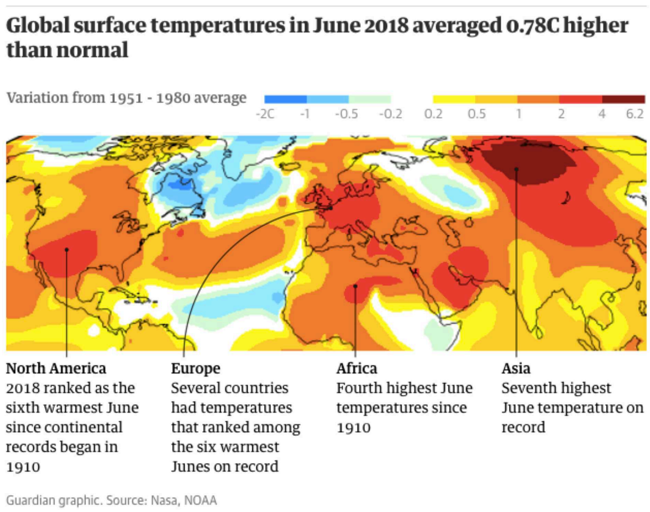
Statistics of Extremes: Overview

Anthony Davison

http://stat.epfl.ch

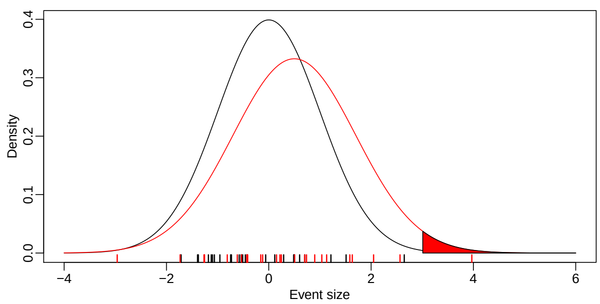
Summer 2018

- Heatwaves, wildfires, drought, heavy rainfall ...



Extrapolation

- For disaster planning, public health, construction (and insurance) we need to **extrapolate** to
  - the tails of distributions, beyond previous events
  - new conditions in a warmer (and more variable?) world.



Motivations for modelling extremes

- **Estimation** of changes in extremes, for better forecasting.
- **Risk assessment** at a single important site.
- **Risk estimation** for particular (compound?) events:
  - What is the risk of crop failure due to drought over a large region?
  - What might the total insurance payout be in case of a major windstorm, or flooding of a major city?
- **Attribution** of events to possible causes: to what extent is a heatwave caused by climate change?
- These involve:
  - accurate interpolation or extrapolation;
  - accurate marginal and/or joint modelling of extreme events.
- Risk estimation can involve **bold extrapolation**:
  - e.g., prediction of 'ten-thousand year event' from 80 years of data.
- **Basic problem**: the events are (used to be!) rare, so there may be little (or no) directly relevant data.

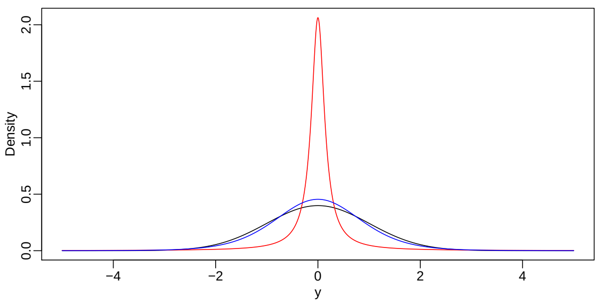
slide 5

Why specialised models?

- Task is **extrapolation** to rarer events.
- The (multivariate) normal distribution is too inflexible for accurate modelling of distribution tails.
- Extrapolation from a fit to the entire distribution can be misleading:
  - there may be regime change in the tails,
  - different fits to the bulk may give very different tail estimates—in particular, the light tails of the Gaussian density can grossly underestimate probabilities of rare events,
  - multivariate Gaussian models predict independence of very rare compound events ('the formula that killed Wall Street').
- Standard copulas can deal with transformations to marginal distributions, but not with joint dependence.

slide 6

Gaussian tails and probabilities

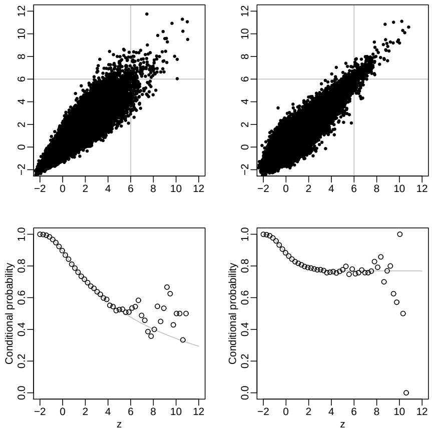


- Gaussian (black), Cauchy (red) and t_{20} (blue) densities matched to have probabilities 0.05 for $|Y| > 1.96$
- Ratios of Cauchy/Gaussian and t_{20}/Gaussian probabilities for $|Y| > y$:

y	2	3	4	5	6	7
Cauchy ratio	1.08	12	387	34247	8.3×10^6	5.5×10^9
t_{20} ratio	1.01	1.7	6.1	58	1589	1.3×10^5

slide 7

Joint Gaussian tails for rare events



- Conditional probabilities $\mathrm{P}(Z_2 > z \mid Z_1 > z)$ as a function of z, for
  - bivariate Gaussian data with correlation 0.9 (left)
  - bivariate extreme-value data with extremal coefficient 1.23 (right)

slide 8

Extremal paradigm

- We need a basis for extrapolation outside the sample, possibly based on a small (extreme) subset of the data.
- The uncertainty will inevitably be large, and must be taken into account.
- Standard models and methods are too limiting, because
  - they cannot accommodate heavy tails
  - their joint tail properties are too inflexible for wide use, and risk mis-(under-?) estimating probabilities for joint events.
- Hence the **extremal paradigm**:
  - Fit asymptotically-justified models which extrapolate 'appropriately',
  - (but check the adequacy of these models carefully!)

slide 9

Max-stability and the GEV

- Basic notion is **max-stability**: 'the maximum of 100 consecutive years of data equals the maximum of the ten decadal maxima'.
- This implies that if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ and there are sequences $\{a_n\} > 0$ and $\{b_n\}$ such that

$$M_n = \frac{\max(X_1, \ldots, X_n) - b_n}{a_n}$$

has a non-degenerate limiting distribution G, then G must satisfy the **stability postulate** (Fréchet, 1927)

$$G^T(b_T + a_T y) = G(y), \quad T > 0.$$

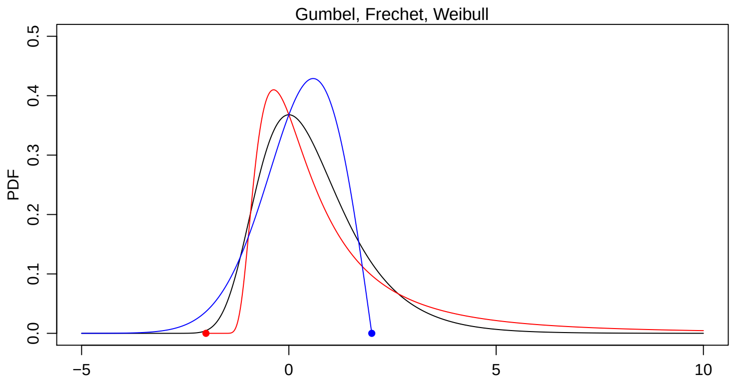
- The **Extremal Types Theorem** (Fisher and Tippett, 1928) states that the only solution is the **generalized extreme-value (GEV)** distribution:

$$G(y) = \exp\{-\Lambda(y)\}, \quad \text{with} \quad \Lambda(y) = \begin{cases} \left(1 + \xi \frac{y-\eta}{\tau}\right)_+^{-1/\xi}, & \xi \neq 0, \\ \exp\left(-\frac{y-\eta}{\tau}\right), & \xi = 0, \end{cases}$$

where $a_+ = \max(a, 0)$, η is a real location parameter, τ is a positive scale parameter, and ξ is a real shape parameter.
- The GEV is a 'universal' law, analogous to use of Gaussian for averages.

slide 11

'Three types'



- ξ determines the rate of tail decay, with:
  - $\xi > 0$ giving the heavy-tailed **Fréchet, Type II**, bounded below;
  - $\xi = 0$ giving the light-tailed **Gumbel, Type I**, with support on $\mathbb{R}$;
  - $\xi < 0$ giving the short-tailed **(reverse) Weibull, Type III**, bounded above.
- Minima have limiting distribution $1 - G(-y)$, so no need to consider them separately.

slide 12

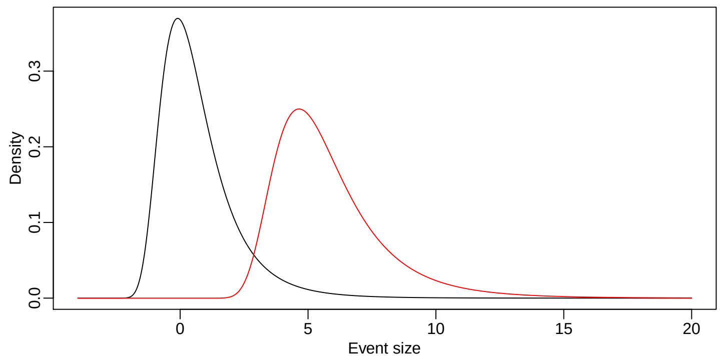
Extrapolation

- To extrapolate to T-year maxima (below, with $T = 50$ in red) from a GEV fitted to annual maxima, we use

$$G^T(y; \eta, \tau, \xi) = G(\eta_T, \tau_T, \xi),$$

where $\eta_T = \eta + \tau(T^\xi - 1)/\xi$ and $\tau_T = \tau T^\xi$; ξ is unchanged.

- In applications the parameters are estimated and the uncertainty may be large.



slide 13

Poisson process

- Random point pattern $\mathcal{P}$ in a state space $\mathcal{E}$ defined by properties of counts

$$N(\mathcal{A}) = |\{x : x \in \mathcal{P} \cap \mathcal{A}\}|, \quad \mathcal{A} \subset \mathcal{E},$$

satisfying two properties:

  – $N(\mathcal{A}_1), \ldots, N(\mathcal{A}_k)$ independent for disjoint $\mathcal{A}_1, \ldots, \mathcal{A}_k$,

  – $N(\mathcal{A}) \sim \text{Poiss}\{\mu(\mathcal{A})\}$,

where the measure μ is non-atomic, and often has an **intensity** $\dot{\mu}$.

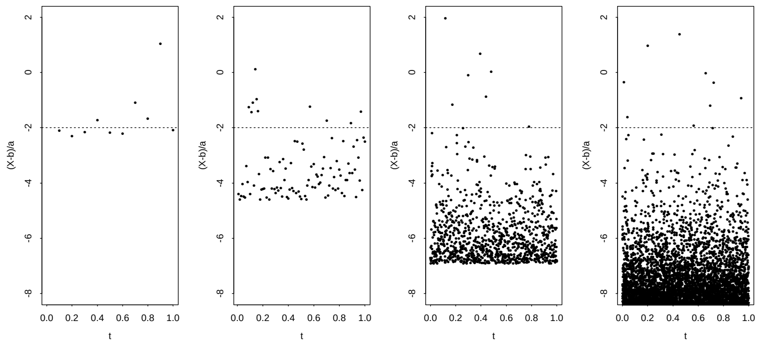
- **Mapping theorem**: if $g : \mathcal{E} \to \mathcal{E}^*$ does not create atoms, then $\mathcal{P}^* = g(\mathcal{P})$ is also a Poisson process.
- Restriction of process $\mathcal{P}$ to $\mathcal{E}' \subset \mathcal{E}$ is also Poisson.
- Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ and for $b_n \in \mathbb{R}$ and $a_n > 0$ define point processes

$$\mathcal{P}_n = \left\{ \frac{X_j - b_n}{a_n} : j = 1, \ldots, \right\}, \quad \mathcal{E} = \mathbb{R}.$$

- Then the rescaled maximum M_n has a non-degenerate limiting distribution G iff $\mathcal{P}_n$ converges to a Poisson process with mean measure $\Lambda(y) \equiv \Lambda\{(y, \infty)\}$ for $y \in \mathbb{R}$.

slide 14

Point process approximation



- Take $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ for which the GEV limit holds and form binomial processes

$$\mathcal{P}_n = \left\{ \left(\frac{j}{n+1}, \frac{X_j - b_n}{a_n} \right) : j = 1, \ldots, n \right\}.$$

- As $n \to \infty$, $\mathcal{P}_n$ converges to a Poisson process $\mathcal{P}$ on $[0,1] \times \mathbb{R}$ with measure

$$\mu\{[t_1, t_2] \times (x, \infty)\} = (t_2 - t_1)\Lambda(x), \quad 0 \le t_1 < t_2 \le 1, x \in \mathbb{R}.$$

slide 15

Threshold exceedances

- Exceedances of a threshold u occur at the times of a homogeneous Poisson process of rate $\Lambda(u)$, and their sizes are independent with the **generalized Pareto distribution (GPD)**

$$H(x) = \begin{cases} 1 - (1 + \xi x/\sigma)_+^{-1/\xi} & \xi \neq 0, \\ 1 - \exp(-x/\sigma), & \xi = 0, \end{cases} \quad x > 0,$$

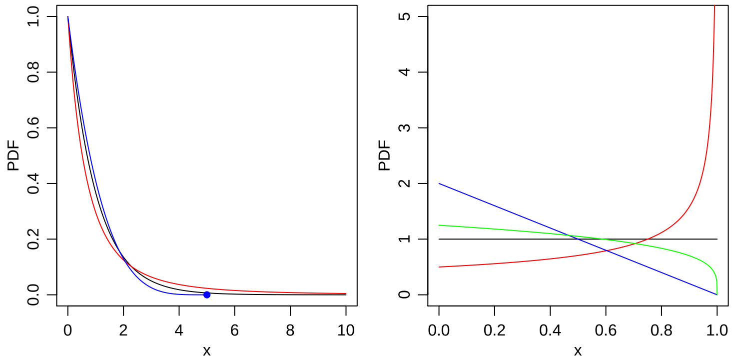
  where $\xi \in \mathbb{R}$ and $\sigma = \tau + \xi(u - \eta) > 0$.
- The GPD can be derived from the ratio of the limiting measures

$$1 - H(x) = \Lambda(x + u)/\Lambda(u), \quad x > 0.$$

- Use of this approximation is called **peaks over threshold (POT)** analysis.
- Analogous to the max-stability of the GEV, the GPD is **threshold-stable**, so it is the natural model for exceedances over high thresholds—and under low ones, by replacing $H(x)$ with $1 - H(-x)$.
- Extrapolation to higher levels is analogous to the GEV.

slide 16

Generalized Pareto distribution



- A flexible distribution whose density can take a variety of shapes.
- Left: exponential density ($\xi = 0$, black), heavy-tailed density ($\xi = 0.5$, red) and light-tailed density ($\xi = -0.2$, blue, with upper terminal shown); all have $\sigma = 1$.
- Right: densities with negative shape parameter and upper terminal $x_F = 1$, with $\xi = -1$ (black), $\xi = -2$ (red), $\xi = -0.5$ (blue) and $\xi = -0.8$ (green).

slide 17

Shape parameter ξ

- ξ has characteristic ranges for different types of data:
  - for rainfall, typically $\xi \approx 0.1$,
  - for extreme hot or cold temperatures, typically $\xi \approx -0.2$,
  - for wind speeds, typically $\xi < 0$,
  - for athletics data, typically $\xi < 0$,
  - for negative financial returns, typically $\xi > 0$, maybe even $\xi > 1$ for very risky assets.
- ξ is difficult to estimate, and its uncertainty dominates extrapolation, so if possible we
  - combine data from different (but compatible!) sources to reduce uncertainty, or
  - use Bayesian methods, if good prior information is available.
- The rth moment of the GEV exists only if $\xi < 1/r$, so the mean exists only if $\xi < 1$, the variance only if $\xi < 1/2$, etc.
- In applications (particularly in finance) some moments may not exist.
- MLE does not have its usual properties if $\xi \le -1/2$.

slide 18

Statistical implementation

- The GEV and GPD are limit models, but are used as approximations for maxima for finite block size m and for finite threshold u, so there is a trade-off:
  - taking m too small/u too low gives more data but estimation may be biased,
  - increasing m or u reduces bias, but may give too little data for useful assessment of uncertainty.
- Methods for automatic choice of m or u have been proposed, but may perform badly, so informal (generally graphical) methods are often used.
- Quantile regression is sometimes used to choose u in big datasets.
- Sensitivity analysis is crucial: conclusions should not depend heavily on the choice of m or u.
- Usually fit models using likelihood or Bayes methods, which are flexible and general.

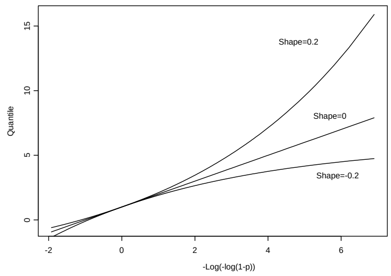
slide 19

Quantiles and return levels

- Assume we use the GEV to model annual maxima, Y.
- Take $0 < p < 1$ and define the GEV quantile y_p by $G(y_p) = p$, giving

$$y_p = \eta + \tau \frac{(-\log p)^{-\xi} - 1}{\xi}.$$

- We call y_p the **return level** associated with the **return period** $1/(1-p)$, so $y_{0.95}$ is the 20-year return level, $y_{0.99}$ is the 100-year return level, etc.
- Is this useful in a non-stationary setting?



slide 20

Estimation or prediction?

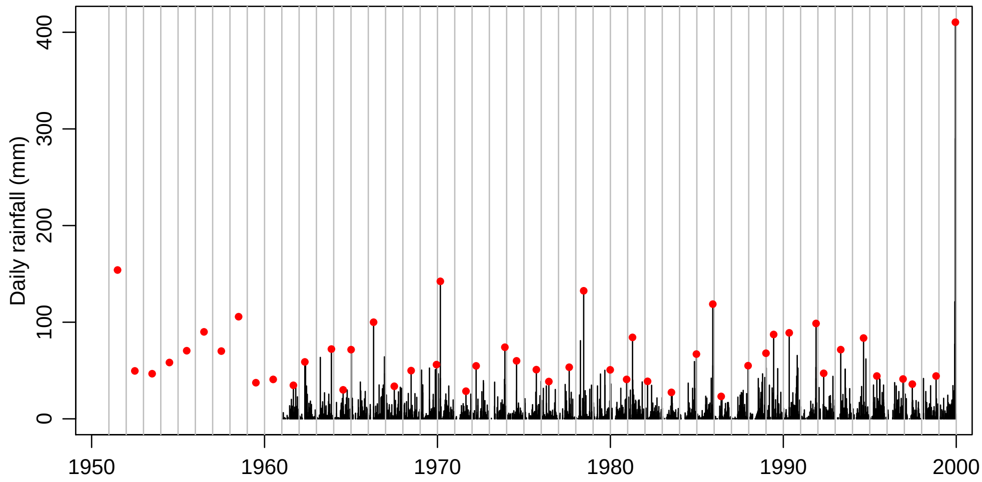
- Return levels, values at risk, expected shortfalls and probabilities would be known if we knew the underlying data generating mechanism—they are parameters to be estimated.
- Very often we are interested in future events, e.g., the largest flood Y_T to be seen in the next T years, which is a random variable—even if we knew the data generating mechanism exactly, we should consider Y_T as random until the T years have passed.
- Should we focus on **prediction** of future events, rather than probabilities for fixed levels?
- In a Bayesian context, this is (in principle) straightforward, we compute the posterior predictive density of Y_T conditional on the observed data $Y = y$, i.e.,

$$f(y_T \mid y) = \frac{\int f(y_T \mid y, \theta) f(y \mid \theta) \pi(\theta)\, \mathrm{d}\theta}{\int f(y \mid \theta) \pi(\theta)\, \mathrm{d}\theta},$$

  where $\pi(\theta)$ is the prior density for parameters θ. We could also compute summaries, such as quantiles of $f(y_T \mid y)$ or its mean $\mathrm{E}(Y_T \mid Y = y)$.
- In a frequentist setting, we may estimate properties of the T-year maximum, such as its median or its expectation (if finite).

slide 21

Annual maxima for Vargas



slide 22

Vargas GEV fit

	1961–1999	1961–1998
Location η	$47.15_{3.77}$	$47.87_{3.73}$
Scale τ	$20.55_{3.29}$	$19.52_{2.92}$
Shape ξ	$0.36_{0.15}$	$0.14_{0.16}$

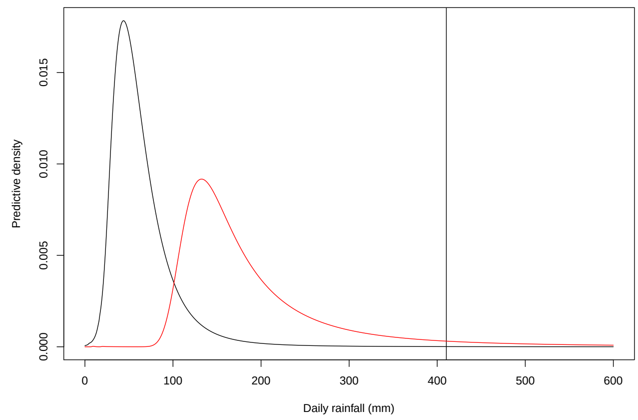
Parameter estimates and standard errors with and without the 1999 maximum.

- Note:
  - the sizes of the standard errors relative to the estimates;
  - the large change in $\widehat{\xi}$ due to dropping the final maximum;
  - the Gumbel distribution ($\xi = 0$) is well inside a 95% confidence interval for ξ if 1999 is dropped, but not otherwise.
- The largest observation has a huge effect on inferences, particularly for ξ.

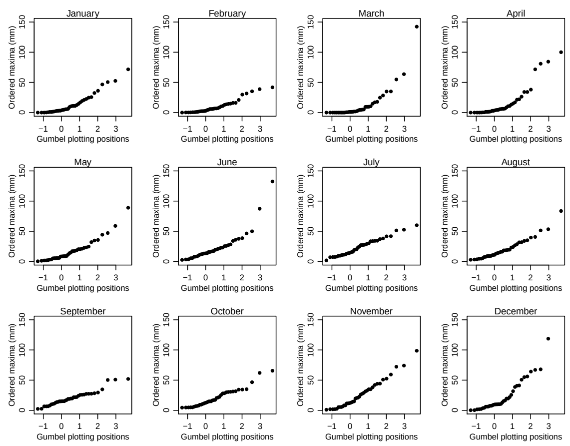
slide 23

Predictive densities

Predictive densities for annual daily maximum (black) and 39-year daily maximum (red), based on data without 1999, with the 1999 daily maximum shown by the vertical line:



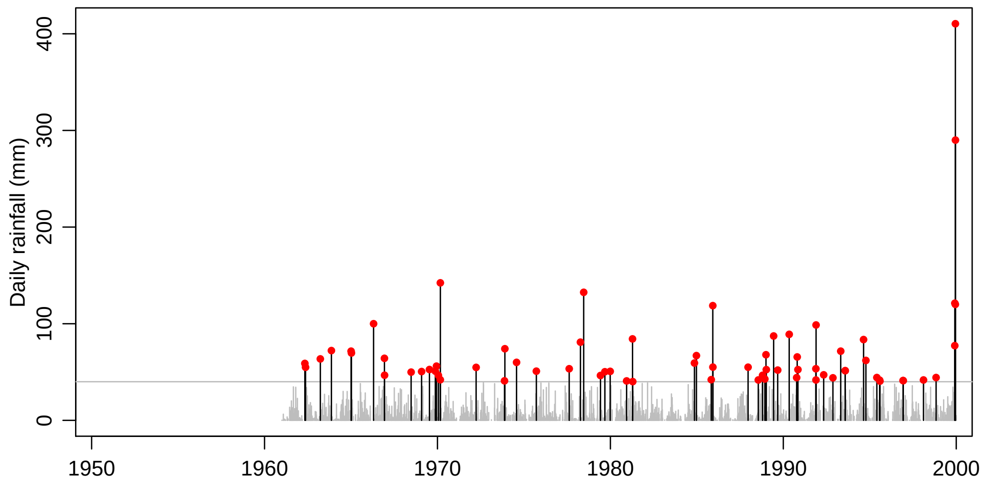
Monthly maxima



Gumbel fits to the monthly maxima would be poor:

- ☐ there can be zero maxima (March, April?);
- ☐ mixtures — some values lie well off the line for the rest;

Peaks over threshold (POT) analysis, $u = 40$mm



slide 26

Exploratory techniques

- The mean of the GPD satisfies

$$\mathrm{E}(X - u \mid X > u) = \frac{\sigma_u}{1-\xi} = \frac{\tau + \xi u}{1-\xi}, \quad \xi < 1.$$

so if the GPD is applicable above some threshold v, a plot of

$$\frac{\sum_{j=1}^{n}(x_j - u)I(x_j > u)}{\sum_{j=1}^{n} I(x_j > u)} \text{ vs } u$$

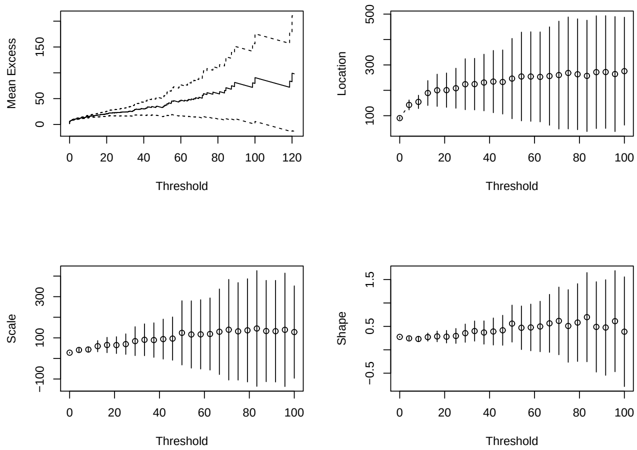
should be a straight line of gradient $\xi/(1-\xi)$ when $u > v$.

- Likewise, if the point process model is appropriate for data above some threshold u, plots of the ML estimates of η, τ and ξ based on data above those thresholds should become constant, above u.
- The main issue with these plots (and many others for extremal data) is that the region of interest usually has too few observations, and hence too wide confidence intervals, to draw firm conclusions.

slide 27

Example: Venezuela rainfall

Mean residual life plot (top left) and parameter stability plots for fits of point process model:



slide 28

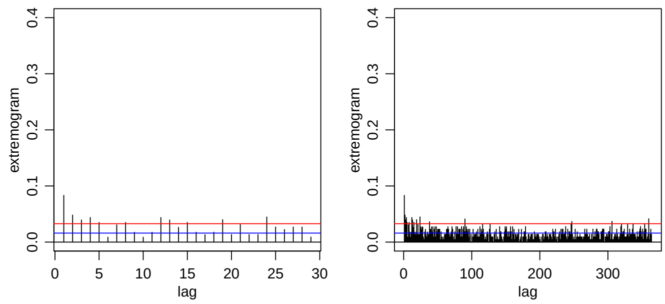
Extremogram

The **extremogram** for a stationary series $\{X_t\}$ estimates

$$\pi_h(u) = \mathrm{P}(X_{t+h} > u \mid X_t > u), \quad h = 1, 2, \ldots.$$

If there is no serial dependence, we should see $\pi_h(u) = \mathrm{P}(X_t > u)$ for all h (blue in picture, upper 95% point is red).

- ☐ This is (almost) the ACF for the time series $I(X_t > u)$;
- ☐ estimated by (almost) the corresponding correlogram;
- ☐ beware poor sampling properties—is there an annual cycle for $X_t > 20\mathrm{mm}$?



slide 29

Comments

- ☐ There is (slight) autocorrelation in large values, leading to some **clustering**, but
  - – under often-plausible conditions on the long-range dependence of extremes, block maxima from a stationary process have limiting GEV distribution
  $$\exp\{-\theta\Lambda(y)\},$$
  where the **extremal index** $\theta \in (0, 1]$, determines the behaviour of clusters;
  - – the basic Poisson process approximation extends to allow clusters of extremes of mean size $1/\theta$ (and otherwise arbitrary configuration).
- ☐ Seasonality in extremes could stem from variation in the numbers of large rainfall days, or in the sizes of large rainfall amounts, or both — need to formulate regression models appropriately if distinguishing these is of interest.
- ☐ Analyzing annual maxima avoids having to deal with any clustering or seasonality.
- ☐ Analysis of the exceedances would allow more detailed modelling of clusters.
- ☐ This dataset ends after the largest event, so there is a **stopping rule**. Ignoring this will bias estimates of risk upwards.

Multivariate extremes

- **Substantive motivation**: many extremal problems are intrinsically multivariate:
  - overwhelming of sea defences by high tides and strong winds;
  - flooding at many locations of a river system;
  - heatwaves have successive very hot days over a wide spatial area.
- **Statistical motivation**: uncertainty may be reduced by combining information from several sources.
- In one dimension it's obvious what is 'extreme', but
  - what is 'extreme' in two or more dimensions?
  - how can we summarize and model extremal dependence?
- One approach is to use a scalar **structure variable**.

slide 32

Structure variable

- A **structure variable** $S = s(X_1, \dots, X_D)$ is a univariate function:
  - for example, insurance loss
  $$S = \sum_{d=1}^{D} a_d(X_d),$$
  where the functions $a_d(\cdot)$ express losses due to risks X_d.
- Then we have a scalar time series $S_1, \dots, S_n$ to which previous ideas apply, using block maxima or threshold exceedances.
- **Advantages**: simple analysis, ignores dependence between $X_1, \dots, X_D$.
- **Disadvantages**:
  - if a new structure variable is introduced, a new analysis is needed—which may disagree with original;
  - missing values of X_d not allowed;
  - don't learn which combinations of $X_1, \dots, X_D$ yield extreme events.

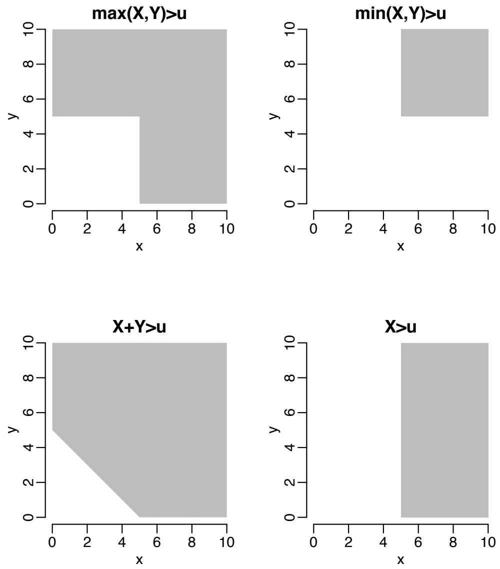
slide 33

Extremes for two variables

- For simplicity, consider bivariate case (X, Y) with the same marginal distributions.
- Given a high threshold u, we might consider any of the following scenarios as extreme:
  - at least one of X and Y exceeds u, i.e., $\max(X, Y) > u$;
  - both X and Y exceed u, i.e., $\min(X, Y) > u$;
  - a function $s(X, Y)$ exceeds u, e.g., $X + Y > u$, though $s(\cdot)$ could also measure distance from some multivariate centre for the data; or
  - given that $X > u$, we consider the distribution of Y, where Y is called a **concomitant** of X; the extremal set is $X > u$.
- There are other possibilities, but these already make life complicated enough.
- The grey regions on the next slide are considered under these four scenarios.

slide 34

Extremes in two dimensions



slide 35

Models for multivariate extremes

- In extension of the univariate case, we ask:

  If non-degenerate limiting distributions exist for maxima of rescaled $(X_1, \ldots, X_D)$, what forms can they have?

- Clearly the limiting margins of suitably rescaled variables must be GEV, so we consider the component-wise transformations

$$\mathcal{P}_n = \left\{ \left(1 + \xi \frac{X_j - b_n}{a_n}\right)_+^{1/\xi} : j = 1, \ldots, \right\} \subset \mathcal{E} = \mathbb{R}_+^D - \{0\},$$

  where $X_1, \ldots, X_n$, $a_n > 0$, b_n and ξ are all $D \times 1$ vectors; we replace $(1 + \xi \ \cdot)_+^{1/\xi}$ by $\exp(\cdot)$ when $\xi = 0$.

- In this case the marginal maxima are unit Fréchet, $\mathrm{P}(Z \leq z) = \exp(-1/z)$, for $z > 0$.
- With this transformation, $\mathcal{P}_n$ converges to a Poisson process $\mathcal{P}$ on $\mathcal{E}$ with mean measure

$$\lim_{n \to \infty} n\mathrm{P} \left\{ \left(1 + \xi \frac{X - b_n}{a_n}\right)_+^{1/\xi} \in \cdot \right\} = \mu(\cdot)$$

  that defines the joint distribution of the maxima (up to marginal transformation).

slide 36

Poisson process limit

- Let $z = (z_1, \ldots, z_D) \in \mathcal{E}$ and let
$$\mathcal{A}_z = \mathcal{E} - [0, z_1] \times \cdots \times [0, z_D].$$
The maximum of $\mathcal{P}_n$ lies below z iff $\mathcal{P}_n \cap \mathcal{A}_z = \emptyset$, and this has limiting probability
$$\mathrm{P}(Z \leq z) = \mathrm{P}(\mathcal{P} \cap \mathcal{A}_z = \emptyset) = \exp\left\{-\mu(\mathcal{A}_z)\right\}, \quad z \in \mathcal{E},$$
where Z denotes the componentwise maximum of the points in $\mathcal{P}$.
- Equivalently we define the **exponent function**
$$V(z_1, \ldots, z_D) = \mu(\mathcal{A}_z),$$
and can show that
  - the marginal unit Fréchet distributions of the Z_d yield $V(z, \infty, \ldots, \infty) = 1/z$ for any permutation of the arguments;
  - the function V is **homogeneous of order -1**, i.e.,
$$V(tz_1, \ldots, tz_D) = t^{-1} V(z_1, \ldots, z_D), \quad z_1, \ldots, z_D > 0, t > 0,$$
which implies **max-stability** of the distribution of Z.

slide 37

Limit distribution of componentwise maxima

- It follows that the limiting distribution of suitably rescaled maxima is of the form
$$\mathrm{P}(Z_1 \leq z_1, \ldots, Z_D \leq z_D) = \exp\left\{-\mu(\mathcal{A}_z)\right\}, \quad z_1, \ldots, z_D > 0,$$
where
$$\mu(\mathcal{A}_z) = V(z_1, \ldots, z_D) = D\mathrm{E}\left\{\max_{d=1}^{D}(W_d/z_d)\right\},$$
and
  - the **angular variable** $W = (W_1, \ldots, W_D)$ lies in the simplex,
$$W \in \mathcal{S}_{D-1} = \{(w_1, \ldots, w_D) : w_d \geq 0, \sum_d w_d = 1\},$$
  - the **angular distribution** ν of W satisfies the marginal constraints
$$\mathrm{E}(W_d) = 1/D, \quad d = 1, \ldots, D,$$
but is otherwise arbitrary.
- Hence (up to marginal transformations) we can model multivariate threshold exceedances using a Poisson process with measure μ, or joint maxima using the distribution $\exp\{-\mu(\cdot)\}$.

slide 38

Extremal functions

- If we write $\mathcal{P} = \{Q_j : j = 1, 2, \ldots\}$ using **extremal functions** Q_j, then

$$Q_j = R_j W_j, \quad R_j > 0, W_j \in \mathcal{S}_{D-1},$$

  where
  - the **pseudo-radii** R_j are points of a Poisson process on $(0, \infty)$ with intensity D/r^2 independent of
  - the **pseudo-angles** $W_j \overset{\text{iid}}{\sim} \nu$,

  and the extremal functions form a Poisson process on $\mathcal{E}$ with intensity

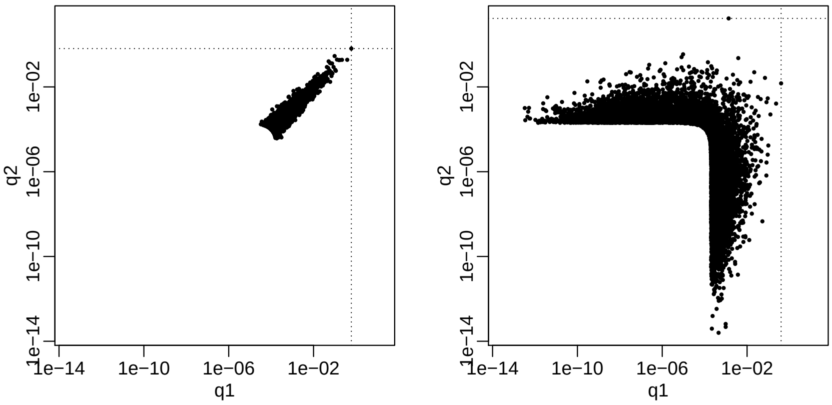
$$\mu(\mathrm{d}q) \equiv \mu(\mathrm{d}r, \mathrm{d}w) = \frac{\mathrm{d}r}{r^2} \times \nu(\mathrm{d}w).$$

- The Q_j represent individual extreme events (storms, heatwaves, ...).
- We can simulate the Q_j by starting with the largest R_j and working downwards.
- The same decomposition applies in more generality, with the W functions lying in a suitable function space.

slide 39

Hüsler–Reiss simulations

Simulated Poisson processes for the Hüsler–Reiss model with $\lambda = 0.5, 4$ (left, right). In each case 10000 points with the largest pseudo-radii have been simulated; the limits appear curved because of the log axes. The intersections of the dotted lines show the componentwise maxima: on the right both arise from a single event, whereas on the left they arise from two separate events.



slide 40

Hüsler–Reiss distribution

- ☐ A natural analogue of the normal distribution in multivariate extremal contexts.
- ☐ The bivariate version has a scalar parameter $\lambda > 0$ and

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{\lambda}{2} + \frac{1}{\lambda}\log\left(\frac{z_2}{z_1}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{\lambda}{2} + \frac{1}{\lambda}\log\left(\frac{z_1}{z_2}\right)\right\}, \quad z_1, z_2 > 0,$$

  where Φ denotes the standard normal cumulative distribution function.
- ☐ Its limits are total independence and total dependence,

$$V(z_1, z_2) \to \begin{cases} 1/z_1 + 1/z_2, & \lambda \to \infty, \\ 1/\min(z_1, z_2), & \lambda \to 0. \end{cases}$$

- ☐ For this model,

$$W^{-1} = 1 + \exp(\lambda\varepsilon + \lambda^2 I/2),$$

  where $I = \pm 1$ with equal probabilities independently of $\varepsilon \sim \mathcal{N}(0, 1)$.

slide 41

Events from subsets of W

- ☐ If V is differentiable, there can be densities on the D-dimensional simplex $\mathcal{S}_{D-1}$ and on each of its sub-faces, defined by setting subsets of the w_d to zero.
- ☐ Hence ν can have $2^D - 1$ components in general — a complicated object!
- ☐ These correspond to particular combinations of extremes and can be viewed as components of a mixture distribution.
- ☐ If $D = 3$, for example, there are three singleton components, three pair components, and one triple component, so perhaps the limiting rare events are

  W_1 only, $\quad$ W_2, W_3 together, or $\quad$ W_1, W_2, W_3 together,

  corresponding to

$$Q_1 > 0, Q_2 = Q_3 = 0, \qquad Q_1 = 0, Q_2, Q_3 > 0, \text{ or } \quad Q_1, Q_2, Q_3 > 0,$$

  with other combinations impossible.
- ☐ In applications we never see $Q_d = 0$, so we have to declare $Q_d \equiv 0$ when $Q_d < \varepsilon$ for some small positive ε.

slide 42

Asymptotic dependence and independence

- All max-stable models are **asymptotically dependent (AD)**, i.e.,

$$\chi(u) = \mathrm{P}\left\{X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)\right\} \to \chi > 0, \quad u \to 1,$$

  or exactly independent if $\chi = 0$.
- In many applications $\chi(u) \to 0$ as $u \to 1$, i.e., the variables are **asymptotically independent (AI)**, and we then use

$$\overline{\chi}(u) = 2\frac{\log \mathrm{P}\{F_2(X_2) > u\}}{\log \ \mathrm{P}\{F_2(X_2) > u, F_1(X_1) > u\}} - 1 \to \overline{\chi}, \quad u \to 1,$$

  to measure the level of AI. The scaling is chosen so that if
  - X and Y are independent, $\overline{\chi} = 0$;
  - if X and Y are perfectly dependent, $\overline{\chi}(u) \equiv 1$;
  - if X and Y are AD, $\overline{\chi} = 1$;
  - $-1 < \overline{\chi}(u) \leq 1$, and $\overline{\chi}$ increases with increasing dependence.
- Both $\chi(u)$ and $\overline{\chi}(u)$ can be estimated non-parametrically from data.
- Can construct AI models from AD ones by **inversion**.
- Models encompassing both AD and AI exist (e.g., Heffernan & Tawn, 2004).

slide 43

Extremal coefficient

- A simple summary of dependence is the **extremal coefficient**,

$$\theta = V(1, \ldots, 1),$$

  which satisfies $\theta = 1$ for perfectly dependent data, and $\theta = D$ for independent data, and is (loosely) interpreted as the 'number of independent maxima' contributing to Z, because

$$\begin{aligned} \mathrm{P}\{\max(Z_1, \ldots, Z_D) \leq z\} &= \mathrm{P}(Z_1 \leq z, \ldots, Z_D \leq z) \\ &= \exp\{-V(z, \ldots, z)\} \\ &= \exp\{-V(1, \ldots, 1)/z\} \\ &= \{\exp(-1/z)\}^{V(1,\ldots,1)}, \quad z > 0, \end{aligned}$$

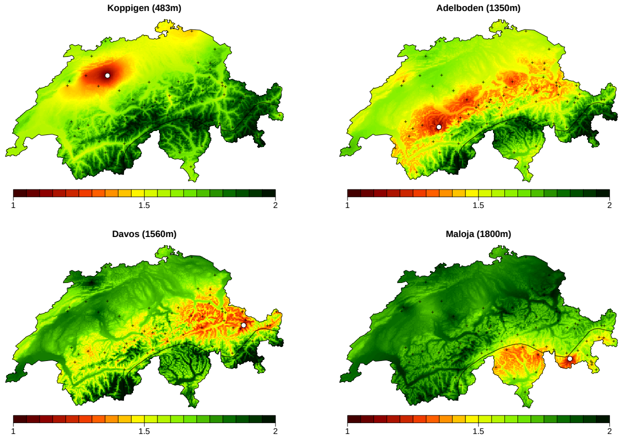
  the distribution of the maximum of $\theta = V(1, \ldots, 1)$ independent Fréchet variables.
- When $D = 2$,

$$\chi = \lim_{z \to \infty} \mathrm{P}(Z_1 > z \mid Z_2 > z) = 2 - \theta, \quad 1 \leq \theta \leq 2,$$

  and $2 - \theta$ is sometimes called the **extremal correlation**.

slide 44

Extremal coefficients for snow depth



Extremal coefficient computed relative to Koppigen, Adelboden, Davos and Maloja (white points), kriged to the whole of Switzerland using a linear trend on absolute altitude difference (Blanchet & Davison, 2011).

slide 45

Concurrence probability

- The **concurrence probability** is the probability p_D that all D components of a multivariate maximum Z arise from a single extreme event, or equivalently that the limiting Poisson process contains a point Q such that $Q = Z$ with probability one.
- This event occurs if $Q = z$ and $\mathcal{A}_z$ is void, or equivalently if the maximum of the other points of the Poisson process is below z, so
$$p_D = \int_{\mathcal{E}} \dot{\mu}(z) \exp\{-\mu(\mathcal{A}_z)\}\, \mathrm{d}z = \mathrm{E}_W \left[\mathrm{E}_{W^*} \left\{ \max_d \left(W_d / W_d^* \right) \right\}^{-1} \right],$$
where $W, W^* \overset{\text{iid}}{\sim} \nu$.
- Total dependence yields $p_D = 1$, whereas independence yields $p_D = 0$.
- p_D has a nice interpretation, but its computation typically involves numerical integration.
- In spatial applications the probability that the same Q leads to extremes at locations s and s' is $p_D \equiv p_2(s, s')$, and
$$I(s) = \int p_2(s, s')\, \mathrm{d}s'$$
measures the mean area of the events leading to extremes at s.

slide 46

Concurrence for high/low temperature extremes

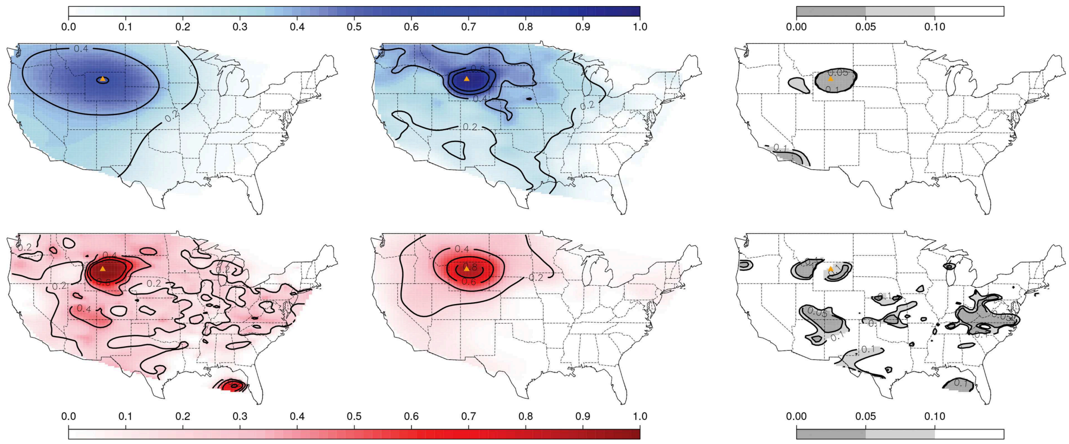


Figure 8. Maps of the extremal concurrence probability for the Worland station (triangle) for the Fall (top) and Spring (bottom) seasons. The left panels show results for the time period 1911–1950 while the middle panels to 1951–2010. The rightmost panels plot the pointwise p-values relative to the hypothesis test of concurrence probabilities difference between the two time periods—see Section 4.5.

From Dombry, Ribatet and Stoev (2018)

slide 47

Concurrence for high/low temperature extremes

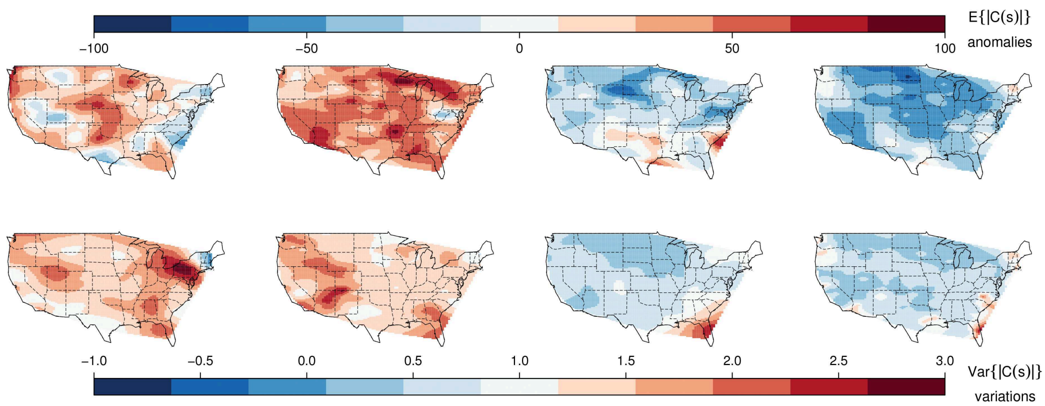


Figure 9. Top: Spatial distribution of the expected extremal concurrence cell areas anomalies, that is, the pointwise difference between $\mathbb{E}\{|C(s)|\}$ for period 1951–2010 and period 1911–1950. Bottom: Spatial distribution of the extremal concurrence cell area variations, that is, the pointwise ratio between var$\{|C(s)|\}$ for period 1951–2010 and 1911–1950. Each column corresponds to one season. From left to right: Fall, Winter, Spring, and Summer.

From Dombry, Ribatet and Stoev (2018)

slide 48

Inference

Typically we

- ☐ fit the GEV or GPD to the margins, allowing for covariates, trends, seasonality . . .;
- ☐ transform the data using the marginal fits $\widehat{\Lambda}_d$, for $d = 1, \ldots, D$;
- ☐ check for AD/AI using $\chi(u)$ and $\overline{\chi}(u)$;
- ☐ fit suitable models to the marginally transformed data,
  - avoiding using exact values of any very small observations;
  - often avoiding full likelihood or Bayesian inference (use pairwise likelihood or other tricks);
  - compare models, check fit, etc.
- ☐ estimate risks/rare event probabilities, . . . ,
- ☐ perform sensitivity analysis as needed.

Better understanding of overall uncertainty if we do all the fitting in a single step (but the computation can be too burdensome).

slide 49

Comments

- ☐ After transformation of the margins to a standard form, the joint distributions have a nonparametric structure imposed by max-stability.
- ☐ Other marginal transformations could be used (e.g., to uniform, Gumbel, exponential, . . . margins); use of unit Fréchet is just for convenience.
- ☐ There are close links between the maximum and point process representations.
- ☐ There are many parametric models for bivariate data, but fewer for $D > 2$.
- ☐ Dependence measures exist:
  - χ and $\overline{\chi}$ measure asymptotic dependence (AD) and asymptotic independence (AI);
  - the extremal coefficient is a scalar summary of dependence, with $\theta = 1$ for fully dependent data and $\theta = D$ for independent data;
  - the concurrence probability is another scalar summary, with a simple interpretation.
- ☐ Inference becomes awkward in (very) high dimensions.

slide 50

Models

- ☐ We consider extremes at the points of a set $\mathcal{X}$ (e.g., an area of a map).
- ☐ We again apply marginal transformations $\Lambda_x(\cdot)$ to a unit Fréchet distribution at each x.
- ☐ Individual multivariate events were represented as

$$Q_j = R_j W_j, \quad W_j \in \mathcal{S}_{D-1} \overset{\text{iid}}{\sim} \nu \perp\!\!\!\perp R_j \sim \text{Poisson process}(1/r),$$

and now the W_j become independent replicates of some spatial process $W(x)$ on $\mathcal{X}$ for which $\mathrm{E}\{W(x)\} = 1$ for every x, so we have

$$Q_j(x) = R_j W_j(x), \quad j = 1, 2, \ldots, \quad x \in \mathcal{X}$$

and $Z(x) = \sup_j Q_j(x)$ for every x.
- ☐ Inference is based on observations at $\mathcal{D} = \{x_1, \ldots, x_D\}$, so follows the same lines as for multivariate data, except that
  - the marginal transformations must pool information over $\mathcal{X}$ for prediction at $x \notin \mathcal{D}$,
  - the distribution of $W(x)$ uses geostatistical ideas (variogram, ...) to model dependence — Brown–Resnick process extends the Hüsler–Reiss model and is particularly useful.

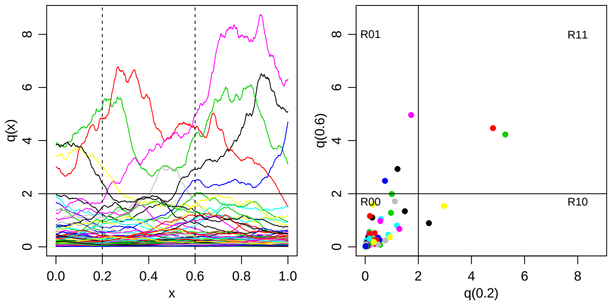
slide 52

Likelihood for events

- ☐ Base extremal modelling on those individual events q falling into extreme set $\mathcal{A}$:
  - allows more detailed modelling and may include more data,
  - if $\mu(\mathcal{A})$ is readily computed, likelihood is

$$\exp\{-\mu(\mathcal{A})\} \times \prod_{q \in \mathcal{A}} \dot{\mu}(q), \quad \dot{\mu}(q) = -\frac{\partial^D V(z_1, \ldots, z_D)}{\partial z_1 \cdots \partial z_D},$$

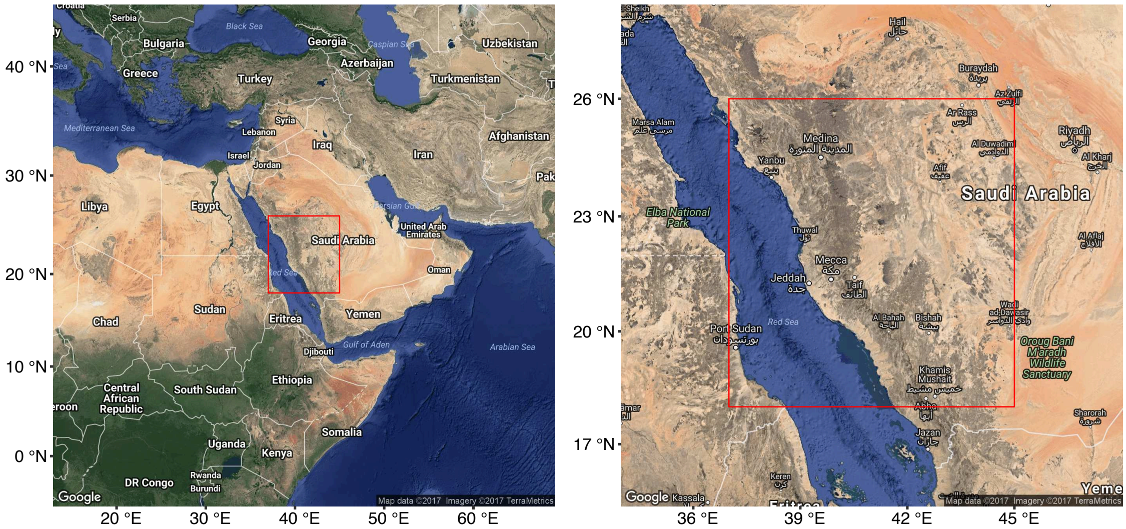
  - but components of some q may be non-extreme, so use a **censored likelihood**.



slide 53

Saudi Arabian rainfall

- Jeddah liable to intense (but rare!) strong convective rainstorms, leading to flash floods, extensive damage and deaths.
- 15-minute radar data available at 750 grid cells over 17 years, so daily annual maxima are space-rich but time-poor.



(Davison et al. 2019)

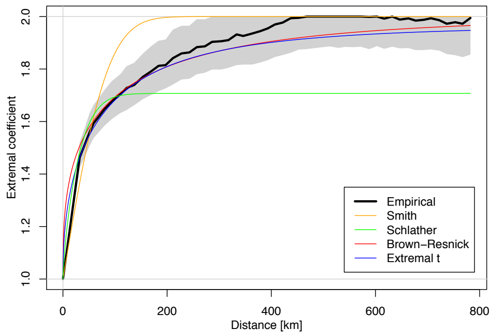
slide 54

Saudi Arabian rainfall

- Censor annual maxima < 3mm.
- Use local likelihood estimates of location and scale parameters, with $\xi \approx 0.14$ constant.
- Transform maxima to standard Fréchet scale, and fit spatial models using censored pairwise local likelihood

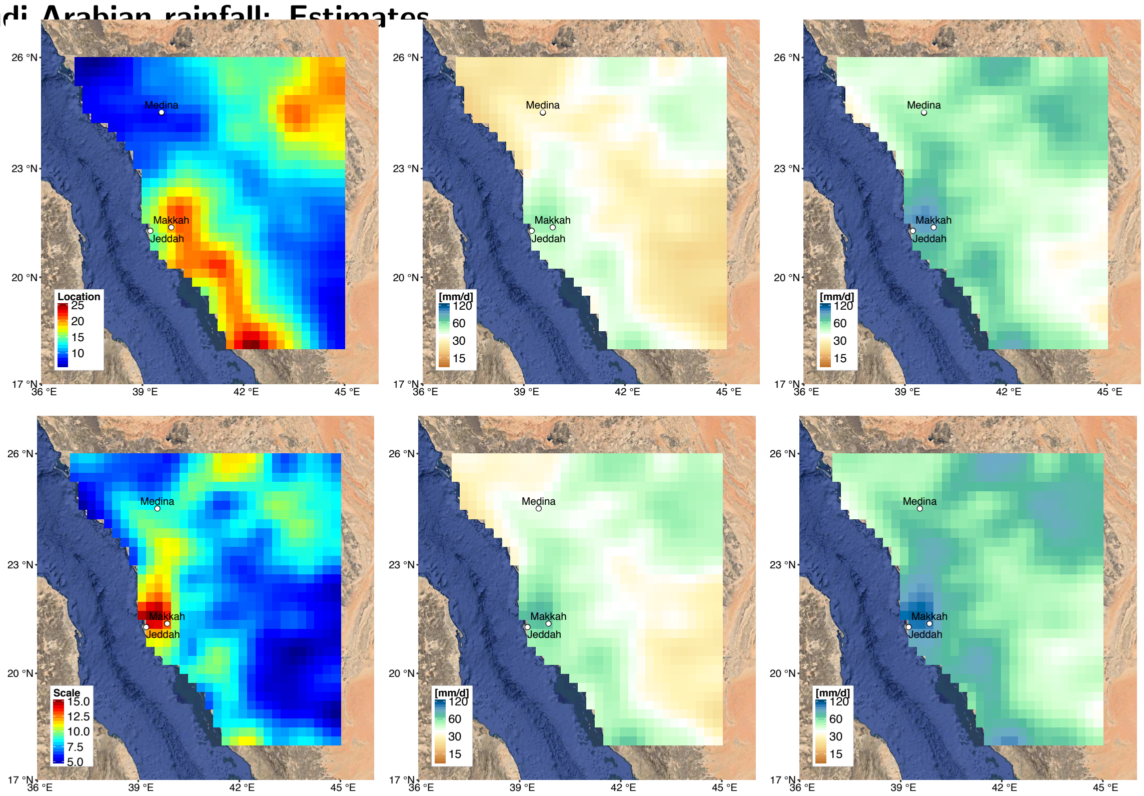
$$\ell(\vartheta) = \sum_{i=1}^{17} \sum_{d'<d} w_{d',d} I(z_{i,d'} > u'_d, z_{i,d} > u_d) \log \left\{ \frac{\exp(-V)(V_1 V_2 - V_{12})}{p(u_{d'}, u_d)} \right\}.$$

- Information criteria suggest reasonable fit of isotopic Brown–Resnick model with variogram $(h/\lambda)^\alpha$, with range $\lambda \approx 13$km range and shape $\alpha \approx 0.7$.



slide 55

Saudi Arabian rainfall: Estimates

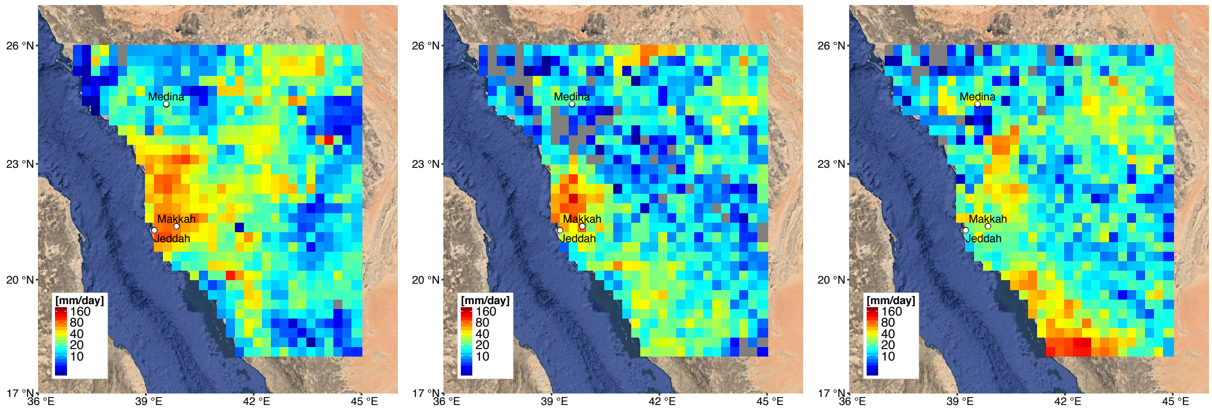


slide 56

Saudi Arabian rainfall: Model fit

Empirical and fitted extremal coefficients $\theta_\mathcal{D} \in [1, D]$ for locations around Jeddah.

Region $\mathcal{D} = \{s_1, \ldots, s_D\}$	D	Empirical $\widehat{\theta}_\mathcal{D}$	Smith	Schl.	B.–R.	Ext.-t
$[39, 40]^\circ\mathrm{E} \times [21, 22]^\circ\mathrm{N}$	14	$4.17_{[1.90,6.44]}$	3.47	3.16	4.41	3.44
$[39, 41]^\circ\mathrm{E} \times [21, 23]^\circ\mathrm{N}$	62	$14.25_{[6.50,22.00]}$	9.73	6.06	11.27	8.71
$[39, 42]^\circ\mathrm{E} \times [21, 24]^\circ\mathrm{N}$	142	$20.90_{[9.54,32.26]}$	19.00	9.01	20.10	15.96



slide 57

Saudi Arabian rainfall: Risk estimation

- Use simulation of individual events to compute probabilities that annual maximum averaged over 14 grid cells $\mathcal{S}$ around Jeddah/Makkah exceeds $v$mm/day, i.e.,

$$p(v) = \mathrm{P}\left\{|\mathcal{S}|^{-1}\sum_{s\in\mathcal{S}} Z(s) > v\right\},$$

  obtaining

$$p(50) = 0.072, \quad p(71.1) = 0.019, \quad p(100) = 0.0048,$$

  with respective return periods around 14, 54 and 208 years.
- Daily rainfall total on 25 November 2009 was 71.1mm/day, leading to 122 deaths.

slide 58

Causality 101

- ☐ The formal study of causality concerns how manipulation of variables (causes) of a set of units affect other variables (effects) of those units.
- ☐ Three broad types of causal statement:
  - **evidence-based mechanism** as in hard science (genetics, physics, ...);
  - **stable association** not explainable by another allowable variable (no confounders); or (intermediate between these)
  - **(potential or actual) experiments** and counterfactuals.
- ☐ Examples:
  - Clausius–Clapeyron mechanism, global heating and rainfall;
  - Granger causality — the predictability of one time series improves (often in information-theoretic terms) if we have access to another time series;
  - attribution of individual events to global heating?
- ☐ Requiring potential for manipulation means that some variables (e.g., time) cannot be causal.

slide 60

Causality 101, II

- ☐ Manipulation of potential causes is sometimes possible (e.g., by randomisation).
- ☐ For a given unit with response Y and possible cause A we would like to know

$$Y_{A=1} - Y_{A=0},$$

where $Y_{A=a}$ is the response when $A = a \in \{0, 1\}$.
- ☐ The key difficulty even in a randomised experiment is that one of $Y_{A=1}$ and $Y_{A=0}$ is unobserved (**counterfactual**), so we need assumptions
  - about the relations between the $Y_{A=1} - Y_{A=0}$ for different units, and
  - to relate observable data on individual units to some average of $Y_{A=1} - Y_{A=0}$.
- ☐ In an **observational study** we need assumptions to link

$$\mathrm{E}(Y_{A=a}) \quad \text{(unobservable)}$$
$$\mathrm{E}(Y \mid A = a) \quad \text{(can be estimated)}.$$

- ☐ A major benefit of formal causal inference is to clarify the background assumptions, whose plausibility can be discussed (and sometimes checked).

slide 61

Causality 101, III

- ☐ Climate scientists often want to compare a real rare event (or series thereof), $Y_{A=1}$, with $Y_{A=0}$, where $a = 1$ corresponds to the real world and $a = 0$ corresponds to an imaginary world without anthropomorphic climate change, e.g., computing

$$\mathrm{P}(Y_{A=1})/\mathrm{P}(Y_{A=0}), \quad \mathrm{P}(Y_{A=1}) - \mathrm{P}(Y_{A=0})$$

  or some variant.
- ☐ Major issues are:
  - both probabilities are typically small, so the details of the computation matter (not using EVT may seriously underestimate them) and the uncertainty may be large;
  - computing $\mathrm{P}(Y_{A=1})$ relies on past data/known mechanisms, so cannot account for regime changes;
  - numerous (but related) climate models could be used to get $Y_{A=0}$, so sensitivity analysis is tricky;
  - the climate system would be non-stationary even with $A = 0$;
  - the event is often defined post-hoc, so there is an element of selection;
  - often the computation takes place just after the event, but this is not accounted for.
- ☐ Much more later ...

slide 62

Closing

Discussion

- ☐ Extreme-value statistics
  - is a well-developed (and growing!) domain of statistics dealing with extrapolation for rare(r) events,
  - relies on point process theory and regular variation (suppressed here),
  - to provide asymptotically-justified models that are never exactly correct,
  - can be fitted (with some effort) to high-dimensional problems/processes, and
  - has many applications to complex problems in climate science.

slide 64

References (mostly review papers)

- ☐ Blanchet & Davison (2011) *Annals of Applied Statistics*.
- ☐ Coles (2001) An Introduction to Statistical Modelling of Extreme Values. Springer.
- ☐ de Haan & Ferreira (2006) Extreme Value Theory. Springer.
- ☐ Dombry, Ribatet & Stoev (2018) *Journal of the American Statistical Assocation*.
- ☐ Davison, Padoan & Ribatet (2012) *Statistical Science*.
- ☐ Davison & Huser (2015) *Annual Review of Statistics and its Applications*.
- ☐ Gelfand, Fuentes, Hoeting & Smith, eds. (2019) Handbook of Environmental and Ecological Statistics. Chapman & Hall/CRC.
- ☐ Huser & Wadsworth (2020) *Wiley Interdisciplinary Reviews (WIREs) Computational Statistics*.

slide 65