducing GEV $\rightarrow$ Gumbel

Univariate extremes – block maxima



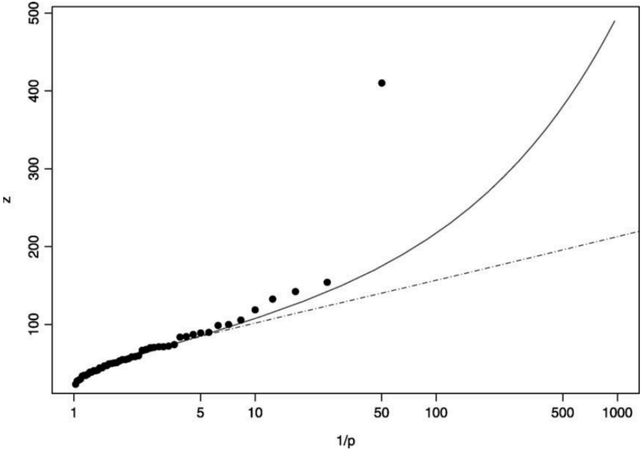
- Slightly improved fit – Gumbel seems a good choice.
- However: only one parameter saved – is this a good thing to do in general?
- Cautionary tale follows ...

Univariate extremes – block maxima



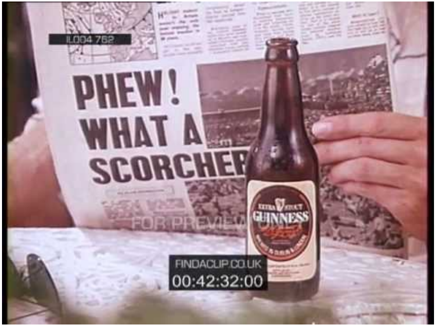
- 50 years of annual rainfall maxima at Maiquetia International Airport, Venezuela
- December 1999 storm way off the charts (extensive damage)
- What is return period of this magnitude of storm (without observing it)? I.e. should we have been prepared for it?
- All statistical tests recommend reducing GEV → Gumbel.

Univariate extremes – block maxima



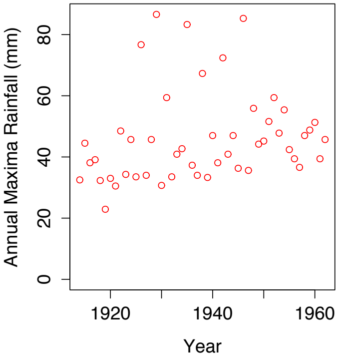
- Return period of Dec 1999 storm (410.4mm)
  - GEV: 4280 year event
  - Gumbel: 17.6 million year event (in ∼50 years of data!)
- Do we really believe this?
  Perhaps just fit GEV, unless 100% certain of Gumbel limit.

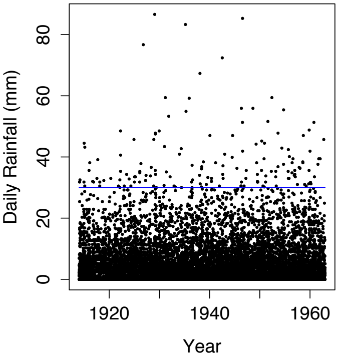
Talk Outline



1. Motivation
2. Univariate extremes
    - Block maxima results
    - Block maxima practice
    - Threshold models
3. Multivariate extremes
4. Spatial extremes
5. Extras
    - Stereological extremes (ABC)
    - Big data extremes (SDA)

Univariate extremes – threshold models





- Daily rainfall SW England 1914–1962.
- 17531 daily data points, 49 annual maxima.
- Analysis of annual maxima only seems wasteful.
- Alternative characterisation of extreme = above a high threshold (e.g. $u = 30$ gives 152 data points here).

Univariate extremes – threshold models

Theorem

Let $X_1, X_2, \ldots \sim F$ (i.i.d) and $M_n = \max\{X_1, \ldots, X_n\}$, and that

$$Pr(M_n \leq z) \approx G(z|\mu, \sigma, \xi) \qquad \text{as before.}$$

Then, for large u, the distribution function of $X - u|X > u$ is approximately

$$H(z) = 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)_+^{-1/\xi}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

- $H(z)$ is in the Generalised Pareto Distribution (GPD) family.
- Precisely: $H(z)$ is the limit distribution of $X - u|X > u$ as $u \to \infty$.
- GPD parameters are function of GEV parameters.
- Increasing u analogous to increasing n for block maxima.

Take away: If block maxima result holds, can model extreme tail by GPD.

Univariate extremes – threshold models

Outline justification:

- If GEV limit holds then for large n

$$F^n(z) \approx \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

  for some μ, σ, ξ.

- Then

$$n \log F(x) \approx -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

- For large z, a Taylor expansion gives

$$\log F(z) \approx -[1-F(z)].$$

- Substituting gives

$$1-F(u) \approx \frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

  for large u.

Univariate extremes – threshold models

- For large u and $y > 0$ we then have

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-1/\xi}$$

for large u.

- Immediately we then have

$$\begin{aligned} Pr(X > u+y \mid X > u) &= \frac{Pr(X > u+y, X > u)}{Pr(X > u)} \qquad (y > 0) \\ &\approx \frac{n^{-1}[1+\xi(u+y-u)/\sigma]^{-1/\xi}}{n^{-1}[1+\xi(u-\mu)/\sigma]^{-1/\xi}} \\ &= \left[1 + \frac{\xi(u+y-\mu)/\sigma}{1+\xi(u-\mu)/\sigma}\right]^{-1/\xi} \\ &= \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]^{-1/\xi} \end{aligned}$$

where $\tilde{\sigma} = \sigma + \xi(u-\mu)$.

Univariate extremes – threshold models

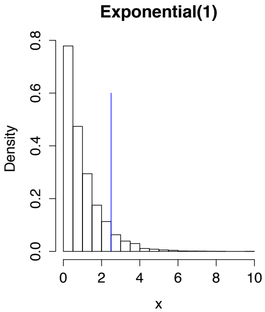
Simple examples (redux):

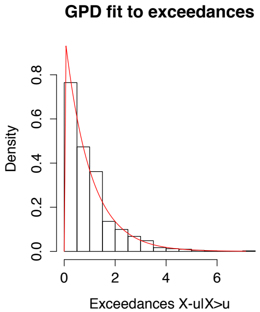
#1 Exponential(1): $F(x) = 1 - e^{-x}$ for $x > 0$.

- By direct calculation

$$\frac{1 - F(u + y)}{1 - F(u)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y} \qquad \text{for } y > 0.$$

- $\Rightarrow$ threshold exceedances are $GPD(\tilde{\sigma} = 0, \xi = 0)$ (=exponential).
- Exact result for all $u > 0$.



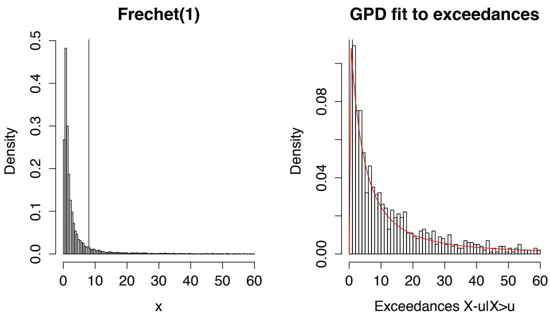


Univariate extremes – threshold models

Simple examples (redux):

#2 Fréchet(1): $F(x) = \exp -1/x$ for $x > 0$.

- $$\frac{1-F(u+y)}{1-F(u)} = \frac{1-\exp\{-(u+y)^{-1}\}}{1-\exp(-u^{-1})} \sim \left(1+\frac{y}{u}\right)^{-1}$$
  as $u \to \infty$.
- $\Rightarrow$ threshold exceedances are $GPD(\tilde{\sigma} = u, \xi = 1)$.

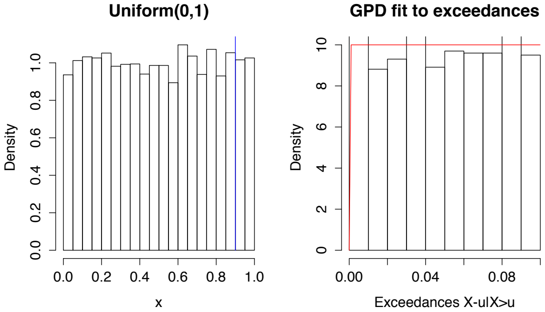


Univariate extremes – threshold models

Simple examples (redux):

#3 Uniform(0,1): $F(x) = x$ for $0 \leq x \leq 1$.

- $$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - (u+y)}{1 - u} = 1 - \frac{y}{1-u}$$
  for $0 \leq y \leq 1 - u$.
- $\Rightarrow$ threshold exceedances are $GPD(\tilde{\sigma} = 1 - u, \xi = -1)$.



Univariate extremes – threshold models

Threshold choice

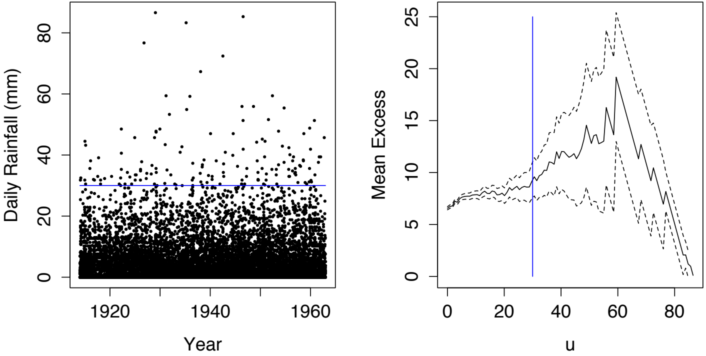
- The GPD is the limiting model as $u \to \infty$
- In practice two contrasting needs:
  - Need to identify large u so that tail of $F \sim GPD$.
    – gives unbiased, but imprecise parameter estimates.
  - Need u low as possible to maximise data exceeding threshold.
    – too low gives precise but biased parameter estimates
  - Ideally, lowest possible u so that GPD limit still approximately holds.
- Two simple methods (many complex methods!):
  - 1) Mean residual life plot
  - 2) Parameter estimates vs threshold plot

Univariate extremes – threshold models

1) Mean-residual life plots:

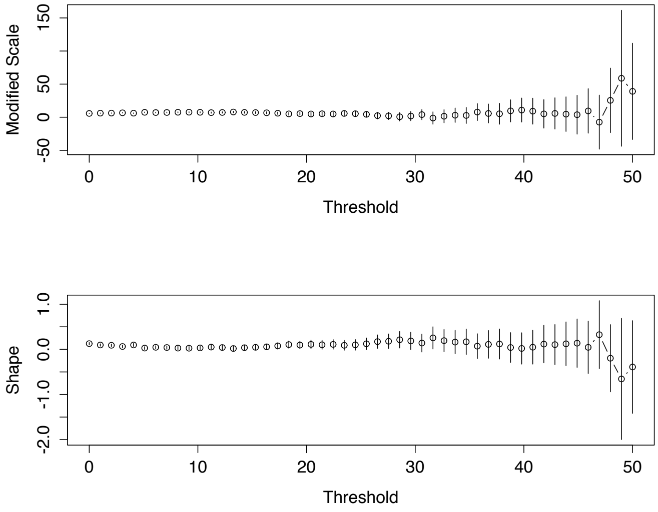
- If $Y \sim GPD(\sigma, \xi)$ then
$$\mathbb{E}[Y] = \frac{\sigma}{1-\xi} \qquad \text{for } \xi < 1.$$
- For threshold modelling
$$\mathbb{E}[X - u_0 | X > u_0] = \frac{\sigma_{u_0}}{1-\xi} \qquad \text{for } \xi < 1.$$
- Now if u_0 is a valid threshold, then any $u > u_0$ is also valid.
- Hence (using the original GPD as F and threshold modelling this):
$$\mathbb{E}[X - u | X > u] = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi u}{1-\xi}.$$
- So for $u > u_0$ $\mathbb{E}[X - u | X > u]$ is linear in u.
- So plot
$$u \quad \text{against} \quad \frac{1}{n_u}\sum_{i=1}^{n_u}(x_{(i)} - u),$$
for $u < x_{max}$. Choose smallest u above which plot is linear.

Univariate extremes – threshold models



- SW England rainfall data
- MRL plot linear after $u = 30$ (6 datapoints above $u = 60$ so ignore)
- Other values also seem credible.

Univariate extremes – threshold models



- 2) Parameter estimates versus threshold plot
- Nothing much happening here, but no reason to overrule MRL plot ($u = 30$)

Univariate extremes – threshold models

Calculation of Return Levels needs a small modification

- If $X - u \sim GPD(\sigma, \xi)$ then for $x > u$

$$Pr(X > x | X > u) = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}.$$

- It follows that

$$Pr(X > x) = Pr(X > x | X > u)Pr(X > u) = \zeta_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}.$$

  as $x > u$, where $\zeta_u = Pr(X > u)$.

- Hence, the level x_m that is exceeded on average once every m observations is the solution of

$$\zeta_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}.$$

Univariate extremes – threshold models

- This gives the *m*-observation return level as

$$x_m = u + \frac{\sigma}{\xi}[(m\zeta_u)^{\xi} - 1].$$

- To get this on a *N*-year return level scale, write

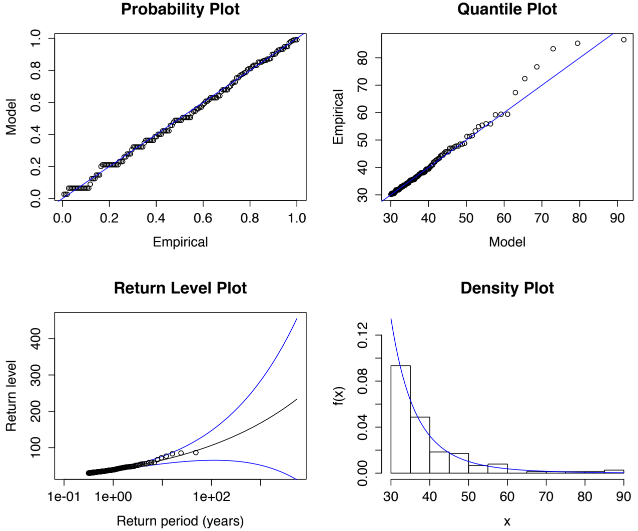
$$m = N \times n_y$$

where n_y is the number of observations in 1 year (extreme or otherwise).

- Finally, need an estimate of ζ_u:

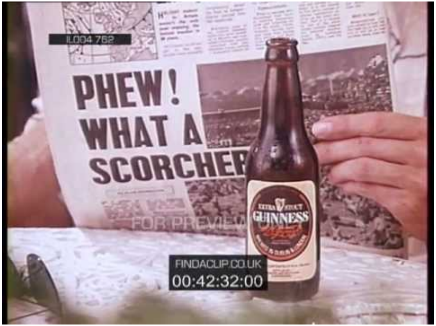
$$\hat{\zeta}_u = \text{proportion of observations above } u.$$

Univariate extremes – threshold models



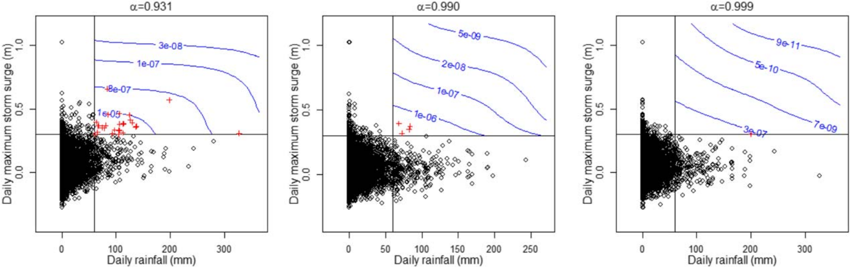
- Semi-ok fit in the tails ($u = 30$)
- Return levels on annual scale (with $n_y = 365.25$, $\hat{\zeta}_u = 152/17531$)

Talk Outline



1. Motivation
2. Univariate extremes
    - Block maxima results
    - Block maxima practice
    - Threshold models
3. Multivariate extremes
4. Spatial extremes
5. Extras
    - Stereological extremes (ABC)
    - Big data extremes (SDA)

Multivariate extremes



- Interest in extremes of more than one variable
- Daily rainfall vs. daily storm surge in Hawkesbury-Nepean catchment (near Sydney)
- Interest in modelling when both (or one) measurement is extreme.
- There are 4 ways to model multivariate extremes:
  - block maxima
  - threshold models
  - point processes
  - conditional modelling (× – not discussed today)

Multivariate extremes

Method #1: Block maxima

Similar to univariate case. If sequences $a_{nd} > 0$, b_{nd}, $d = 1, \ldots, D$, exist such that there is a non-degenerate limit

$$\lim_{n\to\infty} Pr\left(\frac{M_{n1} - b_{n1}}{a_{n1}} \leq z_1, \ldots, \frac{M_{nD} - b_{nD}}{a_{nD}} \leq z_D\right) \to G(z_1, \ldots, z_D)$$

then G is a multivariate extreme value distribution (MEVD), and is max-stable.

- Any univariate margin is GEV
- Fact: domains of attraction are preserved under monotone transformations of margins (Resnick, 1987)
- Common to represent MEVD's with unit Fréchet margins Z_d

Multivariate extremes

Method #1: Block maxima

Under unit Fréchet margins, the joint distribution has the form

$$Pr(Z_1 \leq z_1, \ldots, Z_D \leq z_D) = \exp(-V(z_1, \ldots, z_D))$$

where the exponent measure V takes the form

$$V(z_1, \ldots, z_D) = D \int_{\Delta_D} \max_{d=1,\ldots,D} \frac{w_d}{z_d} H(dw)$$

where

- $\Delta_D = \{w \in \mathbb{R}_+^D : w_1 + \ldots + w_D = 1\}$ is the $D-1$ dimensional simplex
- angular/spectral measure $H(w)$ on Δ_D determines the dependence structure
- satisfies moment conditions $\int_{\Delta_D} w_d H(w) = 1/D$

Multivariate extremes

Method #1: Block maxima

In principle easy to build parametric models for G

- Assume parametric form for $H(w)$
  – must satisfy moment constraints
  – not easy to find closed form for G
- Obtain the density $\partial G/\partial z_1, \ldots, z_D$ to permit inference
  – many, many terms due to partial differentiation.
  E.g. differentiating $\exp(-V)$ gives

$$
\begin{aligned}
\frac{\partial}{\partial z_1}\exp(-V) &= -V_1\exp(-V)\\
\frac{\partial^2}{\partial z_1 z_2}\exp(-V) &= (V_1V_2 - V_{12})\exp(-V)\\
\frac{\partial^3}{\partial z_1 z_2 z_3}\exp(-V) &= (-V_1V_2V_3 + V_{12}V_3 + V_{13}V_2 + V_{23}V_1 - V_{123})\exp(
\end{aligned}
$$

  where V_i is partial derivaive with respect to i.

Inference viable for low D only.

Some parametric families have been worked out. Hard in general.

No single parametric form for G (unlike univariate GEV).

Multivariate extremes

Some models (bivariate):

- Logistic model

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\} \quad x > 0, y > 0$$

which arises (though not obviously!!) through

$$h(w) = \frac{1}{2}(\alpha^{-1} - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1+w)^{-1/\alpha}\}^{\alpha-2}$$

on $0 \leq w \leq 1$.

  - $\alpha \to 1$ gives complete independence

$$G(x, y) = \exp\left\{-\left(x^{-1} + y^{-1}\right)\right\}.$$

  - Smaller $\alpha \to 0$ is increasing dependence
  - One parameter for D-dimensional dependence.

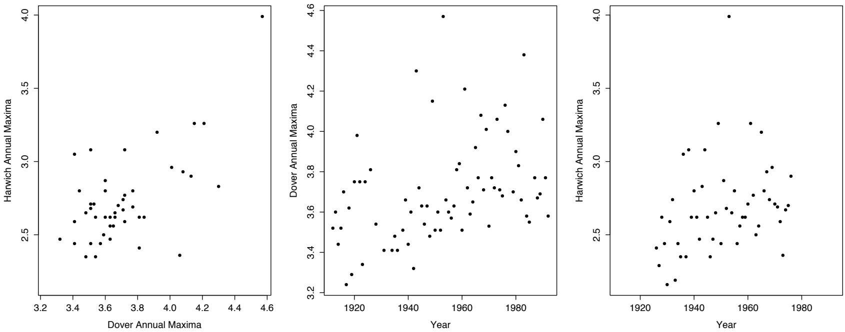
Multivariate extremes

Some other models (bivariate):

- negative logistic
- bilogistic
- negative bilogistic
- dirichlet (multivariate)

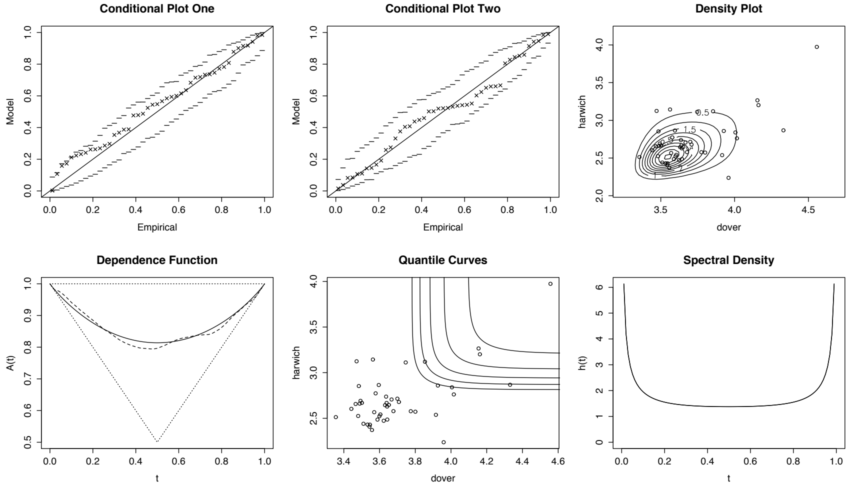
Note: There are not many parametric families for which $h(w)$ can be defined, which satisfies the moment conditions, and for which G has closed form ...

Multivariate extremes



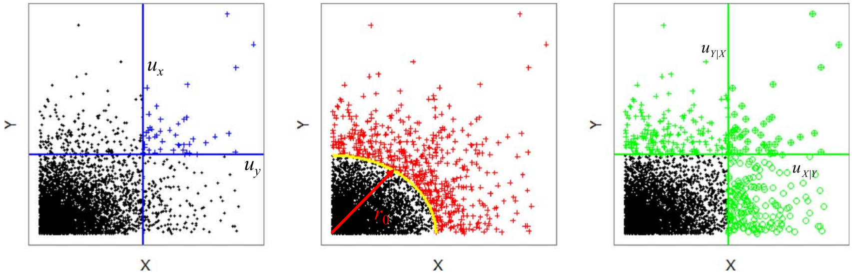
- Annual sea level maxima in Dover & Harwich, 1912–1992.
- Missing data: Dover (9), Harwich (30), both (3).
- Outlier = large storm passing over SE Britain on 1st February 1953.
- Model linear trend in μ in both margins.
- $\hat{\alpha} = 0.70283$ (0.09551) so some dependence.

Multivariate extremes



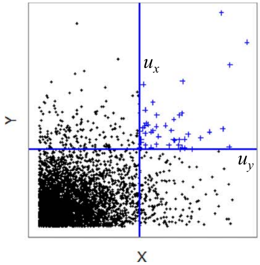
- Annual maxima is very wasteful of data.
- Highly likely $(M_{n1}, \ldots, M_{nD})$ is not an observed event!
- Other approaches are more common.

Multivariate extremes



- Other multivariate EVT methods define "extreme" by some kind of threshold
  - Method #2: threshold models
  - Method #3: point process
  - Method #4: conditional modelling (× – not discussed today)
- Better use of data (but other trade-offs)

Multivariate extremes



- Data $(X_1, \ldots, X_D) \sim F(X_1, \ldots, X_D)$.
- For each margin d, assume $X_d - u_d | X_d > u_d \sim GPD$
- Transform to unit Fréchet margins above u: $X_d \to \tilde{X}_d$.
$\Rightarrow \tilde{F}(\tilde{X}_1, \ldots, \tilde{X}_D)$ is approx Fréchet(1) above u (and something else below u).
- For large n (and above u):

$$\begin{aligned} \tilde{F}(\tilde{X}_1, \ldots, \tilde{X}_D) &= \left[\tilde{F}^n(\tilde{X}_1, \ldots, \tilde{X}_D)\right]^{1/n} \\ &\approx^a \left[\exp\left\{-V(\tilde{X}_1/n, \ldots, \tilde{X}_D/n)\right\}\right]^{1/n} \\ &=^b \exp\{-V(\tilde{X}_1, \ldots, \tilde{X}_D)\}. \end{aligned}$$

- Finally, as $F(X_1, \ldots, X_D) = \tilde{F}(\tilde{X}_1, \ldots, \tilde{X}_D)$ then

$$F(X_1, \ldots, X_D) \approx G(X_1, \ldots, X_D) \qquad \text{above } u.$$

a – As $F^n(nz) = e^{-1/z}$ for unit Fréchet z from slide 16.

b – As $nV(x) = V(x/n)$ is

homogeneous of order -1.

I.e. we can model threshold exceedances using a MEVD.

Multivariate extremes

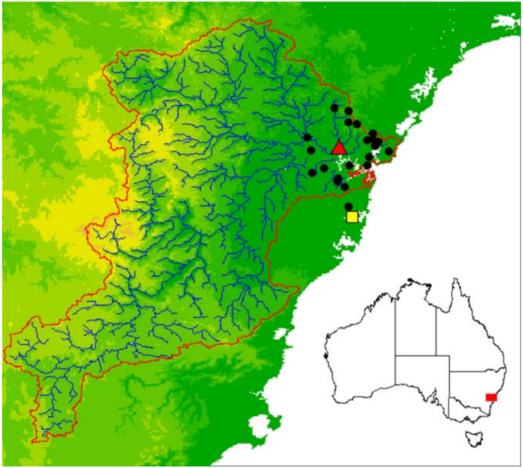
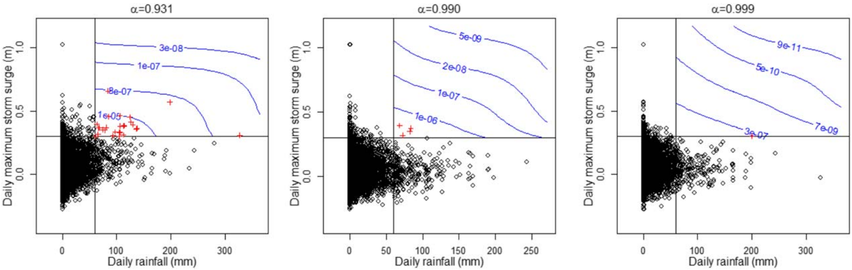


Figure 5. The Hawkesbury-Nepean catchment near Sydney. The catchment is bounded by red lines. The Fort Denison tide gauge and daily rainfall gauges are, respectively, represented by the yellow square and the black dots. The red triangle shows the river cross section at Spencer.

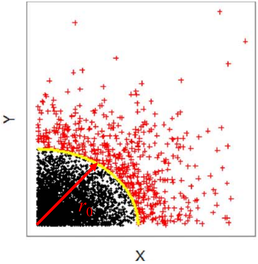
Modelling joint extreme rainfall (dots) and tide surge (square).

Multivariate extremes



- Uses censored likelihood for marginal and joint exceedances $\Rightarrow$ means this method only good for low D.
- Logistic model
- Extremal dependence (α) weakens as distance from tide gauge increases.
- Doesn't require any joint exceedances (RHS), but will estimate compete independence!

Multivariate extremes



- As threshold-exceedance approach is inefficient, consider reworking extremes as point process
- For bivariate X, Y, with block maxima limit $G(X, Y)$.
- Define sequence of point processes $\{N_n\}$ as

$$N_n = \{(x_1/n, y_1/n), \ldots, (x_n/n, y_n/n)\}.$$

- Then

$$N_n \to N \qquad \text{in distribution as } n \to \infty,$$

  to a non-homogeneous Poisson process on $(0, \infty)^2/\{(0, 0)\}$.
- Reparameterising to pseudo-polar $r = x + y$, $w = x/(x + y)$ gives the intensity function of $\lambda(r, w) = 2\frac{dH(w)}{r^2}$
- If we define region $A = \{(x, y) : r > r_0\}$ then convergence is guaranteed (away from the origin) and likelihood function is simply

$$L(\theta|w) \propto \prod_{i=1}^{N_A} h(w_i|\theta), \qquad \text{where } w_i = x_i/(x_i + y_i).$$

Multivariate extremes

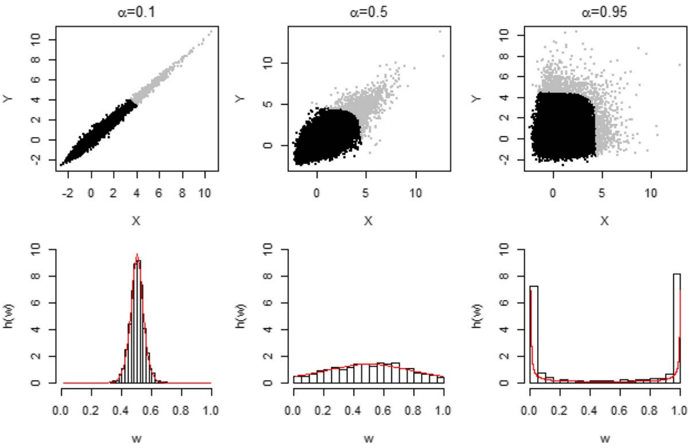
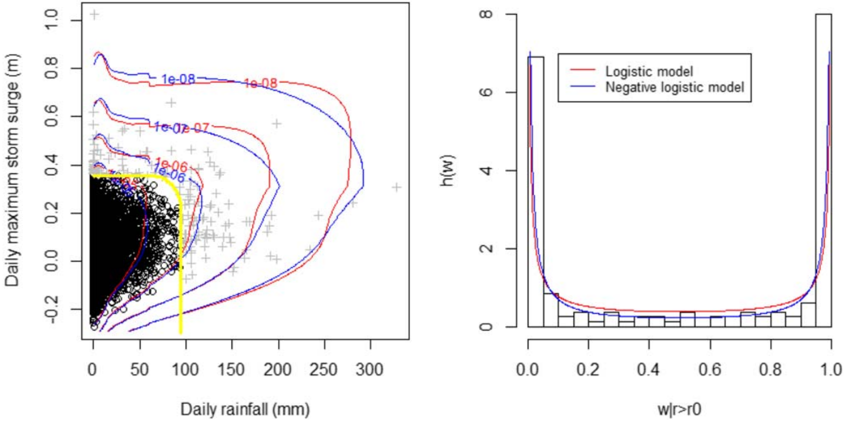


Figure 2. Illustration of the spectral density function, $h(w)$, for three data sets generated from the bivariate logistic model with different levels of dependence ($\alpha = 0.1$, 0.5, and 0.95). (top) Scatterplots of each data set, with extreme events shown in grayed dots. (bottom) Histograms of $w = x/(x + y)$ for the extreme events with superimposed spectral density functions, $h(w)$.

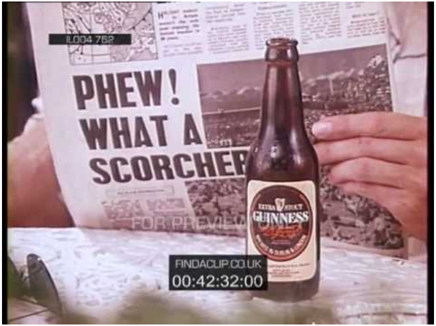
- Illustrations of $h(w)$ capturing extremal dependence.
- Full dependence: $h(w) = \{0.5\}$. Independence: $h(w) = \{0, 1\}$.

Multivariate extremes



- Weak dependence (again).
- AIC Logistic (3710), negative logistic (3683).
- Bump: zero rainfall violates Poisson process assumptions. (No-one really knows how to fix this . . .)

Talk Outline



1. Motivation
2. Univariate extremes
   - Block maxima results
   - Block maxima practice
   - Threshold models
3. Multivariate extremes
4. Spatial extremes
5. Extras
   - Stereological extremes (ABC)
   - Big data extremes (SDA)

Spatial extremes

Recall (univariate block maxima):
If sequences $a_n > 0$ and b_n exist such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) = F^n(a_n z + b_n) \to G(z) \qquad \text{as } n \to \infty,$$

then G is a max-stable distribution (and in particular in the GEV family).

- We are interested in modelling extremes spatially, so it is natural to define a suitable process:

Max-stable process

Let T be an index set and $\{\tilde{Y}_i(t)\}_{t \in T}$, $i = 1, \ldots, n$ be n independent realisations of a continuous stochastic process. If there are sequences of continuous random functions $a_n(t) > 0$ and $b_n(t)$ such that

$$\lim_{n \to \infty} \frac{\max_{i=1}^n \tilde{Y}_i(t) - b_n(t)}{a_n(t)} = Y(t) \qquad t \in T,$$

then $Y(t)$ is a max-stable process.

Spatial extremes

Two properties of the max-stable process:

- For any fixed $t_0 \in T$, the one-dimensional margin $Y(t_0)$ follows a GEV distribution.
- For any finite K-dimensional set $t_1, \ldots, t_K$, the K-dimensional margin $(Y(t_1), \ldots, Y(t_K))$ belongs to the class of K-dimensional extreme value distributions.

(Contrast with similarity to Gaussian process.)

- This standardised process $Z(t)$ has unit Fréchet margins.
- If $Z(t)$ is stationary, it can be expressed through its spectral representation.
- A simple, intuitive version of this is known as the storm profile model.

Spatial extremes

Spectral representation (storm process model):

Let $\{X_j, U_j\}_{j\geq 1}$ be a Poisson process Λ on $\mathbb{R}^d \times \mathbb{R}^+$, with counting measure $\Lambda(\cdot) := \sum_j \mathbb{I}_{(X_j,U_j)}(\cdot)$ and intensity measure $\nu(dx) \times u^{-2}du$, where $\mathbb{I}_{(X_j,U_j)}(A)$ is the indicator function of the random number of points falling in a bounded set $A \subset \mathbb{R}^d \times \mathbb{R}^+$ and ν is a positive measure.

For a non-negative measurable function $f(x-t)$ (for fixed $t \in T$) such that $\int_{\mathbb{R}^d} f(x-t)v(dx) = 1$, the stochastic process

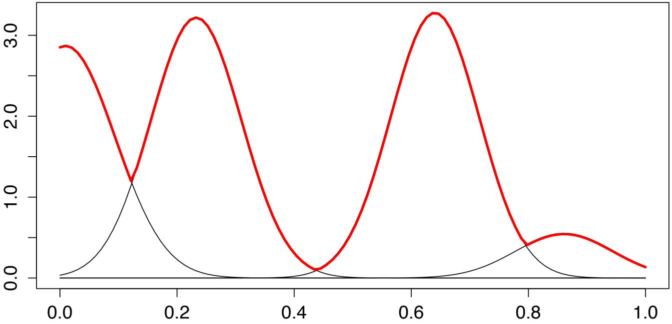
$$Z(t) := \max_{j=1,2,\ldots} \{U_j f(X_j - t)\} \qquad t \in T$$

is a (stationary) max-stable process.

Storm profile model interpretation

- U = storm magnitude
- X = storm centre
- f = storm shape

Spatial extremes



- Gaussian storm shape ("Smith" model), with 4 storms
- Need trick to sample max of $n = \infty$ storms.
- Any point is max of i.i.d. scalar process (Fréchet)
- Clear that spatial dependence is induced

Spatial extremes

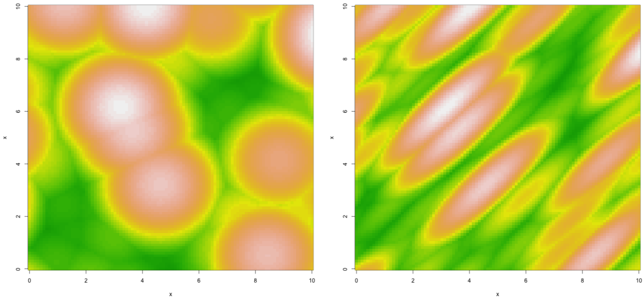


Figure 1.1: *Two simulations of the Smith model with different Σ matrices. Left panel: $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 0$. Right panel: $\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 1$. The max-stable processes are transformed to unit Gumbel margins for viewing purposes.*

- Realisations of Smith model in $D = 2$ dimensions.
- Clear spatial and directional dependence (in a Gaussian sense).

Spatial extremes

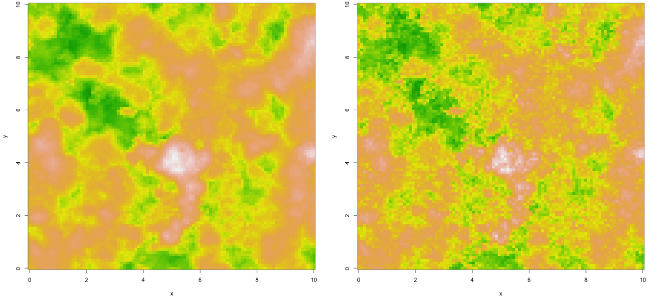


Figure 1.3: *Two simulations of the Schlather model with different correlation functions having approximately the same practical range. Left panel: Whittle–Matérn with $c_1 = c_2 = \nu = 1$. Right panel: Powered exponential with $c_1 = \nu = 1$ and $c_2 = 1.5$. The max-stable processes are transformed to unit Gumbel margins for viewing purposes.*

"Schlather" model based on generalisation of spectral representation

$$Z(t) = \max_j [U_j(t) \max\{0, W_j(t)\}]$$

where $W(t)$ is a stationary process on $\mathbb{R}^d$, $W_j(t)$ are i.i.d replicates, and $\mathbb{E}[\max\{0, W_j(t)\}] = 1$.

- Allows general noise processes (w. geosts. dependence functions)

Spatial extremes

For the storm-profile model, the distribution function of $(Z(t_1), \ldots, Z_(t_K))$ is

$$Pr(Z(t_k) \leq z_k, k = 1, \ldots, K) = \exp\left[- \int_{\mathbb{R}^d} \max_{1 \leq k \leq K} \left\{ \frac{f(x - t_k)}{z_k} \right\} \nu(dx) \right]$$

which has unit Fréchet margins as $(z_k \to \infty)$

$$Pr(Z_t \leq z) = \exp\left(-\frac{1}{z} \int_{\mathbb{R}^d} f(x - t) \nu(dx) \right) = \exp(-1/z).$$

Beyond univariate margins, tractable distribution functions are not really viable for $K \geq 3$.

Spatial extremes

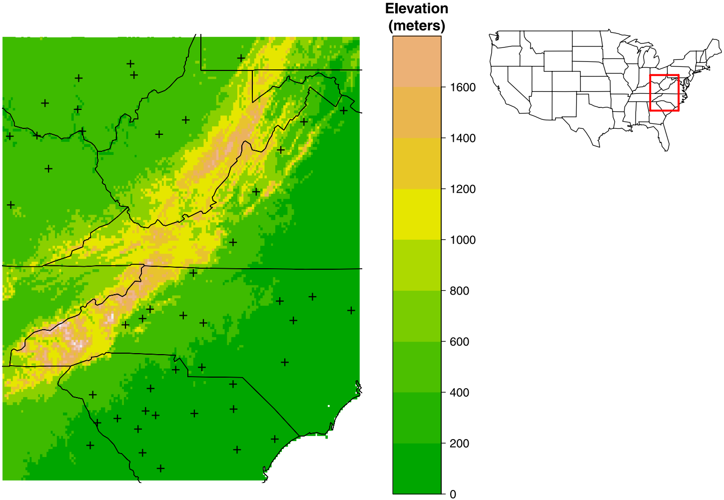
Modelling spatial extremes with the Gaussian (Smith) storm model:

- Two spatial locations $t_i,\ t_j$
- $h = (t_j - t_j)^\top$
- $a(h) = (h^\top \Sigma^{-1} h)^{1/2}$
- Σ is covariance of Gaussian f

$$Pr(Z(0) \leq z_i, Z(h) \leq z_j)$$
$$= \exp\left[-\frac{1}{z_i}\Phi\left(\frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z_j}{z_i}\right) - \frac{1}{z_j}\Phi\left(\frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z_i}{z_j}\right)\right].$$

- So we can model two spatial locations only with this model.
- For $K \geq 2$ need to use (say) composite likelihoods (Padoan et al., 2010) – lots of maths
- Could also use ABC etc.

Spatial extremes



- 46 rainfall gauges East USA, with elevation (Appalachian Mount.)
- Fit Gaussian storm max-stable process using CL
- Spatial surfaces for $\mu(t)$, $\sigma(t)$, $\xi(t)$ (GEV margins)

Spatial extremes

Table 5. Some Gaussian extreme value processes and their corresponding maximized negative composite log-likelihood, degrees of freedom, and the CLIC score

Model		$-\ell_{\mathcal{P}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{y})$	df	CLIC
M_0:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_2(\text{lon})$ $\xi(x) = \gamma_0$	412,110.5	12	825,679
M_1:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_3(\text{lon})$ $\xi(x) = \gamma_0$	412,111.7	11	825,526
M_2:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$ $\xi(x) = \gamma_0$	412,113.6	11	825,459
M_3:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_3(\text{lon})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt}) + \beta_3(\text{lon})$ $\xi(x) = \gamma_0$	412,234.4	11	825,840
M_4:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt}) + \alpha_3(\text{lon})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_3(\text{lon})$ $\xi(x) = \gamma_0$	412,380.9	11	826,177
$\mathbf{M_5}$**:**	$\mu(x) = \alpha_0 + \alpha_1(\text{lat}) + \alpha_2(\text{alt})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$ $\xi(x) = \gamma_0$	**412,113.9**	**10**	**825,327**
M_6:	$\mu(x) = \alpha_0 + \alpha_1(\text{lat})$ $\lambda(x) = \beta_0 + \beta_1(\text{lat}) + \beta_2(\text{alt})$ $\xi(x) = \gamma_0$	412,314.4	9	825,684

- Model choice using CLIC (Varin & Vidoni, 2005)
- $\xi(t)$ typically does not change over space
  (i.e. not enough data to clearly identify variations.)

Spatial extremes

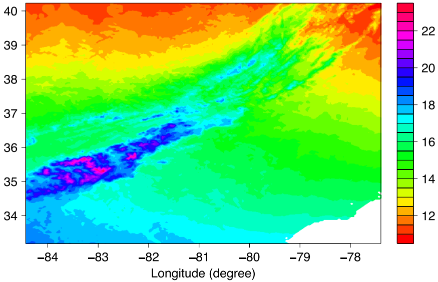
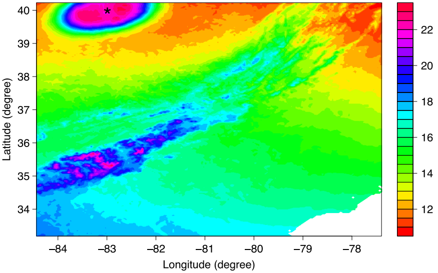


Figure 7. Pointwise 50-year return level map (cm) estimated from the fitted Gaussian extreme value process, model M_5.



- LHS: Unconditional 50-year return levels
- RHS: 50-year return levels conditional on observing 13cm magnitude event at $\star$
- Clear altitude and spatial dependence (in direction of mountain range).

Spatial extremes

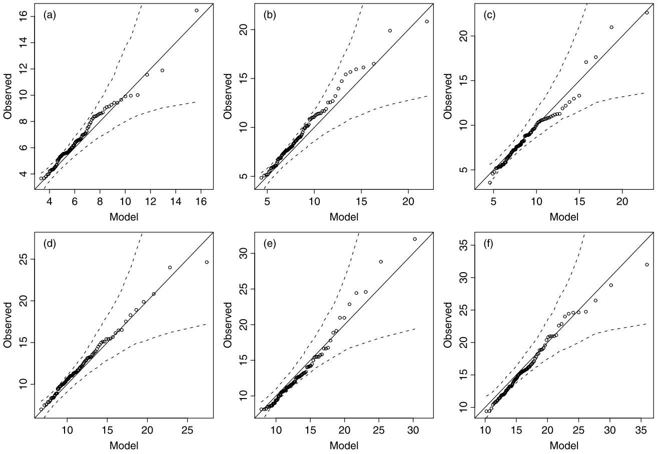
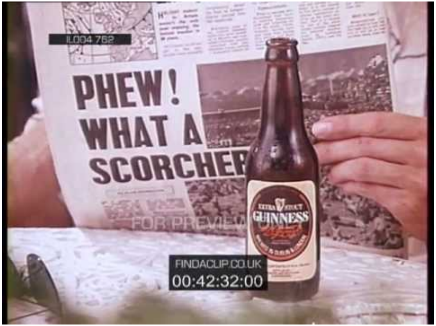


Figure 9. Goodness-of-fit diagnostics for the fitted model for various subsets of site locations, $K' \in \{1, \ldots, K\}$. Panels display quantile plots of observed block-maxima $\tilde{y}_m = \max_{k \in K'}\{y_{m,k}\}$ versus means of their respective distributions, obtained by simulation, under the fitted model (10,000 replications). Dashed lines indicate simulation-based 95% confidence envelopes. Top panels: K' consisting of three pairs of locations separated by (a) short ($\approx$ 20 km), (b) medium ($\approx$ 350 km), and (c) long ($\approx$ 735 km) distances. Bottom panels: K' consisting of (d) the 10 lowest, (e) the 10 highest, and (f) all $K = 46$ sites.

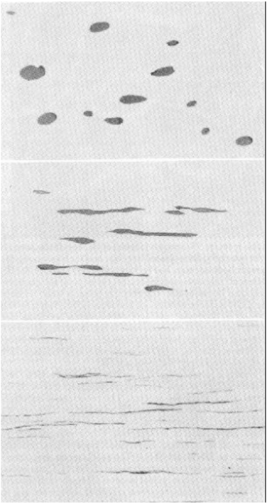
- Always perform goodness-of-fit checks!

Talk Outline



1. Motivation
2. Univariate extremes
    - Block maxima results
    - Block maxima practice
    - Threshold models
3. Multivariate extremes
4. Spatial extremes
5. Extras
    - Stereological extremes (ABC)
    - Big data extremes (SDA)

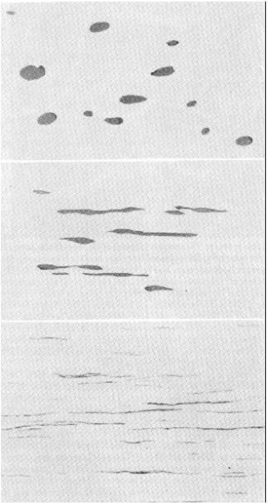
The production of clean steels



Stereological extremes:

- Inclusions (impurities) degrade quality of steel
- Strength related to largest inclusion size
- 2D slice taken from 3D block
- Distribution of size of largest inclusion?
- EVT twist on Wicksell's corpsucle problem
  - Inference on distribution of radii of 3D spheres based on 2D slice
  - We want distribution of largest inclusions
  - Spherical or ellipsoidal inclusions

The production of clean steels



Extremes model

- Point process for inclusion centres, rate λ
- Location-independent inclusion diameters
- Inclusion diameters $(V - v_0)|V > v_0 \sim GPD(\sigma, \xi)$
- Model for observed 2D diameters $D = g(V)$
  - Computationally tractable (just) for spherical inclusions
  - Impractical for ellipsoidal inclusions
- Can use Approximate Bayesian Computation (ABC)
- Bortot et al (2007) use order statistics $S = (D_{(1)}, \ldots, D_{(n)}, n)$ where $n_{obs} = 112$.
- Fan et al (2013) use 3 "Fearnhead & Prangle" statistics.

ABC – 1 slide recap

The basic mechanism:

- Draw samples $(\theta_1, x_1), (\theta_2, x_2), \ldots, (\theta_N, x_N)$ from $\pi(\theta, x) = \pi(x \mid \theta)\pi(\theta)$
- x_k is simulated data from the likelihood with parameter θ_k
- Look at marginal distribution of $\pi(\theta \mid x \approx y)$
- If x is "close enough" to y then $\pi(\theta \mid x \approx y) \approx \pi(\theta|y)$

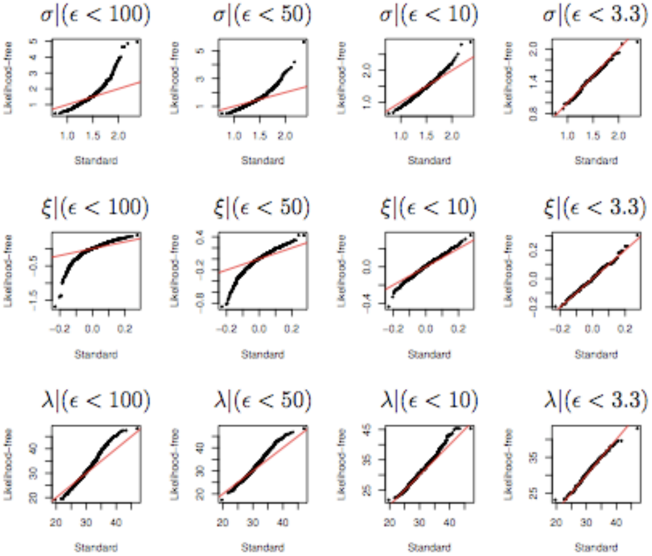
Dimension reduction:

- Condition $x \approx y$ is very unlikely (imagine high-dimensional time series, or high-resolution image)
- Reduce data to summary statistics $s_x = S(x)$
- Idea is then:
  - Draw samples from $\pi(\theta, x) \Rightarrow \pi(\theta, s_x)$
  - Look at the marginal distribution of $\pi(\theta \mid s_x \approx s_y) \approx \pi(\theta|s_y)$

Note: If s_y are sufficient statistics, then $\pi(\theta|s_y) \equiv \pi(\theta|y)$, otherwise some loss of information.

Stereological extremes: variance inflation

Verification via comparison with MCMC output for spherical model



Stereological extremes

Different results under ellipsoidal model

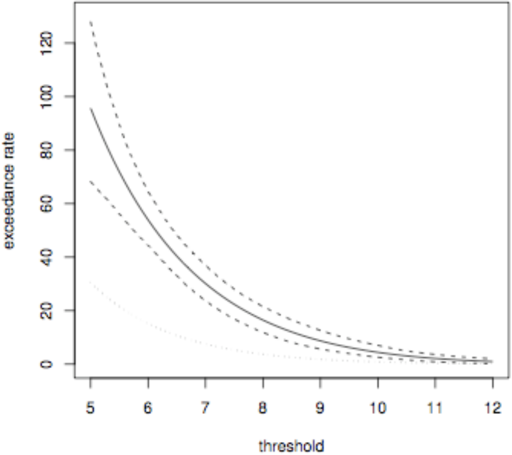
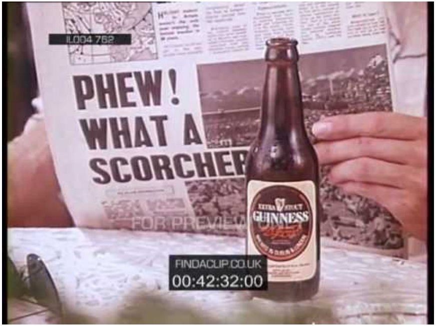


FIGURE 6. Solid curve shows posterior mean of rate of inclusions exceeding a threshold u as a function of u under the ellipsoidal model assumption. The broken curves are pointwise 95% credibility interval limits of the same quantity. The dotted curve shows the marginal mean of the same quantity under the spherical inclusion model

Talk Outline

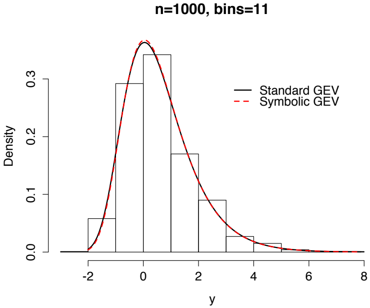


1. Motivation
2. Univariate extremes
   - Block maxima results
   - Block maxima practice
   - Threshold models
3. Multivariate extremes
4. Spatial extremes
5. Extras
   - Stereological extremes (ABC)
   - Big data extremes (SDA)

Fitting a GEV (review)

Suppose $x_1, \ldots, x_n \sim \text{GEV}(\mu, \sigma, \xi)$ for large n.
Create histogram of counts $s = (s_1, \ldots, s_B)$.
Symbolic log-likelihood function is then

$$\ell(s|\mu,\sigma,\xi) \propto \sum_{b=1}^{B} s_b \log\left[F(a_{b+1}|\mu,\sigma,\xi) - F(a_b|\mu,\sigma,\xi)\right]$$



Note that $\ell(s|\mu,\sigma,\xi)$ tends to standard likelihood as # bins gets large (so 1 or 0 observations per bin)

Computation:

- Optimisation of ℓ (v. quick)
- Creation of histogram s (slower)

$\leftarrow$ good fits with moderate bin numbers

Fitting multivariate extremes

Exactly the same as before:

Have $x_1, \ldots, x_n \sim \mathrm{f}(x|\theta)$ for large n where $x_i = (x_i^1, \ldots, x_i^p)^\top \in \mathbb{R}^p$.

Create multivariate histogram of counts $s = (s_1, \ldots, s_B)$ over $\mathbb{R}^p$.

Symbolic log-likelihood function is then

$$\ell(s|\theta) \propto \sum_{b=1}^{B} s_b \log \left[\int_{D_b} f(z|\theta) dz \right]$$

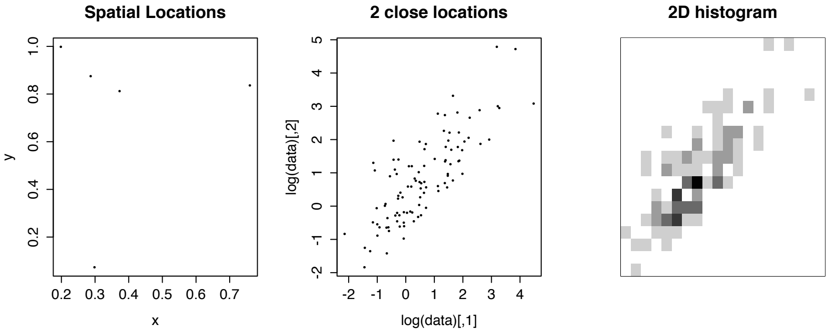
Comments:

- Multivariate histograms become inefficient as p gets large (leading to other symbol types ...)
- Constructing histograms more difficult for large n, p (map-reduce)
- E.g. for bivariate models

  $$\int_{D_b} f(z|\theta) dz = F(b_1, b_2|\theta) - F(a_1, b_2|\theta) - F(b_1, a_2|\theta) + F(a_1, a_2|\theta)$$

  where bin $D_b = [a_1, b_1] \times [a_2, b_2]$.
- So 2^p terms in integral – only practical for moderate p.

Fitting multivariate extremes



Example

- Data from spatial max-stable process ("Smith" model)
- Construct composite pairwise likelihood $\ell_C = \sum_{i \neq j} \ell(x_i, x_j | \theta)$
- Each $\ell(x_i, x_j | \theta)$ follows bivariate extremes model
- Look at bivariate histogram for each $\ell(x_i, x_j)$
  (avoids high-D histogram problems)

Fitting multivariate extremes

# Bins	σ_{11}	σ_{12}	σ_{22}	μ	σ	ξ
2	149.3 (191.4)	53.9 (41.9)	321.1 (286.9)	21.888 (41.37)	5.804 (5.93)	0.392 (0.08)
3	122.3 (163.6)	53.3 (37.6)	265.3 (252.3)	7.969 (20.68)	2.478 (2.64)	0.207 (0.12)
4	125.1 (160.5)	49.4 (35.5)	243.8 (250.5)	1.987 (2.94)	1.876 (2.09)	0.140 (0.10)
5	121.7 (185.8)	41.3 (119.6)	236.1 (245.4)	1.676 (2.78)	2.056 (2.26)	0.105 (0.10)
7	119.3 (155.3)	39.6 (86.6)	229.3 (236.4)	1.545 (3.09)	1.525 (2.36)	0.071 (0.08)
10	114.4 (159.2)	38.3 (61.8)	191.7 (223.1)	1.580 (4.01)	1.458 (3.18)	0.047 (0.07)
classical	98.9 (81.1)	31.8 (12.9)	153.6 (222.6)	0.064 (0.01)	0.131 (0.01)	0.0025 (0.02)

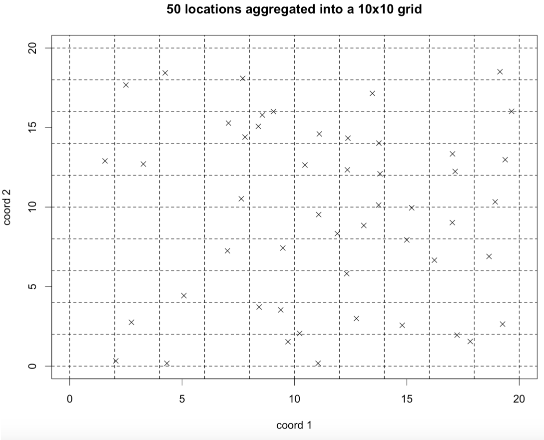
# Bins	σ_{11}	σ_{12}	σ_{22}	μ	σ	ξ
2	51.4 (60.2)	25.3 (20.2)	71.3 (67.4)	7.922 (5.23)	4.838 (2.19)	0.289 (0.06)
4	48.3(71.4)	26.2 (17.3)	65.8 (74.2)	0.618 (1.21)	1.320 (1.98)	0.045 (0.05)
7	44.7 (35.5)	22.6 (26.9)	45.4 (50.1)	0.445 (0.95)	1.791 (2.10)	0.010 (0.02)
10	38.4 (38.1)	19.4 (19.3)	43.6 (38.8)	0.473 (0.80)	0.318 (1.03)	0.003 (0.01)
classical	30.2 (20.4)	12.9 (9.8)	38.9 (22.1)	0.057 (0.01)	0.094 (0.008)	0.0008 (0.005)

Table 7: Histogram mean MSE of the `constrOptim` calculated MLE for each parameter in the Smith model, with associated standard error in brackets. The parameter values of $(\sigma_{11}, \sigma_{12}, \sigma_{22}, \mu, \sigma, \xi)$ used to generate each dataset were (200, 120, 150, 4, 6, 1). The top table was from 5 locations, while the bottom table was from 10 locations.

- Same performance as before
- Dependence parameters a function of # of spatial locations (obv.)
- Need more than 10 bins for acceptable accuracy. Concept ok tho. ✓

Fitting spatial extremes

- Suppose we also have an arbitrarily large number of spatial locations
- Standard: Composite likelihood term $L(x_i, x_j|\theta, L_i, L_j)$ for each location pair (L_i, L_j)
- # pairs (or triples or ...) in composite likelihood would explode



Alternatively:
Consider likelihood where location is binned

Any locations in same bin are considered the same location.

Rather than very many likelihood contributions $L(x_i, x_j|\theta, L_i, L_j)$ we have (loosely)

$$\left[\int_{\tilde{D}_j}\int_{\tilde{D}_i} L(x_i, x_j|\theta, L_i, L_j)dL_i dL_j\right]$$

for bins $\tilde{D}_i$ and $\tilde{D}_j$
evaluated for pooled histogram data.

Fitting spatial extremes

- So replace likelihoods of pointwise pairs L_i, L_j ($n(n-1)/2 =$ huge!) with bin pairs $\tilde{D}_i, \tilde{D}_j$ (far less & controllable)
- And replace pairwise symbolic likelihood (based on bivariate histograms):

$$\int_D f(z|\theta, L_i, L_j)dz = F(b_1, b_2|\theta) - F(a_1, b_2|\theta) - F(b_1, a_2|\theta) + F(a_1, a_2|\theta)$$

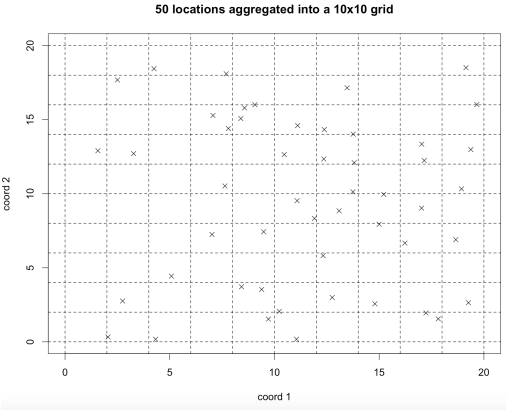
with

$$\int_{\tilde{D}_j}\int_{\tilde{D}_i}\int_D f(z|\theta, L_i, L_j)dz\,dL_i dL_j$$

- So binning data over time (dz) and space (dL_j, dL_i).
- Note 1: pool the n_{ij} histograms of combined locations $\ell(\text{combined}) = \prod_{ij} \ell_{ij}(L_i, L_j)$
- Note 2: need to be able to do integration fast.

Fitting spatial extremes

Comments (positives):



For $b \times b$ grid no need to compute all $b(b-1)/2$ combinations of grid-pairs.

If e.g. isotropic dependence (depending only on distance)

e.g. $(1,1) \leftrightarrow (1,3)$ has same dependence as $(1,2) \leftrightarrow (1,3)$ and $(1,1) \leftrightarrow (3,1)$

So can pool the bivariate histograms

Unique patterns depends on grid, not number of spatial locations.

- As before: as bin sizes shrink in time and space, recover exact composite likelihood approach (e.g. Padon et al, 2010)

Fitting spatial extremes

Comments (negative/neutrals):

- Need to do the 4-dimensional integral many times:

$$\int_{\tilde{D}_j}\int_{\tilde{D}_i}\int_D f(z|\theta)dz\,dL_i\,dL_j$$

a) different locations L_i, L_j, b) different 2d data bins D.

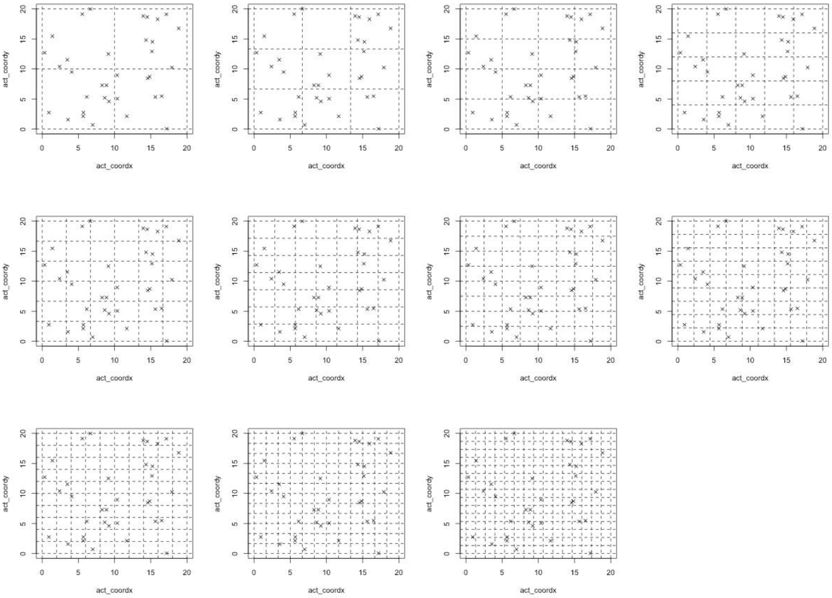
- E.g. Smith:

$$F(z|\theta) = \exp\left(-\frac{1}{z_i}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{z_j}{z_i}\right)-\frac{1}{z_j}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{z_j}{z_i}\right)\right)$$

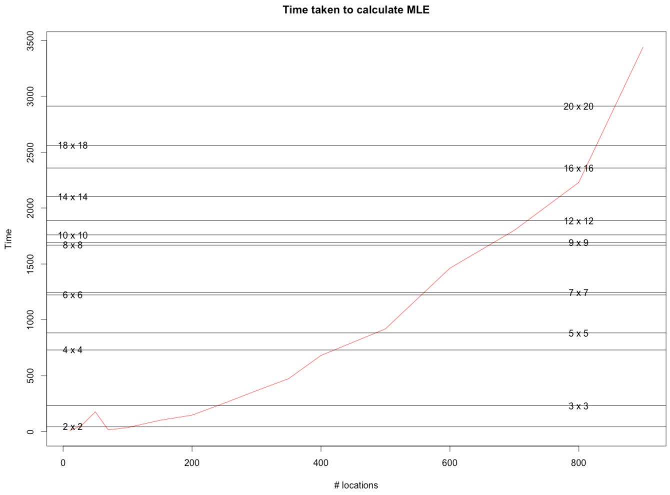
where $a^2 = (L_i - L_j)^\top \Sigma^{-1}(L_i - L_j)$. Or Schlather geostat models etc.

- Integrating over $z = (z_i, z_j)^\top$ is easy as cdf $F(z|\theta)$ is known
- Not sure how to do space (L_i, L_j) yet.
- Currently using numerical method but too slow.
  Needs faster method. (In progress!)

Fitting spatial extremes

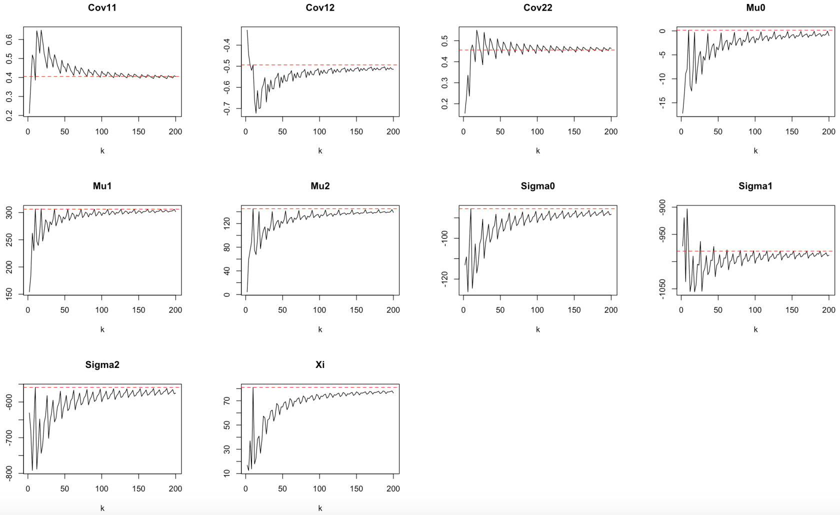


Fitting spatial extremes



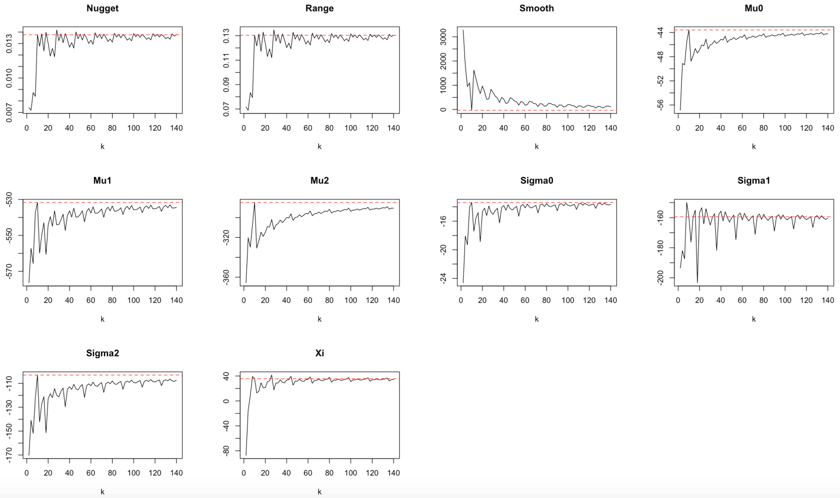
- Method will be more efficient than standard at some point.
- Currently not so favourable due to numerical integration (working on that!)

Fitting spatial extremes



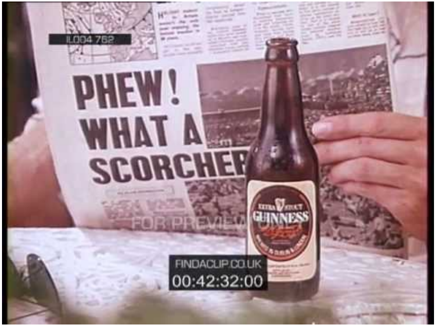
- Smith model with covariate dependent parameters
- Approaches true parameters as bin size shrinks

Fitting spatial extremes



- Schlather model with covariate dependent parameters
- Approaches true parameters as bin size shrinks

Talk Outline



Phew! What a (relief this talk is over)!

THANK YOU!