Spatial Extremes

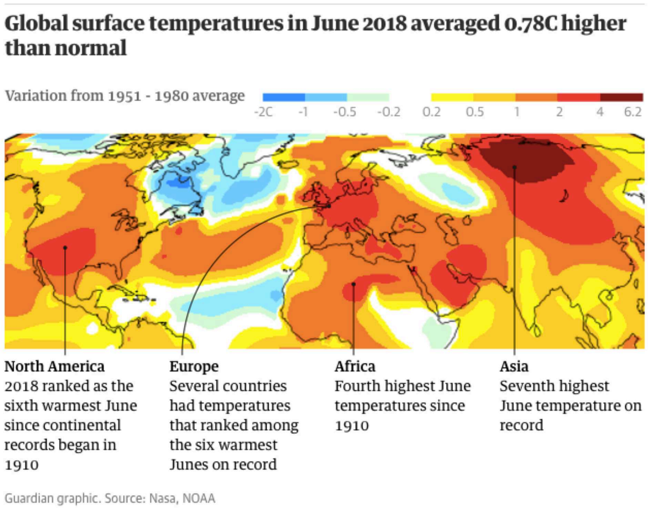
Anthony Davison

Thanks: past PhD students and post-docs, especially Raphaël de Fondeville, Raphaël Huser, Simone Padoan, Mathieu Ribatet, Emeric Thibaud

Funding: Swiss National Science Foundation, Swissnuclear, Swiss Federal Office of the Environment

Summer 2018

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



Motivations for spatial modelling

- ☐ **Attribution** of events to possible causes: to what extent is a heatwave/fire season/flood/. . . due to climate change?
- ☐ **Estimation** of changes in extremes of time series, accounting for dependence between related series.
- ☐ **Risk assessment** at a single important site, borrowing strength from sites nearby.
- ☐ **Risk estimation** for large spatial 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 windstorm or flood?
- ☐ These involve trade-offs:
  - need for accurate space/time interpolation or extrapolation;
  - accurate marginal and/or joint modelling of extreme events.
- ☐ Risk estimation typically involves **extrapolation**:
  - e.g., prediction of 'ten-thousand year event' from 80 years of data

Setup

☐ Focus on extremes of $Y(x)$ for x in some space or space/time domain $\mathcal{X}$.

☐ Often $x = (s, t)$ has spatial component $s \in \mathcal{S}$ and temporal component $t \in \mathcal{T}$, and $\mathcal{X} = \mathcal{S} \times \mathcal{T}$.

☐ Aim to estimate probabilities of the form

$$\mathrm{P}\left\{Y(x) \in \mathcal{R}\right\},$$

where $\mathcal{R}$ is extreme in some sense:

- total rainfall in an upstream catchment large enough to cause major floods downstream;
- daily minimum temperatures over an urban area remain above a danger level for several consecutive days.

☐ Data are available at a finite subset $\mathcal{X}'$:

- a few long series (long-term observations, space-poor/time rich)
- many short series (satellite data, space rich/time poor)
- many longer series (15-minute radar data, space rich/time rich)

Plan

☐ Extreme-value theory: classical models, max-stable processes and their properties, asymptotic dependence and independence; inversion

☐ Latent process models

☐ Max-stable models: a little geostatistics; standard max-stable models and their extreme coefficients

☐ Inference from maxima: graphical methods; joint and separate estimation; pairwise likelihood; example

☐ Inference from exceedances: relation to Poisson process; censored likelihood; gradient score; example

☐ Closing

Poisson process

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

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

  - $N(\mathcal{A}_1), \ldots, N(\mathcal{A}_k)$ 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 (diffuse), and often has an **intensity** $\dot{\mu}$.
- ☐ Think of $\dot{\mu}$ as a PDF that does not have unit integral.
- ☐ **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.

Classical extremal models

- ☐ Use random sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ and for $b_n \in \mathbb{R}$ and $a_n > 0$ define point process

$$\mathcal{P}_n = \{(X_j - b_n)/a_n : j = 1, \ldots, n\}, \quad \mathcal{E} = \mathbb{R}.$$

- ☐ Then the rescaled maximum $\{\max(X_1, \ldots, X_n) - b_n\}/a_n$ has a non-degenerate limiting distribution iff $\mathcal{P}_n$ converges to a Poisson process with mean measure

$$\Lambda\{(y, \infty)\} = \left[1 + \xi\left(\frac{y - \eta}{\tau}\right)\right]_+^{-1/\xi}, \quad y \in \mathbb{R},$$

  where $u_+ = \max(u, 0)$, and η and τ are location and scale parameters.
- ☐ The shape parameter ξ determines the rate of tail decay, with
  - $\xi > 0$ giving the heavy-tailed (Fréchet) case,
  - $\xi = 0$ giving the light-tailed (Gumbel) case—corresponds to Gaussian data,
  - $\xi < 0$ giving the short-tailed (reverse Weibull) case.
- ☐ Limiting distributions:
  - for maxima, **generalized extreme-value (GEV)**, $G(y) = \exp\{-\Lambda(y)\}$;
  - for excesses over threshold u, **generalized Pareto (GPD)**, $H(y) = 1 - \Lambda(y + u)/\Lambda(u)$.

Extrapolation

- Extreme value theory gives **limiting** models:
  - GEV applies for maxima of an infinite sample,
  - GPD applies for exceedances of an 'infinite' threshold.
- Extrapolation to high levels is based on the fact that the GEV is **max-stable**:

$$G(y)^t = G(b_t + a_t y), \quad t > 0,$$

or equivalently

$$\max(X_1, \ldots, X_t) \stackrel{\mathrm{D}}{=} b_t + a_t X_1$$

for known functions $a_t > 0$ and b_t.
- For the unit Fréchet, GEV(1,1,1), distribution, $e^{-1/z}$, $(z > 0)$, we have $b_t \equiv 0$, $a_t = t$.
- Likewise the GPD is **threshold-stable**.
- Could fit other models, but with weaker mathematical justification.
- In practice we have finite samples/finite thresholds, so the extremal models are approximate and extrapolation may be vulnerable.
- Now generalize the above **extremal paradigm** to complex settings ...

Max-stable processes

- Probability integral transformation gives maxima limiting unit Fréchet distribution, so

$$\max(Z_1, \ldots, Z_n) \stackrel{\mathrm{D}}{=} nZ, \quad n = 1, 2, \ldots$$

- Want processes $Z(x)$ with unit Fréchet margins such that if $Z_1(x), \ldots, Z_n(x) \stackrel{\mathrm{iid}}{\sim} Z(x)$, we can base extrapolation on max-stability property

$$\max\{Z_1(x), \ldots, Z_n(x)\} \stackrel{\mathrm{D}}{=} nZ(x), \quad x \in \mathcal{X}.$$

- Let $W(x)$ be a non-negative random process with $\mathrm{E}\{W(x)\} = 1$ $(x \in \mathcal{X})$, and consider the Poisson process on $\mathbb{R}_+ \times \mathcal{C}_+(\mathcal{X})$:

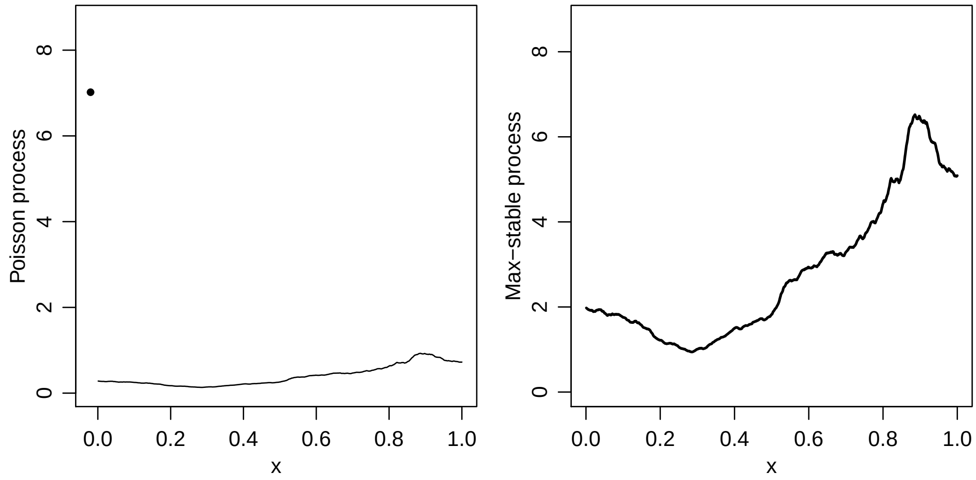
$$\{(R_j, W_j(x)) : j = 1, 2, \ldots, \}, \quad R_j = (E_1 + \cdots + E_j)^{-1}, \quad E_i \stackrel{\mathrm{iid}}{\sim} \exp(1) \perp\!\!\!\perp W_j \stackrel{\mathrm{iid}}{\sim} W.$$

- Setting $Q_j(x) = R_j W_j(x)$ gives a Poisson process on $\mathcal{C}_+(\mathcal{X})$, and any max-stable process has a **spectral representation**

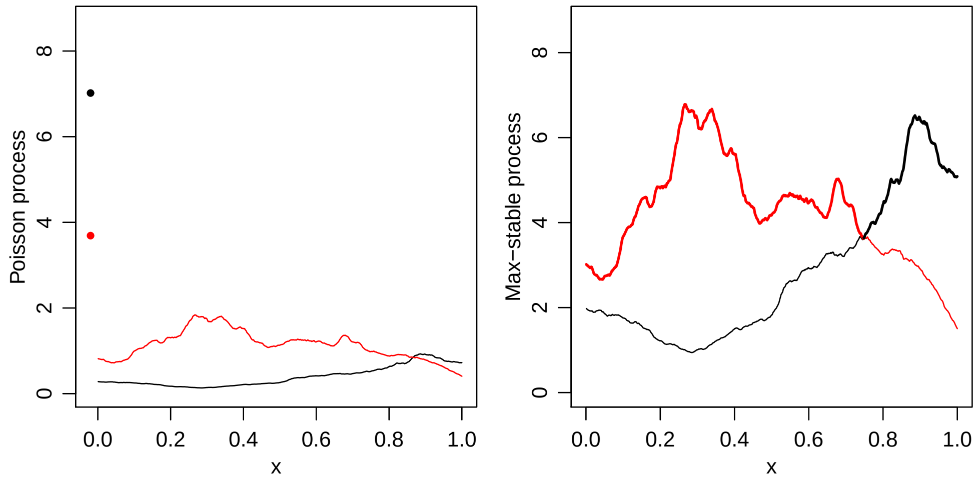
$$Z(x) = \sup_{j=1}^{\infty} Q_j(x), \quad x \in \mathcal{X}, \tag{1}$$

with $Q_j(x)$ interpreted as the jth event, with overall size R_j and profile $W_j(x)$.

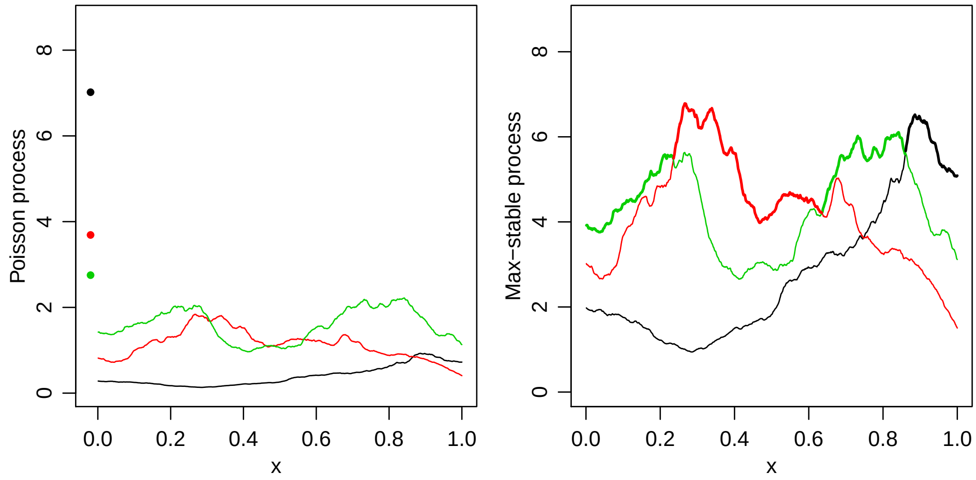
Spectral representation



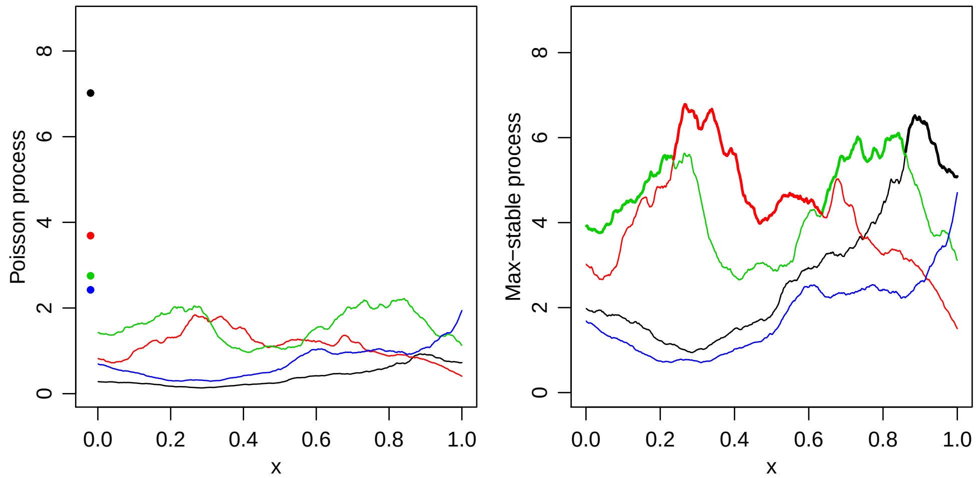
Spectral representation



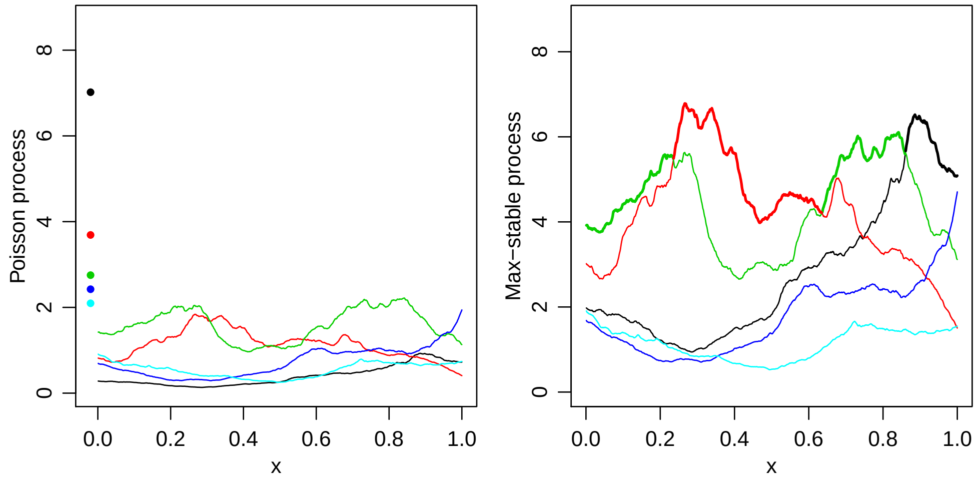
Spectral representation



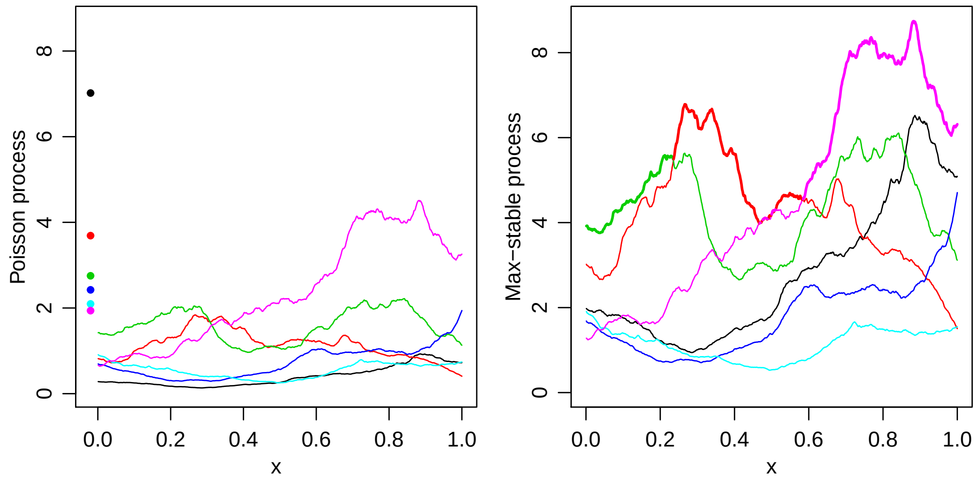
Spectral representation



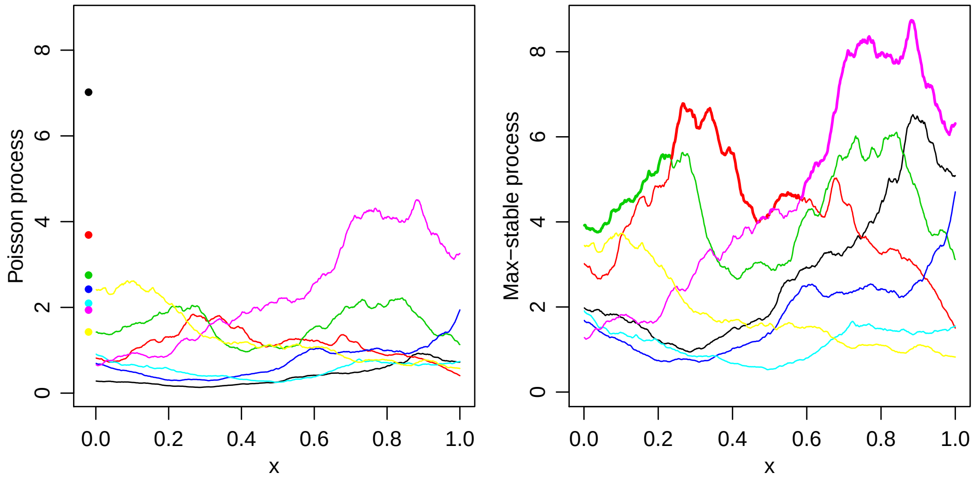
Spectral representation



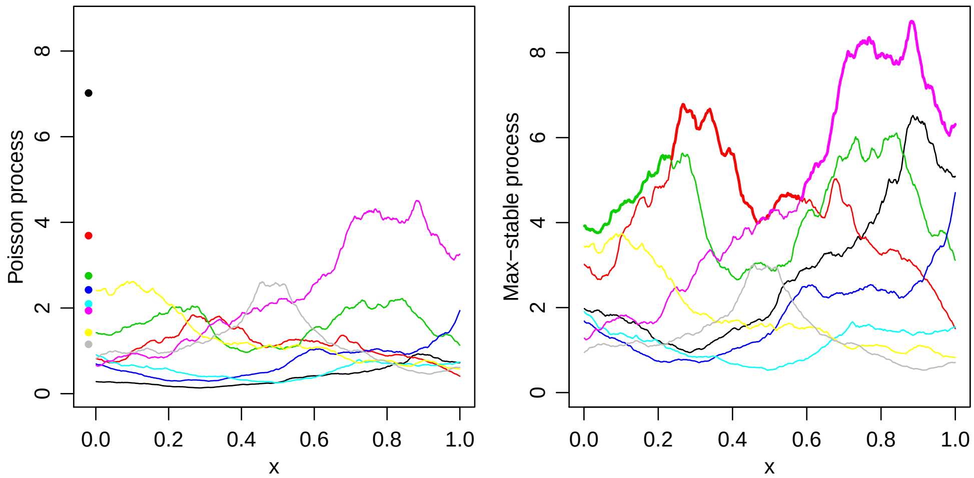
Spectral representation



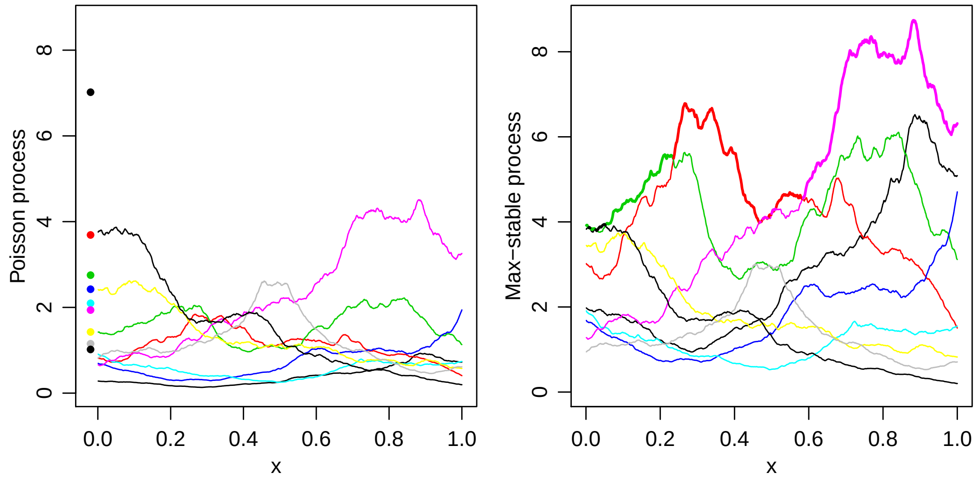
Spectral representation



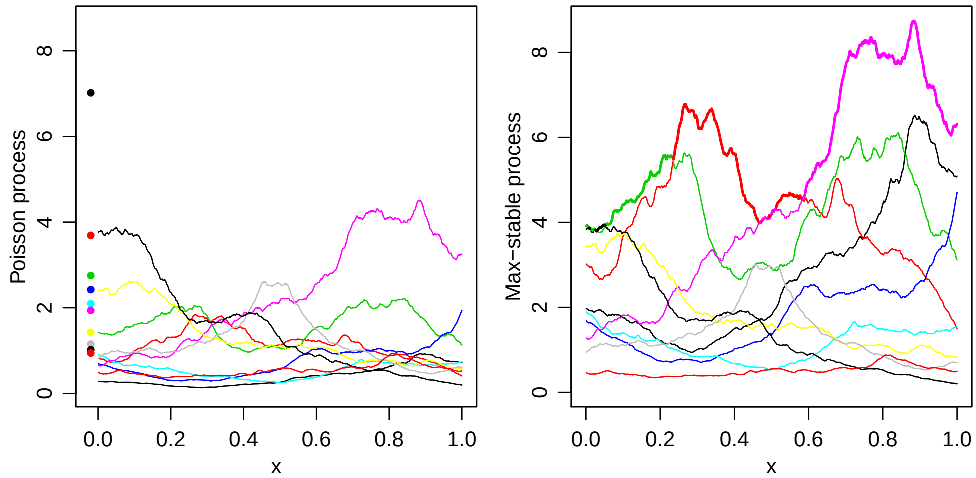
Spectral representation



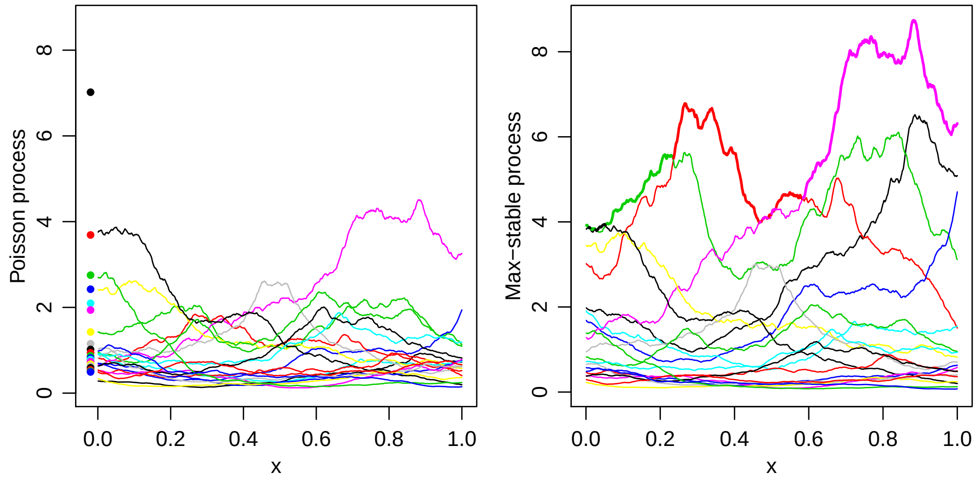
Spectral representation



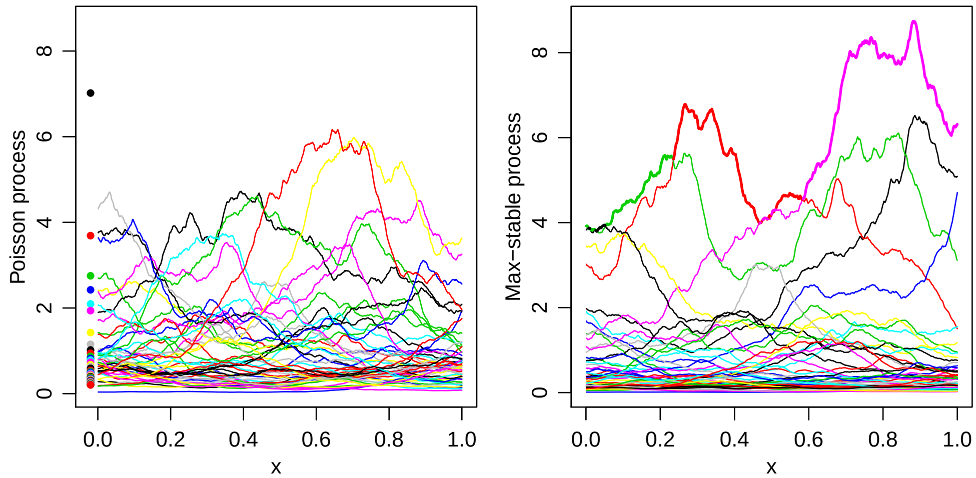
Spectral representation



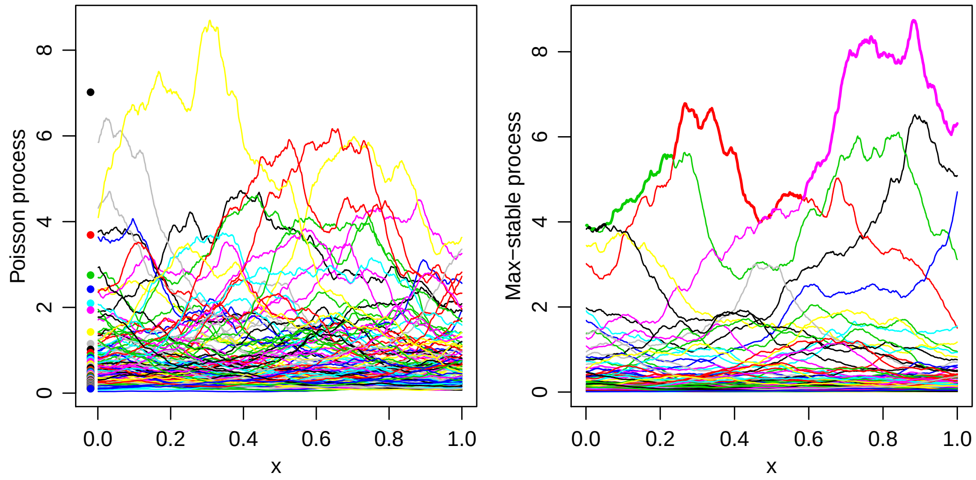
Spectral representation



Spectral representation



Spectral representation



Exponent function

- Extending the argument for scalar maxima, one can show that for a function $z(x)$,

$$\mathrm{P}\{Z(x) \le z(x), x \in \mathcal{D}\} = \exp\left[-V\{z(x) : x \in \mathcal{D}\}\right], \quad \mathcal{D} \subset \mathcal{X},$$

  where the **exponent function**

$$V\{z(x) : x \in \mathcal{D}\} = \mathrm{E}\left[\sup_{x \in \mathcal{D}} \left\{\frac{W(x)}{z(x)}\right\}\right] = \mu[\{q : q(x) \le z(x), x \in \mathcal{D}\}^c]$$

  is derived from the mean measure μ of the Poisson process $\{Q_j\}$, and expectation is over the so-called **angular measure** of W.
- Can show that
  - $Z(x)$ is **simple**, i.e., has unit Fréchet margins, and
  - μ and V are **homogeneous of order** -1, i.e., $\mu(\mathcal{R}) = t \times \mu(t\mathcal{R})$, ($\mathcal{R} \subset \mathcal{E}$, $t > 0$), so we can extrapolate by 'pulling down' extreme risk sets $\mathcal{R}$ to observable levels.
- The case $\mathcal{D} = \{x_1, \ldots, x_D\}$ is key to inference, because data are observed on finite sets, and then we write $z_d = z(x_d)$,

$$V(z_1, \ldots, z_D) = \mu(\mathcal{A}_z), \quad \mathcal{A}_z = ([0, z_1] \times \cdots \times [0, z_D])^c \subset \mathcal{E}' = [0, \infty)^D \setminus \{0\}.$$

Derivation of the exponent function in the process case

- The event $Z(x) \le z(x)$ for all $x \in \mathcal{X}$ is equivalent to

$$R_j W_j(x) \le z(x), \quad x \in \mathcal{X}, j \in \mathbb{N} \quad \Leftrightarrow \quad R_j \le z(x)/W_j(x), \quad x \in \mathcal{X}, j \in \mathbb{N},$$

  or equivalently

$$R_j \le \inf_{x \in \mathcal{X}} z(x)/W_j(x), \quad j \in \mathbb{N}.$$

- Consider (R, W) as a Poisson process taking values in $(0, \infty) \times \mathbb{R}_+^{\mathcal{X}}$ with product measure $\mathrm{d}r/r^2 \times \nu(\mathrm{d}w)$, Then the event $R_j \le \inf_{x \in \mathcal{X}} z(x)/W_j(x)$ for $j = 1, 2, \ldots$ occurs if and only if the set

$$\left\{(r, w) : r > \inf_{x \in \mathcal{X}} z(x)/w(x)\right\}$$

  is void. The corresponding void probability is $\exp[-V\{z(x)\}]$, where

$$\begin{aligned} V\{z(x)\} &= \int \nu(\mathrm{d}x) \int_{\inf_{x \in \mathcal{X}} z(x)/w(x)}^{\infty} \frac{\mathrm{d}r}{r^2} \\ &= \int \nu(\mathrm{d}x) \left[-\frac{1}{r}\right]_{\inf_{x \in \mathcal{X}} z(x)/w(x)}^{\infty} \\ &= \int \nu(\mathrm{d}x) \sup_{x \in \mathcal{X}} \left\{\frac{w(x)}{z(x)}\right\} \\ &= \mathrm{E}\left[\sup_{x \in \mathcal{X}} \left\{\frac{W(x)}{z(x)}\right\}\right]. \end{aligned}$$

- For max-stability, note that with $a_n(x) = n$ and $b_n(x) = 0$, we have

$$\mathrm{P}\left[\max_{j=1}^{n} \left\{\frac{Z_j(x)}{n}\right\} \le z(x), x \in \mathcal{X}\right] = \mathrm{P}\left\{Z_1(x) \le n z(x), x \in \mathcal{X}\right\}^n = \exp\left(-n\mathrm{E}\left[\sup_{x \in \mathcal{X}} \left\{\frac{W(x)}{n z(x)}\right\}\right]\right),$$

  which obviously equals $\mathrm{P}\{Z(x) \le z(x), x \in \mathcal{X}\}$.

- If we set $z(x) = z/\delta(x')$, where $\delta(\cdot)$ is the Kronecker delta, then

$$\mathrm{P}\{Z(x) \le z(x)\} = \mathrm{P}\{Z(x') \le z\} = \exp\left[-\sup_{x \in \mathcal{X}} \mathrm{E}\left\{\frac{W(x)}{z(x)}\right\}\right] = \exp\left[-\frac{\mathrm{E}\{W(x')\}}{z}\right] = \exp(-1/z),$$

  so the process has unit Fréchet margins.

Extremal coefficient

- ☐ Homogeneity of V yields

$$\mathrm{P}\left\{Z(x) \le z, x \in \mathcal{D}\right\} = \exp\left\{-V_{\mathcal{D}}(z)\right\} = \exp\left\{-V_{\mathcal{D}}(1)/z\right\} = \left(e^{-1/z}\right)^{V_{\mathcal{D}}(1)}, \quad z > 0,$$

and the **extremal coefficient**

$$\theta_{\mathcal{D}} = V_{\mathcal{D}}(1)$$

summarises the dependence of extremes within $\mathcal{D}$.

- ☐ The pairwise version,

$$\theta(x_1, x_2) = \mathrm{E}\left[\max\left\{W(x_1), W(x_2)\right\}\right], \quad x_1, x_2 \in \mathcal{X},$$

can be regarded as an analogue of the correlation coefficient, with

$$\text{(total dependence)} \quad 1 \le \theta(x_1, x_2) \le 2 \quad \text{(independence)},$$

and the conditional probability (**extremogram**) interpretation

$$\lim_{z\to\infty} \mathrm{P}\left\{Z(x_2) > z \mid Z(x_1) > z\right\} = 2 - \theta(x_1, x_2).$$

- ☐ $\theta(x_1, x_2)$ is estimated nonparametrically by the **F-madogram** .

Concurrent events

- ☐ Another measure of dependence is the extent to which maxima in $x_{\mathcal{D}}$ occur together, so we define the **concurrence probability** for $\mathcal{D}$ as

$$p(x_{\mathcal{D}}) = \mathrm{P}\left\{Z(x_{\mathcal{D}}) = Q(x_{\mathcal{D}}) \text{ for some } Q(x)\right\}.$$

In terms of the Poisson measure this equals

$$\int_{\mathbb{R}_+^D} \dot{\mu}(q) \exp\left\{-\mu(\mathcal{A}_q)\right\} \mathrm{d}q = \mathrm{E}_{W'}\left\{\frac{1}{V_{\mathcal{D}}(W')}\right\},$$

which can be estimated by simulation.

- ☐ Can likewise define the expected size of a concurrence cell at x as

$$\int_{\mathcal{X}} p(x, x')\,\mathrm{d}x'.$$

Asymptotic dependence and independence

- Max-stable models are **asymptotically dependent (AD)**: at high levels,

$$\lim_{z\to\infty} \mathrm{P}\left\{Z(x_2) > z \mid Z(x_1) > z\right\} = 2 - \theta(x_1, x_2),$$

  either with $\theta(x_1, x_2) \in [1, 2)$ (dependence) or $\theta(x_1, x_2) = 2$ (exact independence).

- In many applications, spatial dependence decreases as the events become rarer, corresponding to **asymptotic independence (AI)**,

$$\mathrm{P}\left\{Z(x_2) > z \mid Z(x_1) > z\right\} \to 0, \quad z \to \infty.$$

- Interpretation: as they become rarer, AD events stay the same size, but AI events shrink.
- An AD model $Z(x)$ can be transformed to an AI model $Z'(x)$ by **inversion**,

$$Z'(x) = -1/\log\left[1 - \exp\left\{-1/Z(x)\right\}\right], \quad x \in \mathcal{X},$$

  so that $\{Z'(x)\}$ has and then

$$\mathrm{P}\left\{Z'(x_2) > z \mid Z'(x_1) > z\right\} \sim L(z) z^{1-\theta(x_1,x_2)}, \quad z \to \infty,$$

  where $1/\theta(x_1, x_2)$ is called the **coefficient of tail dependence** and the function $L(z)$ is **slowly-varying at infinity**, i.e., $\lim_{z\to\infty} L(az)/L(x) = 1$ for any $a > 0$.

Latent process models

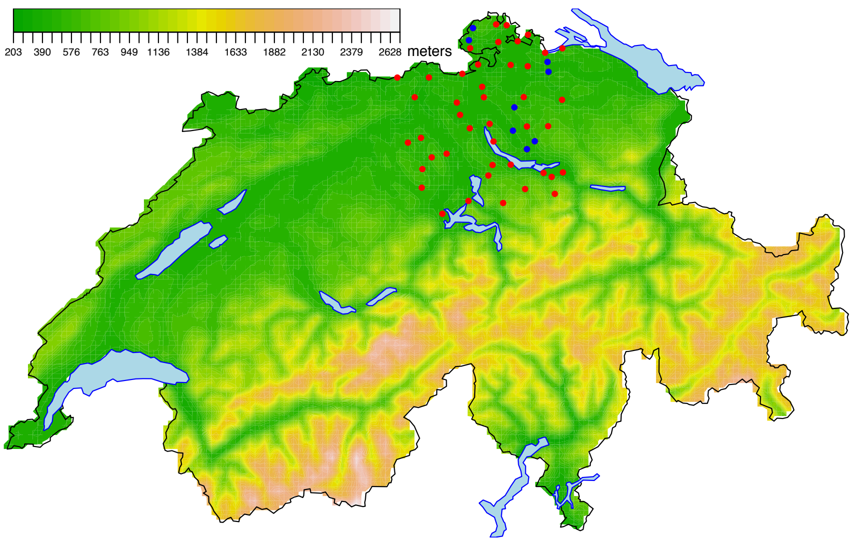
- Conditional on latent process $L(x)$, observations $Y(x)$, for $x \in \mathcal{X}$ follow GEV (or other extremal distribution)
- Example:

$$Y(x) \quad | \quad L(x) = (\eta(x), \tau(x), \xi(x)) \overset{\text{ind}}{\sim} \text{GEV}\{y; L(x)\}, \quad L(x) \sim N_3\{\mu(x), \Omega(x)\}$$

where $\Omega(x)$ represents spatial (usually Gaussian) process allowing variation in GEV parameters.

- **Properties**:
  - + computationally feasible for large-scale problems using MCMC simulation;
  - + possibility of estimating quantiles spatially;
  - + full assessment of uncertainty based on MCMC output;
  - – all extremal dependencies are incorporated through $L(x)$;
  - – marginal distributions are not extremal;
  - – does not capture spatial dependence for episodes.
- Better models based on α-stable random fields can be used at large scale.

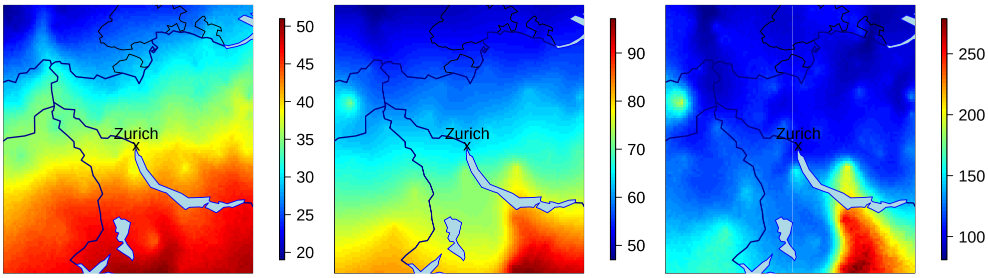
Example: Rainfall in northern Switzerland



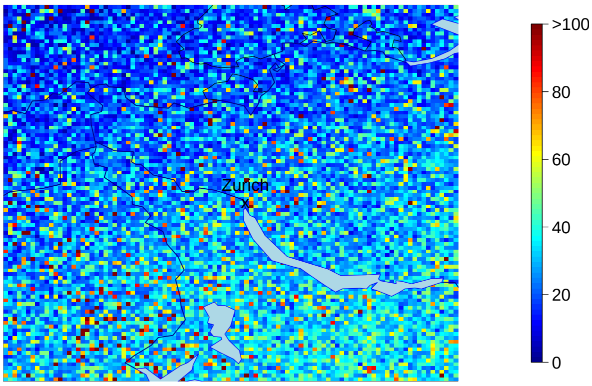
Example: Latent variable model results

Pointwise 25-year return levels (better, 0.96 quantiles) for annual maximum daily rainfall (mm) obtained from a latent variable model.

- ☐ Left and right: estimated 0.025 and 0.975 posterior credible limits for the return levels.
- ☐ Centre: posterior mean return level.



Example: Simulated annual maximum rainfall



- ☐ No spatial pattern, because of conditional independence assumption, so risk estimation is too optimistic—need to include spatial structure of extremes.
- ☐ Can fix this to some extent using copulas, but extremal properties remain inappropriate at very high levels.

Models

- The choice of process W determines event size, orientation, smoothness, etc., subject to the constraints $W \geq 0$ and $\mathrm{E}\{W(x)\} = 1$ for all $x \in \mathcal{X}$.
- Best-known (unfortunately!) is the **Smith model**, which takes

$$W(x) = \phi_k(x - T; \Omega), \quad x \in \mathcal{X} = \mathbb{R}^k,$$

where T is chosen randomly on $\mathcal{X}$ and $\phi_k(x; \Omega)$ is the k-dimensional Gaussian density function with zero mean and covariance matrix Ω.
- The Smith model is too smooth to be realistic, and should not be used in applications.
- Other densities have been proposed, but have the same drawback.
- More realistic models are based on **Gaussian processes**, for which a well-developed theory exists and which are used in many geostatistical applications.
- A random process $\{\varepsilon(x) : x \in \mathcal{X}\}$ is said to be **Gaussian** if all its finite-dimensional distributions are Gaussian, i.e., there exist a **mean function** $\mu(x) = \mathrm{E}\{\varepsilon(x)\}$ and a **covariance function** $\gamma(x_1, x_2) = \mathrm{cov}\{\varepsilon(x_1), \varepsilon(x_2)\}$ such that for any $\{x_1, \dots, x_D\} \subset \mathcal{X}$,

$$(\varepsilon(x_1), \dots, \varepsilon(x_D))^{\mathrm{T}} \sim \mathcal{N}_D \left\{\mu_{D\times 1}, \Omega_{D\times D}\right\},$$

where $\mu_d = \mu(x_d)$ and $\Omega_{cd} = \gamma(x_c, x_d)$.

Stationarity

- For a finite set $\mathcal{D} \subset \mathcal{X}$, let $\varepsilon_{\mathcal{D}}$ and $\mathcal{D} + h$ denote $\{\varepsilon(x) : x \in \mathcal{D}\}$ and $\{x + h : x \in \mathcal{D}\}$, and if it exists, let

$$\gamma(x, x + h) = \mathrm{cov}\{\varepsilon(x), \varepsilon(x + h)\}, \quad x, x + h \in \mathcal{X}.$$

- Then a random process $\{\varepsilon(x)\}$ for which $\mathcal{D} + h \subset \mathcal{X}$ is
  - **strictly stationary** if $\varepsilon_{\mathcal{D}+h}$ and $\varepsilon_{\mathcal{D}}$ have the same (joint) distribution;
  - **weakly stationary** if $\varepsilon_{\mathcal{D}+h}$ and $\varepsilon_{\mathcal{D}}$ have the same first and second (joint) moments;
  - **isotropic** if $\gamma(x, x + h)$ depends only on the distance $\|h\|$.
- For Gaussian processes, the first two coalesce, so we just use the term **stationary**.
- The requirement $\mathrm{var}\{\sum_{d=1}^{D} a_d \varepsilon(x_d)\} = \sum_{c,d} a_c a_d \gamma(x_c, x_d) \geq 0$ for any real $a_1, \dots, a_D$ and any $\{x_1, \dots, x_D\} \subset \mathcal{X}$ constrains possible covariance functions. Here are some isotopic ones valid in any $\mathbb{R}^k$, with scale and smoothness parameters $\lambda > 0$ and κ:
  - the **stable (power exponential)**, $\exp\{-(\|h\|/\lambda)^\kappa\}$, $0 < \kappa \leq 2$;
  - the **Cauchy**, $1/\{1 + (\|h\|/\lambda)^2\}^\kappa$, $\kappa > 0$;
  - the **Matérn**, $(2\sqrt{\kappa}\|h\|/\lambda)^\kappa K_\kappa(2\sqrt{\kappa}\|h\|/\lambda)/\{\Gamma(\kappa)2^{\kappa-1}\}$, $\kappa > 0$, where K_κ is the modified Bessel function of the second kind of order κ.

Intrinsic stationarity

- A stationary process has bounded covariance function, because

$$\mathrm{cov}\{\varepsilon(x+h), \varepsilon(x)\} \le \mathrm{var}\{\varepsilon(x)\} < \infty,$$

  but this is too strong for many applications.
- A process is **intrinsically stationary** if the **increment** $\varepsilon(x+h) - \varepsilon(x)$ has mean zero and variance $2\gamma(h)$.
- We define the **semivariogram** $\gamma(h)$ in terms of the **variogram**

$$\mathrm{var}\,\{\varepsilon(x+h) - \varepsilon(x)\} = 2\gamma(h), \quad x, x+h \in \mathcal{D}.$$

- An intrinsically stationary process has

$$\mathrm{cov}\{\varepsilon(x_1) - \varepsilon(x'), \varepsilon(x_2) - \varepsilon(x')\} = \tfrac{1}{2}\left\{\gamma(x_1 - x') + \gamma(x_2 - x') - \gamma(x_1 - x_2)\right\},$$

  for some $x' \in \mathcal{X}$, and this can be unbounded, which is very useful in applications.
- A standard isotropic semivariogram is the **stable**,

$$\gamma(h) = (\|h\|/\lambda)^{\kappa}, \quad \lambda > 0, 0 < \kappa < 2,$$

  which can yield processes with very different roughnesses and scales of dependence.

Anisotropy and non-stationarity

- There are not many options for anisotropic or non-stationary modelling.
- **Geometric anisotropy** replaces $\|h\|$ in an isotropic covariance function or semivariogram by the **Mahalanobis distance** $h_{\mathcal{M}} = (h^{\mathrm{T}}\Omega h)^{1/2}$. In $\mathbb{R}^2$, for example, we take

$$\Omega = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 & 0 \\ 0 & a^2 \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix},$$

  with degree of anisotropy $a > 0$ and angle $\alpha \in (-\pi, \pi]$.
- For **non-stationarity**, one possibility is to set

$$\mathrm{cov}\{\varepsilon(x'), \varepsilon(x)\} = \sigma^2 |\Omega(x)|^{1/4} |\Omega(x')|^{1/4} \left| \frac{\Omega(x) + \Omega(x')}{2} \right|^{1/2} \rho(h^{1/2}), \quad x, x' \in \mathcal{X},$$

  where
  - $\Omega(x)$ is any symmetric positive definite matrix that varies with x,
  - ρ is any valid correlation function,
  - $h = (x' - x)^{\mathrm{T}}\{\Omega(x') + \Omega(x)\}(x' - x)/2$ is a Mahalanobis distance between x and x'.

  Often ρ is taken to be the Matérn function.

Popular models

- The **Brown–Resnick process** takes $\varepsilon(x)$ to be stationary or intrinsically stationary, and sets

$$W(x) = \exp\left[\varepsilon'(x) - \mathrm{var}\{\varepsilon'(x)\}/2\right],$$

  where (in the intrinsically stationary case) $\varepsilon'(x) = \varepsilon(x) - \varepsilon(x')$ for some $x' \in \mathcal{X}$ and (in the stationary case) $\varepsilon'(x) = \varepsilon(x)$.
- **Extremal t** processes take $\varepsilon(x)$ stationary and set

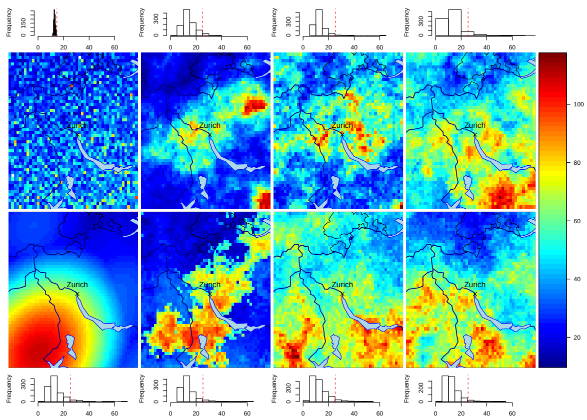
$$W(x) \propto \varepsilon(x)_+^\alpha, \quad \alpha > 0,$$

  where the constant of proportionality gives $\mathrm{E}\{W(x)\} = 1$.
- $\alpha = 1$ gives the **Schlather process**.
- **Smith model** has $W(x) = \phi_k(x - T; \Omega)$, where T is uniformly chosen on $\mathcal{X}$ and $\phi_k(\cdot; \Omega)$ is the k-dimensional Gaussian density with covariance matrix Ω.
- Skew-Gaussian, skew-t, and hierarchical processes can also be constructed.
- One way to obtain long-range independence is to use a random set $\mathcal{B}$, e.g.,

$$W(x) \propto I(x \in \mathcal{B})\varepsilon(x)_+,$$

  which induces dependence only for points x_1, x_2 both lying within $\mathcal{B}$

Realisations from spatial models



Top: latent variable, Student t copula, Hüsler–Reiss copula and extremal-t copula models. Bottom: Smith, Schlather, geometric Gaussian and Brown–Resnick models. The histograms are of 1000 realisations of a summary of rainfall centred on Zürich, and the vertical lines correspond to the realizations shown.

Brown–Resnick model

- ☐ The bivariate exponent function equals

$$V_{\mathrm{BR}}(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{z_2}{z_1}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{z_1}{z_2}\right)\right\},$$

where Φ is the standard normal CDF, $a = \sqrt{\gamma(x_1, x_2)}$.
Hence

$$\theta_{\mathrm{BR}}(x_1, x_2) = 2\Phi\left[\left\{\frac{\gamma(x_1, x_2)}{2}\right\}^{1/2}\right].$$

- ☐ θ_{BR} can attain 2 only as $\gamma \to \infty$, so modelling dependence at long ranges can be awkward.

Extremal t model

- ☐ Defined by setting $W(x) \propto \varepsilon(x)_+^{\alpha}$ for some $\alpha > 0$.
- ☐ The bivariate exponent function $V_{\mathrm{ET}}(z_1, z_2)$ equals

$$\frac{1}{z_1}T_{\alpha+1}\left\{-\frac{c}{b} + \frac{1}{b}\left(\frac{z_2}{z_1}\right)^{1/\alpha}\right\} + \frac{1}{z_2}T_{\alpha+1}\left\{-\frac{c}{b} + \frac{1}{b}\left(\frac{z_1}{z_2}\right)^{1/\alpha}\right\}, \quad z_1, z_2 > 0,$$

where $T_\nu(\cdot)$ is the cumulative distribution function of the Student t distribution with ν degrees of freedom, and

$$c = \mathrm{cov}\{\varepsilon(x_1), \varepsilon(x_2)\}, \quad b^2 = (1 - c^2)/(\alpha + 1).$$

- ☐ Hence

$$\theta_{\mathrm{ET}}(x_1, x_2) = 2T_{\alpha+1}\left[\left\{\frac{(1-c)(1+\alpha)}{(1+c)}\right\}^{1/2}\right].$$

- ☐ For finite α, θ_{ET} is bounded strictly below 2.

Schlather model and random set

- ☐ Extremal t model with $\alpha = 1$ and isotropic correlation function $\rho(h)$, and pairwise exponent function

$$V(z_1, z_2) = \tfrac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \left[1 - 2\frac{\{\rho(h)+1\}z_1 z_2}{(z_1+z_2)^2}\right]^{1/2}\right).$$

- ☐ Corresponding extremal coefficient can only represent positive dependence, as it is bounded in $[1, 1 + 1/\sqrt{2}]$.
- ☐ Taking $W(x) \propto \varepsilon(x)_+ I(x \in \mathcal{B}_y)$ with (bounded) random set $\mathcal{B}_y$ yields

$$V(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left\{1 - \frac{\alpha(h)}{2}\left(1 - \left[1 - 2\frac{\{\rho(h)+1\}z_1 z_2}{(z_1+z_2)^2}\right]^{1/2}\right)\right\},$$

where

$$\alpha(h) = \mathrm{E}\{|\mathcal{B}_y \cap (h + \mathcal{B}_y)|\}/\mathrm{E}(|\mathcal{B}_y|).$$

- ☐ The extremal coefficient for the random set model can take any value in $[1, 2]$.

Generalities

- ☐ Extremal models are always mis-specified—so inferences are likely to be biased.
- ☐ Must check stability of inferences and possible presence of AI, so must vary rarity of chosen events (threshold, ...).
- ☐ Extremal coefficients useful for
  - exploratory analyses based on sub-groups (often pairs) of observation sites $\mathcal{D}$,
  - simple parameter estimates, e.g., by least squares
  - model-checking based on sub-sets of $\mathcal{D}$ (preferably **not used for fitting**).
- ☐ Summarise uncertainty by envelopes based on simulations from fitted models.
- ☐ Semiparametric inference would be preferable in principle, but
  - models are already quite flexible;
  - low power for falsifying models, because data are often limited;
  - simulation is usually needed for risk assessment.
- ☐ Mainly likelihood, but also use gradient scores (later).

Exploratory procedures

- ☐ Exploratory procedures are mostly based on estimates of $\theta(x_1, x_2)$, with the
  - **extremogram** $2 - \widehat{\theta}(x_1, x_2)$ in time series,
  - **madogram** in spatial cases: analogous to variogram for Gaussian processes, which estimates (isotropic) covariance function $\gamma(h)$ as a function of h.
- ☐ Prefer the **F-madogram**,

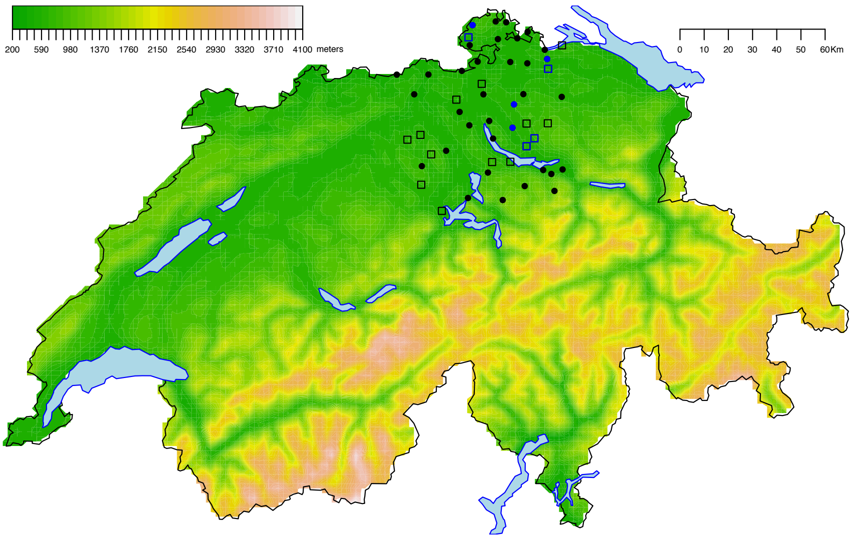
$$\psi_{\mathcal{D}} = \tfrac{1}{2}\mathrm{E}[|F\{Z(x_1)\} - F\{Z(x_2)\}|] = \frac{1}{2}\frac{\theta_{\mathcal{D}} - 1}{\theta_{\mathcal{D}} + 1},$$

  where F is unit Fréchet CDF and $\theta_{\mathcal{D}}$ is the extremal coefficient for $\mathcal{D} = \{x_1, x_2\}$.
- ☐ To estimate $\theta_{\mathcal{D}}$ from independent pairs $(Z_{1,1}, Z_{2,1}), \ldots, (Z_{1,n}, Z_{2,n})$ we replace $F(Z_{1,1})$ by $R_{1,j}/(n+1)$ $(j = 1, \ldots, n)$, where $R_{1,j}$ is the rank of $Z_{1,j}$ among $Z_{1,1}, \ldots, Z_{1,n}$, and likewise with $F(Z_{2,1})$, etc., and then set

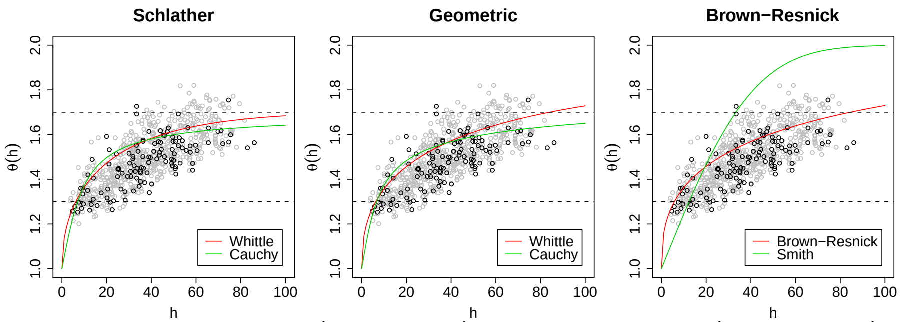
$$\widehat{\psi}_{\mathcal{D}} = \{2n(n+1)\}^{-1} \sum_{j=1}^{n} |R_{1,j} - R_{2,j}|, \quad \widehat{\theta}_{\mathcal{D}} = \frac{1 + 2\widehat{\psi}_{\mathcal{D}}}{1 - 2\widehat{\psi}_{\mathcal{D}}}.$$

- ☐ For a stationary process, $\theta_{\mathcal{D}}$ depends only on $h = x_1 - x_2$, so a plot of the $\widehat{\theta}_{\mathcal{D}}$ for all possible pairs $\mathcal{D}$ can be smoothed or binned as a function of h.

Example: Rainfall in northern Switzerland



Example: Empirical and fitted F-madograms



F-madogram estimates for the fitting (grey points) and the validation (black points) data sets and the estimated extremal coefficient functions for different max-stable models.

Model assessment

- If a max-stable model has been fitted, then for any subset of points $\{x_d : d \in \mathcal{D}\}$,

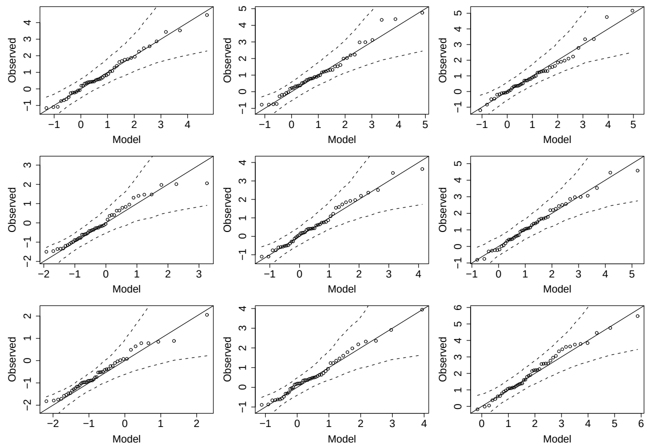
$$\mathrm{P}\{Z(x_d) \le z, d \in \mathcal{D}\} = \exp\{-V_{\mathcal{D}}(z, \ldots, z)\} = \exp(-\theta_{\mathcal{D}}/z), \quad z > 0,$$

  where $\theta_{\mathcal{D}} = V_{\mathcal{D}}(1, \ldots, 1)$, so

$$Z_{\mathcal{D}} = \max\{Z_d : d \in \mathcal{D}\} \sim \text{Fréchet}(\theta_{\mathcal{D}}).$$

- Hence we can compare the empirical distributions of maxima for different subsets $\mathcal{D}$ with those of simulations from a fitted model.
- Assess uncertainty using simulation envelopes.
- If pairs $\{x_1, x_2\}$ are used for fitting, we should use triplets and other subsets for model assessment.

Example: Fit of Brown–Resnick rainfall model



Top row: pairwise maxima estimated from the model and observed maxima for pairs of stations separated by 7 km (left), 45 km (middle) and 83 km (right). Middle row: observed and predicted minima (left), mean (middle) and maxima (right) for a group of 5 validation stations. Bottom row: as middle row, but for all 16 validation stations.

Estimation: Joint or separate?

- ☐ Extremal model involves three steps:
  - fit marginal model for maxima or exceedances at different points of $\mathcal{X}$,
  - transform observed data to standard scale (Fréchet, Gumbel, exponential, Pareto),
  - fit joint model to allow for dependence of transformed variables.
- ☐ Ideally we would perform **joint estimation** of margins and dependence structure, using a likelihood in which both appear.
- ☐ This gives correct standard errors etc., but can be infeasible in large applications.
- ☐ If joint estimation is infeasible, we estimate marginal model first, using **independence likelihood**, which (incorrectly) treats all the observations as independent.
- ☐ Then treat marginal transformation as fixed, transform the observations to unit Fréchet, and estimate the dependence structure.
- ☐ Uncertainty estimates based on separate fitting will be too small, but often this approach is the only one feasible in applications, and we use bootstrap for uncertainty assessment.
- ☐ Balancing good marginal and joint fits can be tricky (easier in some Bayesian formulations).

Marginal fitting

- Let ϑ denote all parameters.
- Extremes $Y(x)$ of original data at $\mathcal{D} = \{x_1, \dots, x_D\}$ will (should!) have GEV/GPD distributions.
- For maxima, use **marginal transformation**

$$Z(x) = \left\{1 + \xi(x;\vartheta)\frac{Y(x) - \eta(x;\vartheta)}{\tau(x;\vartheta)}\right\}_+^{1/\xi(x;\vartheta)}$$

  to the unit Fréchet scale for use with joint model, with
  - splines for space-varying location, scale and shape parameters,

$$\eta(x;\vartheta), \quad \tau(x;\vartheta), \quad \xi(x;\vartheta),$$

  - or local likelihood estimation (later),
  - and, often, constant shape, $\xi(x;\vartheta) \equiv \xi$.
- Similar transformation for exceedances.
- Use standard diagnostics to to check marginal fits (and choose thresholds).

Likelihood for maxima

- Given independent annual maxima observed at $\mathcal{D} = \{x_1, \dots, x_D\}$ for n years, the maxima for each year have joint distribution

$$\mathrm{P}\{Z(x_1) \le z_1, \dots, Z(x_D) \le z_D\} = \exp\{-V(z_1, \dots, z_D)\}, \quad z_1, \dots, z_D > 0.$$

- To compute the likelihood we must differentiate $e^{-V(z_1,\dots,z_D)}$ with respect to $z_1, \dots, z_D$, leading to combinatorial explosion:

$$-V_1 e^{-V}, \quad (V_1V_2 - V_{12})e^{-V}, \quad (-V_1V_2V_3 + V_{12}V_3[3] - V_{123})e^{-V}, \quad \dots,$$

  with about 10^5 terms for $D = 10$. As this is infeasible in applications, we instead
  - use a composite (usually a pairwise) likelihood;
  - use event timings (if known) to determine the required term, e.g., with $D = 3$,

$$(-V_1V_2V_3 \quad + \quad V_{12}V_3 + V_{13}V_2 + V_{23}V_1 \quad - V_{123})e^{-V};$$

  - or base inference on threshold exceedances.
- We need the exponent function V and its derivatives, or the Poisson process intensity $\dot{\mu}$ and its integrals ...

Likelihood for maxima II

- In general we can write the likelihood for observed maxima $z_1, \ldots, z_D$ as

$$\exp\{-V(z)\} \sum_{\Pi \in \mathcal{P}} \prod_{\pi_k \in \Pi} \{-V_{\pi_k}(z)\},$$

  where
  - $\mathcal{P}$ is the set of all partitions of $\mathcal{D} = \{1, \ldots, D\}$,
  - Π denotes a partition with blocks $\{\pi_1, \ldots, \pi_K\}$, and
  - V_{π_k} denotes the partial derivatives of V with respect to the indices in π_k.
- Interpretation: the overall maxima $z_1, \ldots, z_D$ arise from K independent events, and π_k corresponds to maxima that occur together in the k event.
- For the Brown–Resnick process:
  - $V(z) = \sum_{k=1}^{D} z_j^{-1} \Phi_{D-1}(\cdot)$, where Φ_{D-1} is the CDF of a Gaussian vector of dimension $D-1$;
  - V_{π_k} is known, and depends on $\Phi_{D-|\pi_k|}$ (later).
- Similar computations are possible for the extremal t process.
- The terms V_{π_k} also arise when modelling multivariate threshold exceedances.

Pairwise likelihood

- Base inferences on **pairwise log likelihood**

$$\ell_2(\vartheta) = \sum_{\text{block}} \sum_{i>j}^{D} \log f(z_{\text{block},i}, z_{\text{block},j}; \vartheta),$$

  constructed from all distinct disjoint pairs of observations within 'independent' blocks.
- For the score statistic $U_2(\vartheta) = \partial \ell_2(\vartheta)/\partial\vartheta$, we have $\mathrm{E}\{U_2(\vartheta)\} = 0$, so under mild conditions, solving $U_2(\vartheta) = 0$ gives a consistent estimator $\tilde{\vartheta}$ of ϑ.
- We also have

$$\tilde{\vartheta} \doteq \vartheta + \left\{-\frac{\partial U_2(\vartheta)}{\partial \vartheta^{\mathrm{T}}}\right\}^{-1} U_2(\vartheta),$$

  so in large samples,

$$\tilde{\vartheta} \mathrel{\dot\sim} \mathcal{N}\left\{\vartheta, J(\tilde{\vartheta})^{-1} K(\tilde{\vartheta}) J(\tilde{\vartheta})^{-1}\right\},$$

  where $J(\tilde{\vartheta}) = \mathrm{E}\{-\partial U_2(\vartheta)/\partial\vartheta^{\mathrm{T}}\}$ and $K(\vartheta) = \mathrm{var}\{U_2(\vartheta)\}$: a 'sandwich' variance matrix.
- Base model selection on the **composite likelihood information criterion**

$$\mathrm{CLIC} = -2\ell_2(\tilde{\vartheta}) + 2\mathrm{tr}\{K(\tilde{\vartheta}) J(\tilde{\vartheta})^{-1}\}.$$

Ingredients for pairwise fitting

- The pairwise densities for Z_1, Z_2 are of the form

$$f(z_1, z_2; \vartheta) = \left\{ \frac{\partial V(z_1, z_2; \vartheta)}{\partial z_1} \frac{\partial V(z_1, z_2; \vartheta)}{\partial z_2} - \frac{\partial^2 V(z_1, z_2; \vartheta)}{\partial z_1 \partial z_2} \right\} \exp\left\{-V(z_1, z_2; \vartheta)\right\}, \quad z_1, z_2 > 0.$$

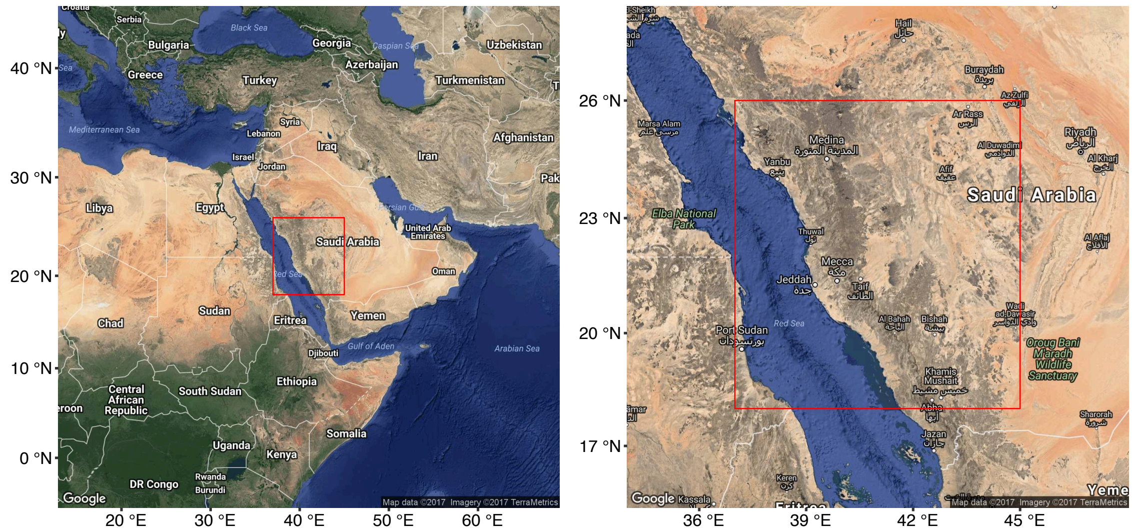
- For the Brown–Resnick, extremal t, Schlather, ..., models, V and its second derivatives are straightforward to compute.
- The density for the original variables $Y(x_1)$, $Y(x_2)$ is

$$\frac{\partial z_1}{\partial y_1} \frac{\partial z_2}{\partial y_2} \times f(z_1, z_2; \vartheta),$$

  where the Jacobian terms also depend on the marginal parameters in ϑ.
- Can't fit complex marginal structures using pairwise likelihood.

Example: 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.



Saudi Arabian rainfall: Marginal fitting

- The domain is spatial, $s \in \mathcal{S}$, with 17 replicates of annual maxima for each cell.
- Annual maximum rainfall $Z(s)$ is zero in some cells, so assume values over $u = 3$mm follow GEV, i.e.,

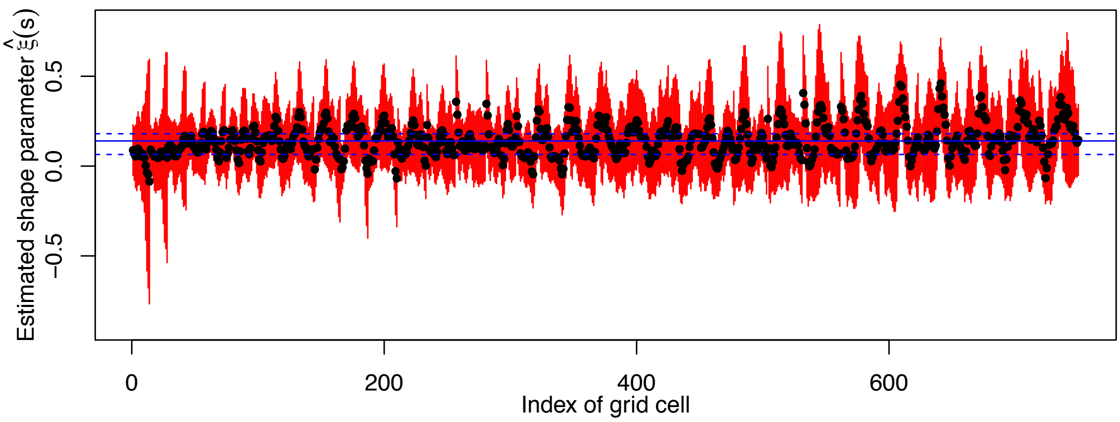
$$\mathrm{P}\{Z(s) \le z\} = G\{z; \eta(s), \tau(s), \xi(s)\} = \exp\left[-\left\{1 + \xi(s)\frac{z-\eta(s)}{\tau(s)}\right\}_+^{-1/\xi(s)}\right], \quad z > u.$$

- Use **local independence (log) likelihood** parameter estimates, which maximise

$$\ell_{s_0}(\eta, \tau, \xi) = \sum_{i=1}^{17} \sum_{d \in \mathcal{N}(s_0)} \omega(\|s_d - s_0\|) \log g_u(z_{i,d}; \eta, \tau, \xi), \quad s_0 \in \mathcal{S},$$

where
  - $z_{i,d}$ is the ith observed annual maximum observed at the dth nearby station,
  - the set $\mathcal{N}(s_0)$ indexes grid cells within a small neighborhood of s_0,
  - the weight function $\omega(h)$ depends on the distance $h = \|s_d - s_0\|$, and
  - $g_u(z; \eta, \tau, \xi)$ is the censored GEV likelihood contribution, $g(z; \eta, \tau, \xi)$ if $z > u$, and $G(u; \eta, \tau, \xi)$ if $z \le u$, where g denotes the GEV PDF.

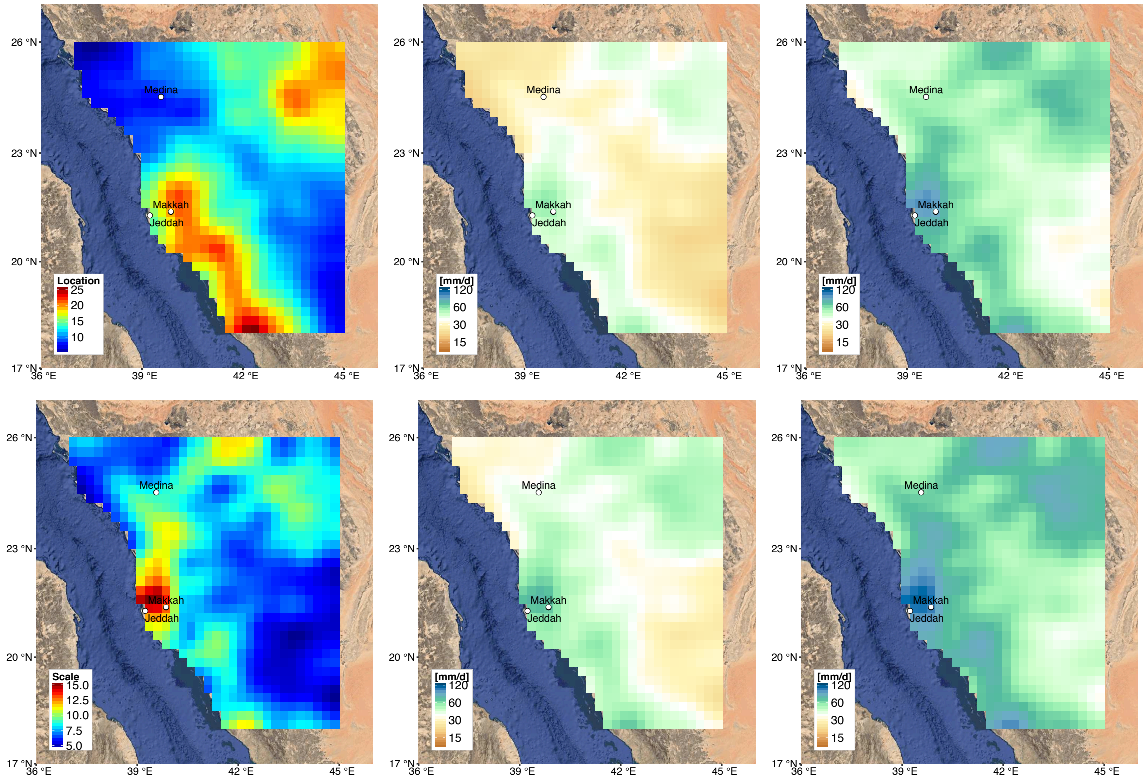
Saudi Arabian rainfall: Shape parameter



- Estimated GEV shape parameter $\widehat{\xi}(s)$ at all grid cells (black dots) with 95% confidence intervals (red segments), from a local likelihood fit using a biweight function $\omega(h) = \{1 - (h/b)^2\}_+^2$ with $b = 80$km.
- Estimates are highly correlated across grid cells. The horizontal blue lines show $\widehat{\xi}$ (solid) when assumed to be constant over the study region, with 95% confidence intervals.
- Next page shows estimated location and scale, and M-year return levels

$$z_M(s) = G^{-1}\{1 - 1/M; \eta(s), \tau(s), \xi(s)\}, \quad s \in \mathcal{S}, \quad M = 10, 20, 50, 100.$$

Saudi Arabian rainfall: Marginal estimates



stat.epfl.ch slide 60

Saudi Arabian rainfall: Joint fit

- Transform maxima to unit Fréchet scale, and fit spatial models using censored pairwise local likelihood

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

  with a random selection of 5100 pairs of locations less than 800km apart, and conditioning on observed maxima being greater than $u = 3$mm.

- Fit four stationary isotropic models:
  - Brown–Resnick process with variogram $\gamma(s_1, s_2) = (\|s_1 - s_2\|/\lambda)^\kappa$, $\lambda > 0$, $0 < \kappa < 2$;
  - Schlather process with correlation function $c(s_1, s_2) = \exp\{-\gamma(s_1, s_2)\}$;
  - extremal t process with same correlation and degrees of freedom $\alpha > 0$; and
  - Smith process with Gaussian density kernels defined through the diagonal covariance matrix $\Omega = \lambda^2 I_2$ ($\equiv$ BR process with $\kappa = 2$).
- We also fitted geometrically anisotropic models, obtained by replacing the Euclidean distance $\|s_1 - s_2\|$ in the models above by the Mahalanobis distance $h_{\mathcal{M}}$.

stat.epfl.ch slide 61

Saudi Arabian rainfall: Pairwise fit

Isotropic max-stable models

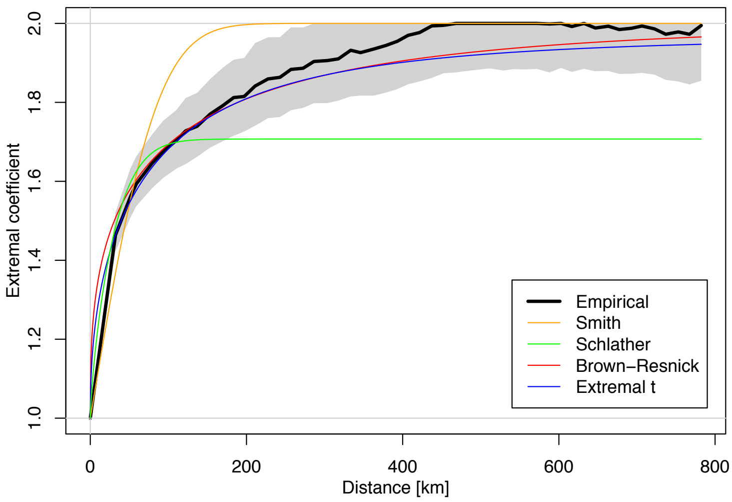
Model	λ [km]	κ	α	a	θ	CLIC
Smith	$34_{[26,39]}$					124
Schl.	$44_{[34,53]}$	$1.46_{[1.19,1.84]}$				362
B.–R.	$13_{[8,16]}$	$0.71_{[0.52,0.94]}$				23
Ext.-t	$333_{[165,1357]}$	$0.90_{[0.63,1.13]}$	$5.9_{[3.9,13.1]}$			0

Anisotropic max-stable models

Model	λ [km]	κ	α	a	θ	CLIC
Smith	$31_{[28,37]}$			$0.82_{[0.72,1.33]}$	$0.19_{[-1.48,2.02]}$	119
Schl.	$42_{[31,54]}$	$1.47_{[1.20,1.82]}$		$0.89_{[0.70,1.28]}$	$0.23_{[-0.32,1.26]}$	362
B.–R.	$12_{[7,21]}$	$0.72_{[0.53,0.95]}$		$0.71_{[0.52,1.81]}$	$-0.12_{[-0.25,1.41]}$	21
Ext.-t	$424_{[176,1352]}$	$0.90_{[0.64,1.10]}$	$6.2_{[4.1,14.8]}$	$1.37_{[0.55,1.66]}$	$1.37_{[-0.30,1.42]}$	41

- No strong evidence against isotropy.
- Extremal t model looks best, but $\widehat{\alpha}$ and $\widehat{\lambda}$ are very unstable, so (maybe) BR model is preferred.
- Smith and Schlather models are worst.

Saudi Arabian rainfall: Joint fit

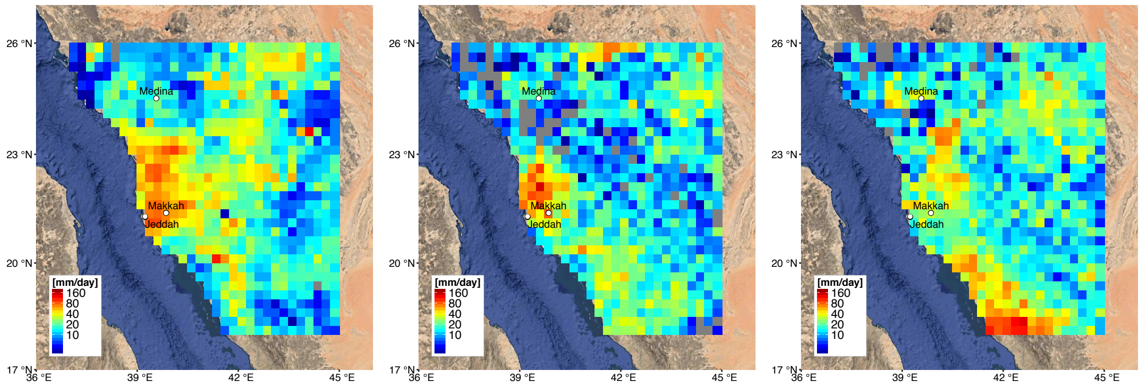


Empirical (black curve) and fitted (colored curves) bivariate extremal coefficients $\theta(s_1, s_2)$, plotted as functions of distance $\|s_1 - s_2\|$, for the isotropic Smith (yellow), Schlather (green), Brown–Resnick (red) and extremal-t (blue) models. The empirical extremal coefficients are binned by distance class, and the grey shaded area is a bootstrap 95% pointwise confidence band.

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
$[39, 40]°\mathrm{E} \times [21, 22]°\mathrm{N}$	14	$4.17_{[1.90,6.44]}$	3.47	3.16	4.41	3.44
$[39, 41]°\mathrm{E} \times [21, 23]°\mathrm{N}$	62	$14.25_{[6.50,22.00]}$	9.73	6.06	11.27	8.71
$[39, 42]°\mathrm{E} \times [21, 24]°\mathrm{N}$	142	$20.90_{[9.54,32.26]}$	19.00	9.01	20.10	15.96



Annual rainfall maxima for 2009 (left), and two simulated maps (middle and right) based on the fitted isotropic Brown–Resnick model. Grey grid cells represent values below $u = 3$mm.

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, 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, leading to 122 deaths—roughly a 50-year event, in a stable climate.

Maxima and exceedances

- Inference may be based on
  - replicates of $\{Z(x) : x \in \mathcal{D}\}$, e.g., annual maximum temperatures at sites in $\mathcal{D}$,
  - individual events $\{Q_j(x) : x \in \mathcal{D}\}$, e.g., hurricanes or droughts.
- Extremal approximations may be better for maxima, but more detailed modelling is possible based on individual events.
- Choose 'extreme' events using **risk functional** ρ and retaining only those falling into

$$\mathcal{E}' = \{q : \rho(q) > 1\}.$$

- Examples involving threshold function $u(x)$:

$$\rho_1(Q) = \sup_{x\in\mathcal{D}} Q(x)/u(x), \quad \rho_2(Q) = \inf_{x\in\mathcal{D}} Q(x)/u(x), \quad \rho_3(Q) = \int_{\mathcal{D}} Q(x)/u(x)\,\mathrm{d}x.$$

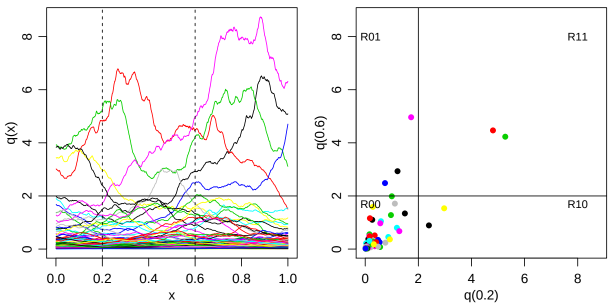
- Inference based on Poisson process likelihood for $\{q_j : q_j \in \mathcal{E}'\}$ involves $\mu(\mathcal{E}')$, which must be finite and computable.
- If $\rho(aQ) = a\rho(Q)$ for $a > 0$, then $\rho(Q) > 1$ gives $R\rho(W) > 1$; then $\mu(\mathcal{E}') = \mathrm{E}\{\rho(W)\}$ depends only on the distribution of W.

Likelihood for events

- Base extremal modelling on those individual events $q(x)$ falling into $\mathcal{E}' = \{q : \rho(q) > 1\}$, where ρ only uses $q(x)$ for $x \in \mathcal{D}$:
  - allows more detailed modelling and may include more data,
  - if Poisson process measure $\mu_\vartheta(\mathcal{E}')$ is computable, likelihood is (in principle),

$$\exp\left\{-\mu_\vartheta(\mathcal{E}')\right\} \times \prod_{q\in\mathcal{E}'} \dot{\mu}_\vartheta(q), \quad \dot{\mu}_\vartheta(q) = -\frac{\partial^D V(q_1,\ldots,q_D)}{\partial q_1 \cdots \partial q_D},$$

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



Censored likelihood

- ☐ In two-dimensional case, $D = 2$, partition $\mathbb{R}_+^2$ into four regions $\mathcal{R}_{I_1,I_2}$, corresponding to the values of the indicators $I_1 = I(Q_1 > u_1)$ and $I_2 = I(Q_2 > u_2)$ of the events that the Q_d exceed thresholds u_d:

$$
\begin{aligned}
\mathcal{R}_{00} &= \{(q_1, q_2) : q_1 \le u_1, q_2 \le u_2\},\\
\mathcal{R}_{10} &= \{(q_1, q_2) : q_1 > u_1, q_2 \le u_2\},\\
\mathcal{R}_{01} &= \{(q_1, q_2) : q_1 \le u_1, q_2 > u_2\},\\
\mathcal{R}_{11} &= \{(q_1, q_2) : q_1 > u_1, q_2 > u_2\},
\end{aligned}
$$

  with the extremal model taken to be fully valid only in $\mathcal{R}_{1,1}$.
- ☐ Then the likelihood contribution from (q_1, q_2) is taken to be
  - $\dot\mu_\vartheta\{(q_1, q_2)\}$ in $\mathcal{R}_{1,1}$,
  - $\int_0^{u_2} \dot\mu_\vartheta\{(q_1, y)\}\,\mathrm{d}y$ in $\mathcal{R}_{1,0}$,
  - $\int_0^{u_1} \dot\mu_\vartheta\{(x, q_2)\}\,\mathrm{d}x$ in $\mathcal{R}_{0,1}$,
  - $\int_0^{u_1}\int_0^{u_2} \dot\mu_\vartheta\{(x, y)\}\,\mathrm{d}x\,\mathrm{d}y$ in $\mathcal{R}_{0,0}$.
- ☐ Obviously the calculations here require that $\mu_\theta(q)$ and its derivatives be readily computable, which is the case for the Brown–Resnick and extremal t models, provided the dimension D is not too high.

Brown–Resnick likelihood

- ☐ If $z_d > u$ for $d = 1, \ldots, C$ and $z_d < u$ for $d \in \mathcal{C}' = \{C+1, \ldots, D\}$, and $\mathcal{C} = \{2, \ldots, C\}$, the censored likelihood contribution has form

$$
\frac{1}{z_1^2 z_2 \cdots z_C} \times \phi_{C-1}(\log \tilde z_{\mathcal{C}}; \tilde\Omega_{\mathcal{C},\mathcal{C}}) \times \Phi_{D-C}\left(\tilde\mu_{\mathcal{C}'|\mathcal{C}}; \tilde\Omega_{\mathcal{C}'|\mathcal{C}}\right),
$$

  where ϕ_k and Φ_k denote the k-dimensional normal density and distribution functions, Ω is defined in terms of the variogram γ, and

$$
\begin{aligned}
\log \tilde z_d &= \log z_d - \log z_1 + \Omega_{d,1}/2, \quad d = 2, \ldots, C,\\
\tilde\Omega_{c,d} &= \tfrac{1}{2}\{\Omega_{c,1} + \Omega_{1,d} - \Omega_{c,d}\}, \quad c, d \in \{2, \ldots, D\},\\
\mu_{\mathcal{C}'|\mathcal{C}} &= (\log u - \log z_1 + \tfrac{1}{2}\Omega_{1,\mathcal{C}'}) - \tilde\Omega_{\mathcal{C}',\mathcal{C}}\tilde\Omega_{\mathcal{C},\mathcal{C}}^{-1}\log \tilde z_{\mathcal{C}},\\
\tilde\Omega_{\mathcal{C}'|\mathcal{C}} &= \tilde\Omega_{\mathcal{C}',\mathcal{C}'} - \tilde\Omega_{\mathcal{C}',\mathcal{C}}\tilde\Omega_{\mathcal{C},\mathcal{C}}^{-1}\tilde\Omega_{\mathcal{C},\mathcal{C}'}.
\end{aligned}
$$

- ☐ Feasible for $D \le 100$, with modified `R` function for Φ.
- ☐ Gradient score needed for higher D:
  - differentiate with respect to data, so normalising constants not needed;
  - use weight function to downweight effects of observations near thresholds.
- ☐ Similar computations are possible for extremal-t processes.

Gradient score

- ☐ Gradient scoring allows statistical inference using derivatives with respect to $q_1, \ldots, q_D$ of the log-density function,

$$
\begin{aligned}
\delta_w(\dot{\mu}_\vartheta, q) \quad = \quad & \sum_{d=1}^{D} \left(2w_d(q) \frac{\partial w_d(q)}{\partial q_d} \frac{\partial \log \dot{\mu}_\vartheta(q)}{\partial q_d} \right. \\
& \left. + w_d(q)^2 \left[\frac{\partial^2 \log \dot{\mu}_\vartheta(q)}{\partial q_d^2} + \frac{1}{2} \left\{ \frac{\partial \log \dot{\mu}_\vartheta(q)}{\partial q_d} \right\}^2 \right] \right),
\end{aligned}
$$

  where $w : \mathcal{E}' \to \mathbb{R}_+^D$ is a weighting function differentiable on the region $\mathcal{E}'$ and vanishing on the boundaries of $\mathcal{E}'$.

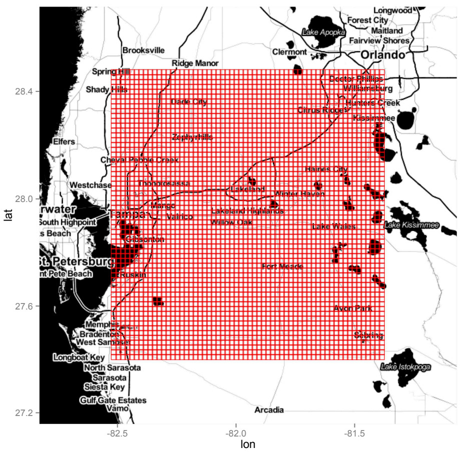
- ☐ The scoring rule $\delta_{\mathrm{w}}(\dot{\mu}_\vartheta, \cdot)$ is strictly proper, i.e., the estimator

$$
\widehat{\vartheta}_\delta = \mathrm{argmax}_\vartheta \sum_{q \in \mathcal{E}'} \delta(\dot{\mu}_\vartheta, q),
$$

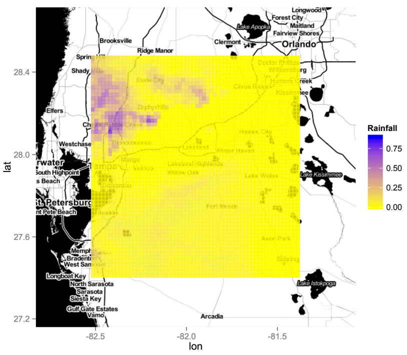
  where q has normalized marginals, is consistent and asymptotically normal as $n \to \infty$ and the threshold $u_n \to \infty$ with $|\{q : q \in \mathcal{E}'\}| = o(n) \to \infty$.

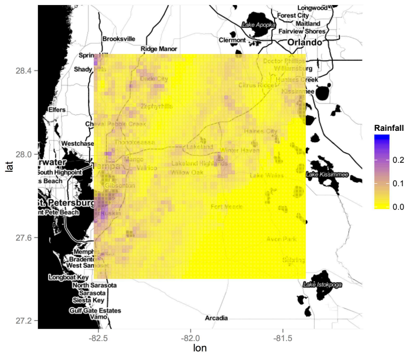
Example: Extreme rainfall over Florida

- ☐ 15-minute radar rainfall measurements over Florida from 1994–2010
- ☐ We focus on a 120 km× 120 km square south-west of Orlando and on the wet season, i.e., June to September.



Example: Florida rainfall





Local rainfall   Spatially dispersed rainfall

Example: Marginal modelling

- Generalized Pareto distributions are fitted for each of the locations s_d using exceedances over the 99 percentile.
- A model with common shape parameter $\xi_0 = 0.124$ is retained.
- Margins are then transformed to unit Fréchet:

$$X^*(s_d) = -1/\log \tilde{F}_d\{X(s_d)\},$$

where

$$\tilde{F}_d\{X(s_d)\} = \begin{cases} \widehat{F}_d\{X(s_d)\}, & X(s_d) \leqslant q_{99}(s_d), \\ 1 - H_{\{\xi_0, \widehat{\sigma}(s_d), q_{99}(s_d)\}}\{X(s_d)\}, & X(s_d) > q_{99}(s_d), \end{cases}$$

and

  – $\widehat{F}_d$ is the empirical cumulative distribution function at location s_d,

  – $H_{\{\xi_0, \widehat{\sigma}(s_d), q_{99}(s_d)\}}$ is the distribution function of a generalized Pareto random variable with shape ξ_0, scale $\widehat{\sigma}(s_d)$ and location $q_{99}(s_d)$.

Example: Risk functionals

- We define two risk functionals

$$\rho_{\max}(X^*) = \left[\sum_{d=1}^{D} \{X^*(s_d)\}^{20}\right]^{1/20}, \quad \rho_{\text{sum}}(X^*) = \left[\sum_{d=1}^{D} \{X^*(s_d)\}^{\xi_0}\right]^{1/\xi_0},$$

where $D = 3600$ is the number of grid cells.

- Here

  – $\rho_{\max}$ is a continuous and differentiable approximation of $\max_{d=1,\ldots,D} X^*(s_d)$ which satisfies the requirements for the gradient score,

  – ρ_{sum} selects events with large spatial cover. The power ξ_0 approximately transforms the data X^* back to a scale where summing observations has a physical meaning.

Example: Spatial model

☐ Non-separable semi-variogram model

$$\gamma(x_d, x_j) = \left\| \frac{\Omega(x_d - x_{d'})}{\tau} \right\|^{\kappa}, \quad x_d, x_{d'} \in [0, 120]^2, \quad d, d' \in \{1, \ldots 3600\},$$

with $0 < \kappa \leqslant 2, \tau > 0$ and anisotropy matrix

$$\Omega = \begin{bmatrix} \cos\eta & -\sin\eta \\ a\sin\eta & a\cos\eta \end{bmatrix}, \quad \eta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right], \quad a > 1.$$

☐ Fitted parameters obtained for both risk functionals with exceedances of $\rho_{\max}(X^*)$ and $\rho_{\text{sum}}(X^*)$ over the 99 quantile:

	κ	τ	η	a
$\rho_{\max}$	$1.192_{0.02}$	$9.06_{0.19}$	$0.08_{0.61}$	$1.008_{0.005}$
ρ_{sum}	$0.326_{0.007}$	$46.67_{0.018}$	$-0.30_{0.10}$	$1.064_{0.017}$

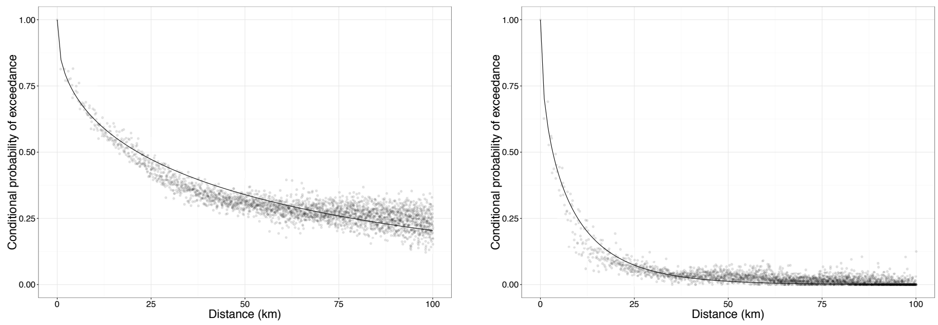
– $\rho_{\max}$ estimates are quite smooth with a small scale, they capture high quantiles and induce a model similar to that in earlier work.

– For ρ_{sum}, the semi-variogram is rougher but with a much larger scale, which is consistent with large-scale events.

– Anisotropy does not seem significant.

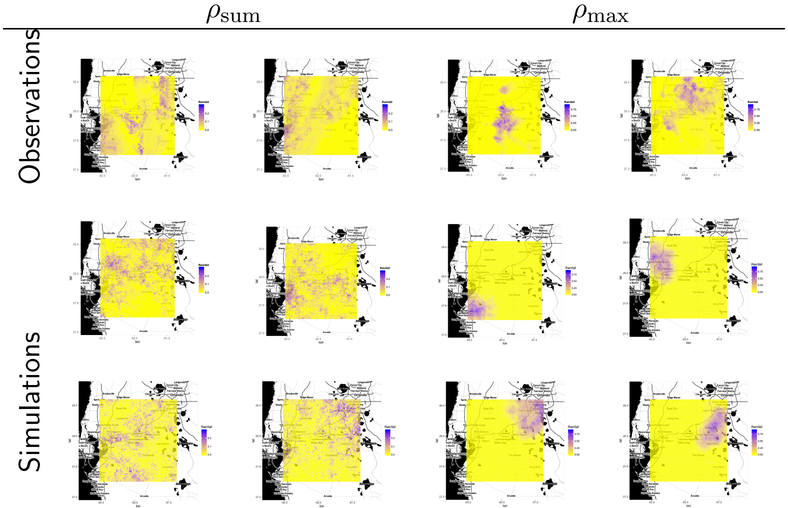
Example: Model checking

$\mathcal{R}_{\text{sum}}$ $\mathcal{R}_{\text{max}}$



Estimated pairwise extremogram (dots) as function of the distance for the risk functionals $\mathcal{R}_{\text{sum}}$ (left) and $\mathcal{R}_{\text{max}}$ (right). The solid black line represents the theoretical extremogram for a power variogram, whose parameters have been estimated using the gradient score.

Example: Simulations



15-minute cumulated rainfall (inches): observed (first row) and simulated (second and third rows) for the risk functionals ρ_{sum} (left) and $\rho_{\max}$ (right) with intensity equivalent to the 0.99 quantile.

Closing

- Basic ideas on maxima and exceedances extend to spatial (and space-time) settings.
- Max-stable processes give asymptotic dependence models—asymptotic independence also seen, and can be modelled using inverted max-stable processes.
- Can fit such models using
  - composite (usually pairwise) likelihood,
  - full likelihood (difficult with large dimension D),
  - Bayesian methods, or
  - gradient score methods.
- Latent variable models also available.
- Model-checking possible, using simulation from fitted models and other techniques—but difficult to validate far into tails, because of lack of data.
- Active research area (e.g., threshold models, non-stationarity, downscaling, semiparametric inference, networks, . . .).

Some reading

- Coles (2001), *Introduction to the Statistical Modeling of Extreme Values*, Springer
- Cooley, Naveau & Poncet (2006) In *Lecture Notes in Statistics*, vol 187
- Cooley, Naveau & Nychka (2007) *JASA*
- Davison & Huser (2015) *Annual Review of Statistics and its Applications*
- **Davison, Huser & Thibaud (2019)** In *Handbook of Environmental and Ecological Statistics*, CRC Press
- Davison, Padoan & Ribatet (2012) *Statistical Science*
- de Fondeville & Davison (2018) *Biometrika*
- de Haan (1984) *Annals of Probability*
- de Haan & Ferreira (2006) *Extreme Value Theory: An Introduction*, Springer
- Huser & Wadsworth (2018+) *JASA*
- Reich & Shaby (2012) *Annals of Applied Statistics*
- Schlather (2002) *Extremes*
- Thibaud & Opitz (2015) *Biometrika*
- Wadsworth & Tawn (2012) *Biometrika*
- Wadsworth & Tawn (2015) *Biometrika*