Statistical Models for Complex Extremes

Anthony Davison

Joint with
Peiman Asadi, Raphaël de Fondeville, Sebastian Engelke, Maria Süveges

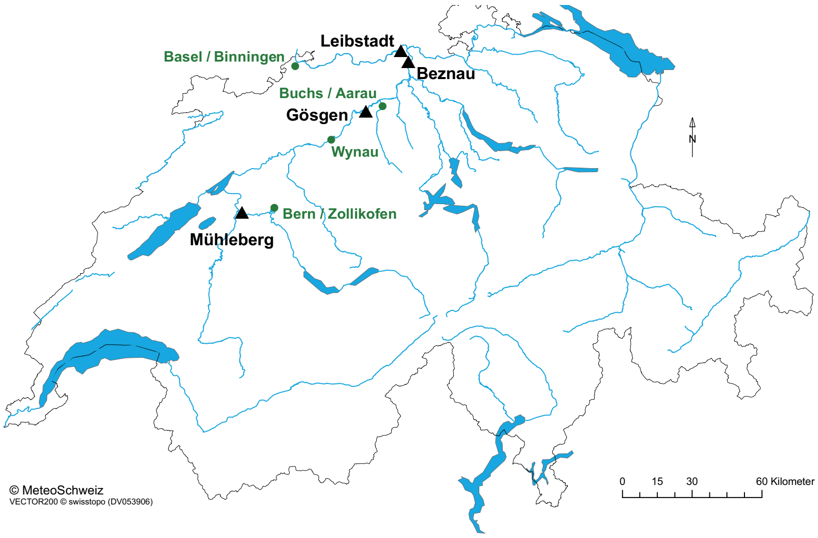
Funding: Swiss National Science Foundation, Swissnuclear,
Swiss Federal Office of the Environment

Fukushima, March 2011

Nuclear power safety

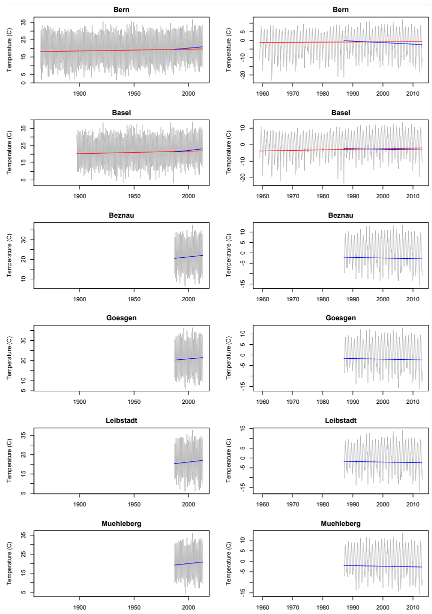
- Fukushima ⇒ nuclear power safety concerns worldwide
- Swiss nuclear regulator asked for (re-)assessment of vulnerability of the four nuclear plants to
  - high and low air temperatures
  - high and low river water temperatures
  - high winds (and tornados)
  - intense rainfall, snowload, lightning strikes,
  - earthquakes and any tsunamis are dealt with separately!
- Task: estimate quantiles for probabilities 10^{-4} per year (and 10^{-7} for high winds), and give their uncertainties
  - based on 25 years of data or so at the plants themselves, and (at very most, and only for comparison) 150 years of data nearby

Swiss nuclear plants



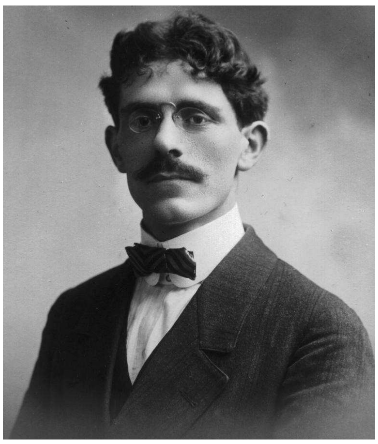
Muhleberg

Air temperature maxima and minima



Founders

Maurice René Fréchet (1878–1973)
Ronald Alymer Fisher (1890–1962)
Leonard Henry Caleb Tippett (1902–1985)



Modelling maxima

- A distribution G for maxima must satisfy the **max-stability** relation

$$G^m(b_m + a_m y) = G(y), \quad m = 1, 2, \ldots, \quad \{a_m\} > 0, \{b_m\} \subset \mathbb{R}.$$

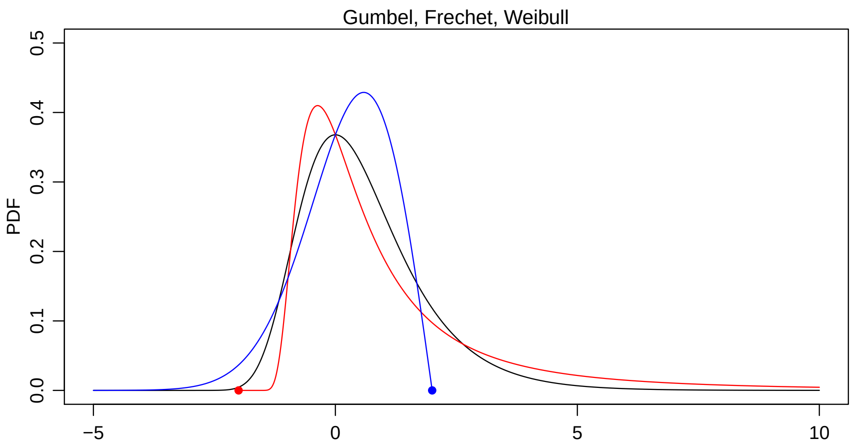
- Only non-trivial solution is the **generalized extreme-value (GEV) distribution**,

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

where $u_+ = \max(u, 0)$, and μ and τ are location and scale parameters.

- ξ is a shape parameter determining 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.
- ξ is hard to estimate, but crucial because it controls probabilities of large events.

GEV and shape parameter ξ



- PDFs of the **Gumbel** ($\xi = 0$), the **Fréchet** ($\xi = 0$) and the (reverse) **Weibull** ($\xi < 0$).
- The Fréchet is bounded below, and the reverse Weibull is bounded above.
- The standard Weibull is a distribution for minima.

Data analysis

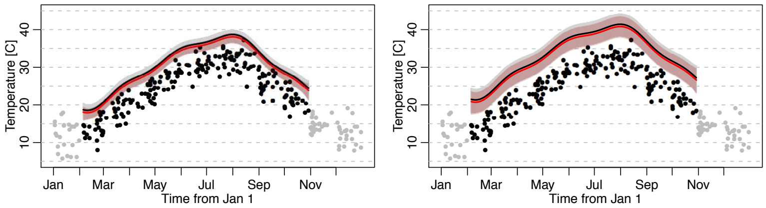


Figure 9: Seasonal 50- and 10000 year return levels (red and black, respectively) as a function of time in 1998 (left, coinciding with the reference model in Figure 7) and in 2050 (right). 95% confidence intervals from parametric bootstrap are shown as pink and light grey bands.

- ☐ Fitted GEV to monthly maxima for winter/summer seasons, allowing for monthly variation in location and (linear!) time trend
- ☐ Estimated shape parameter $\widehat{\xi} < 0$ implies upper bound on maximal temperature
- ☐ Attempt to allow for uncertainty due to
  - parameter estimation
  - stochastic variation of future events
  - number of observations contributing to maximum $(30 \neq \infty)$

General remarks

- ☐ Extreme value theory is based on **limiting** models for tails of distributions:
  - Generalised extreme-value distribution (GEV) applies for maxima of an infinite sample,
  - Generalized Pareto distribution (GPD) applies for peaks over an 'infinite' threshold,

  both satisfying notions of stability from mathematical considerations.
- ☐ Could of course fit many other models, but with weaker mathematical justification.
- ☐ In practice GEV/GPD fitted to finite samples, so extrapolation may be worrisome.
- ☐ Relevant data often limited, so need to combine information from elsewhere.
- ☐ Do we trust the mathematical models for real phenomena?

Black swans . . .

☐ To base extrapolation on max-stability, we assume that the unseen tail has no surprises (aka black swans):

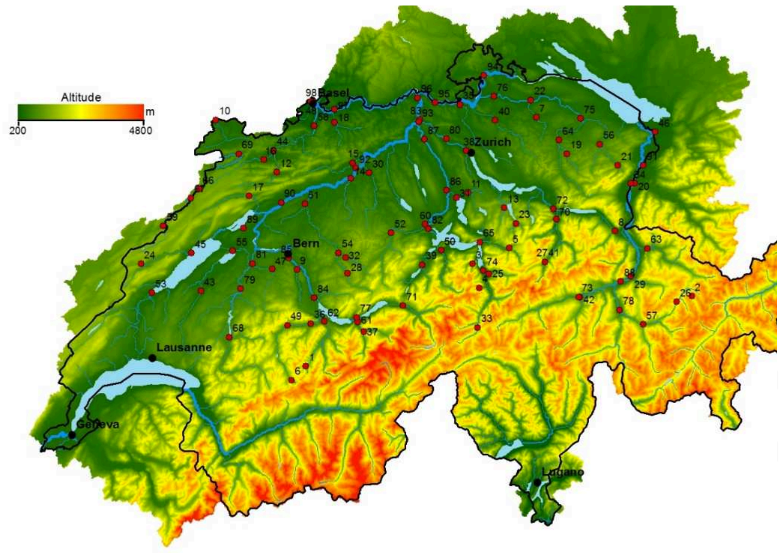
Black swans . . .

☐ To base extrapolation on max-stability, we assume that the unseen tail has no surprises (aka black swans):

A More Complex Problem

slide 15

River gauges in Northern Switzerland



Problem

- Aim to estimate flood risk on Aare–Rhine river network up to 2100, taking into account climate change
- How to assess combined flooding risk, based on short time series?
- Probabilities needed for events on network with annual probability 10^{-4}.
- Several university institutes involved (hydrology, flood hydraulics, geography, climate science, ...)

Max-stable processes

- The GEV distribution is **max-stable**: maxima of independent GEV variables are also GEV—in fact, this is the defining property of the GEV distribution, and allows extrapolation to rare events.
- For the unit Fréchet, GEV(1,1,1), distribution, this means that if $Z, Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \exp(-1/z)$, then for any n,

$$\max\{Z_1, \ldots, Z_n\} \overset{D}{=} nZ.$$

- For space/space-time problems we need a process analogue of the GEV, i.e., we seek a process $Z(x)$ such that if $Z_1(x), \ldots, Z_n(x) \overset{\text{iid}}{\sim} Z(x)$, then

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

  where $\mathcal{X}$ represents a space/space-time domain of interest (e.g., the Rhine watershed within Switzerland over the years 2020–2100).
- In the process case we first transform the process so that its marginal distributions are standard Fréchet at every x.

Construction of a max-stable process

- Let $W(x)$ be a non-negative random process with $\mathrm{E}\{W(x)\} = 1$ $(x \in \mathcal{X})$, and let

$$Z(x) = \sup_j R_j W_j(x), \quad x \in \mathcal{X}, \tag{1}$$

  with $\{R_j\}$ a Poisson process on $\mathbb{R}_+$ of rate $\mathrm{d}r/r^2$ and $\{W_j\}$ replicates of W.
- Then

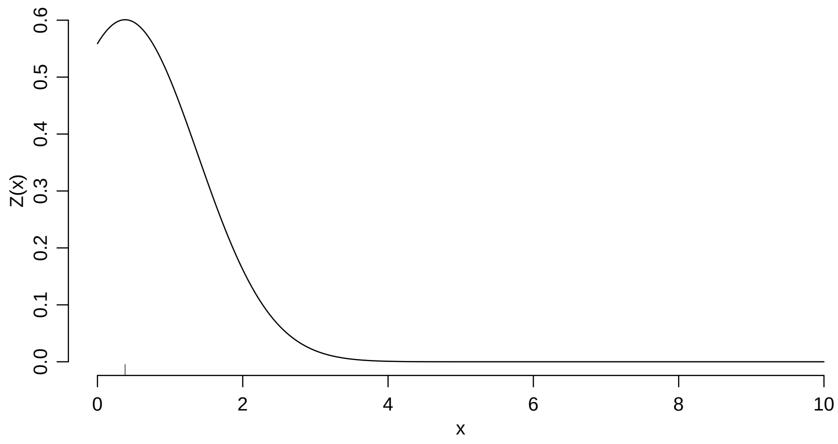
$$\mathrm{P}\{Z(x) \le z(x), x \in \mathcal{X}\} = \exp\left(-\mathrm{E}\left[\sup_{x \in \mathcal{X}}\left\{\frac{W(x)}{z(x)}\right\}\right]\right) = \exp\left[-V\{z(x)\}\right],$$

  say, and this gives:
  - a **max-stable process** $\{Z(x) : x \in \mathcal{X}\}$, i.e., there exist functions $\{b_n(x)\}$ and $\{a_n(x)\} > 0$ such that

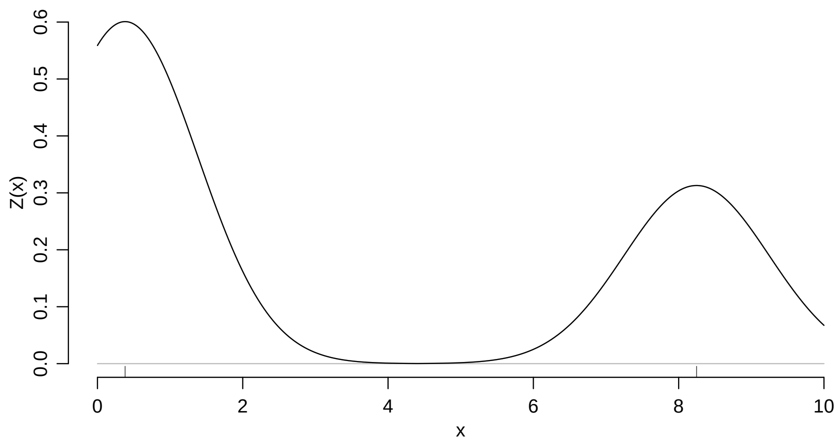
$$Z(x) \overset{D}{=} \max_{j=1}^{n}\left\{\frac{Z_j(x) - b_n(x)}{a_n(x)}\right\}, \quad x \in \mathcal{X}.$$

  - $Z(x) \sim$ unit Fréchet at each $x \in \mathcal{X}$.
- In fact any max-stable process can be written using the **spectral representation** (1).
- Example: Smith (1990) model ...

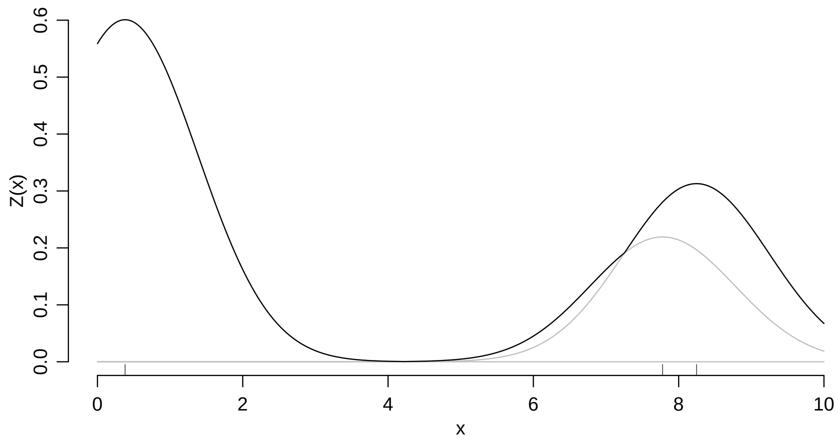
Making a Smith process



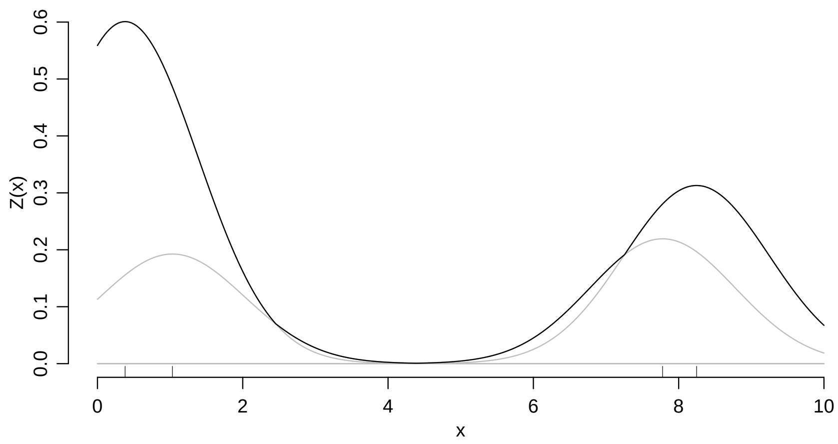
Making a Smith process



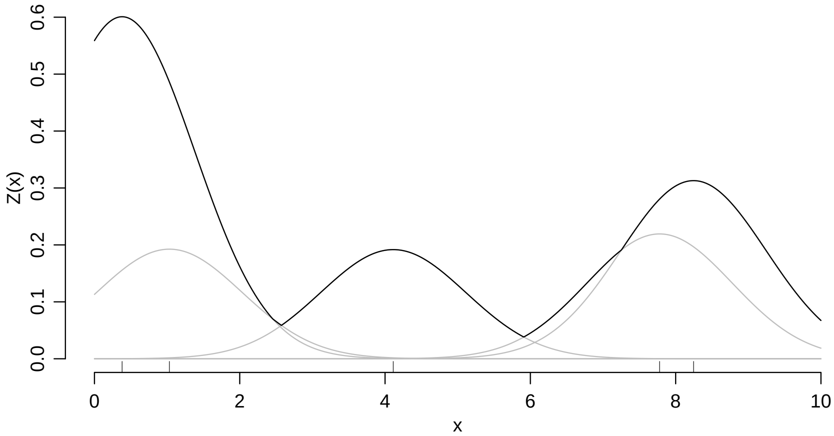
Making a Smith process



Making a Smith process

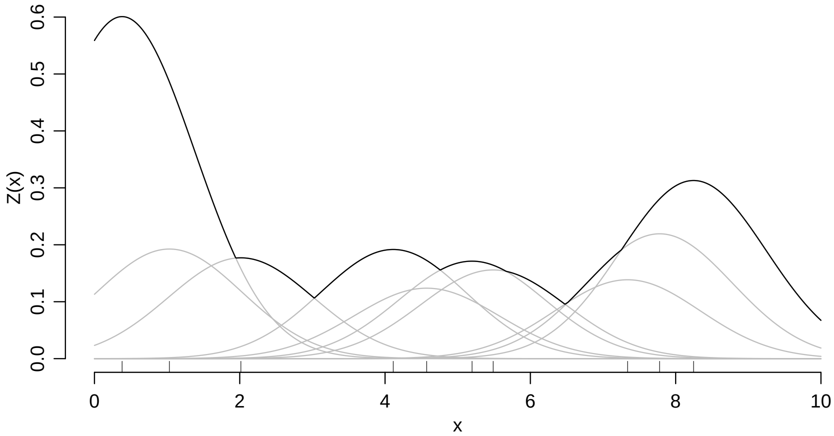


Making a Smith process



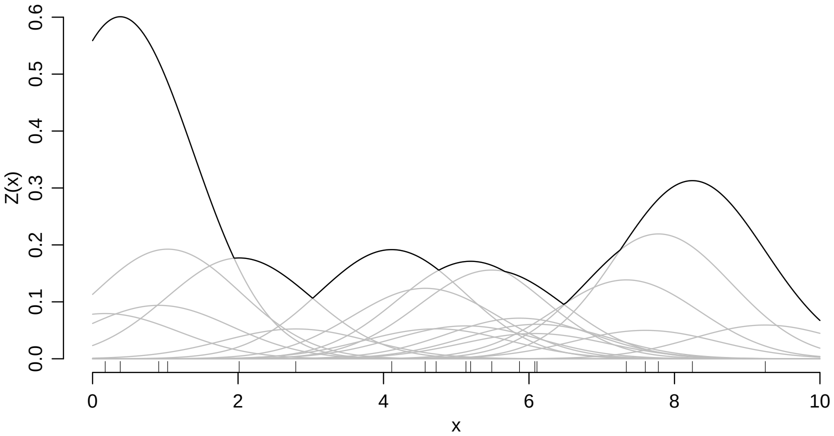
http://stat.epfl.ch CLADAG—Milano – slide 24

Making a Smith process

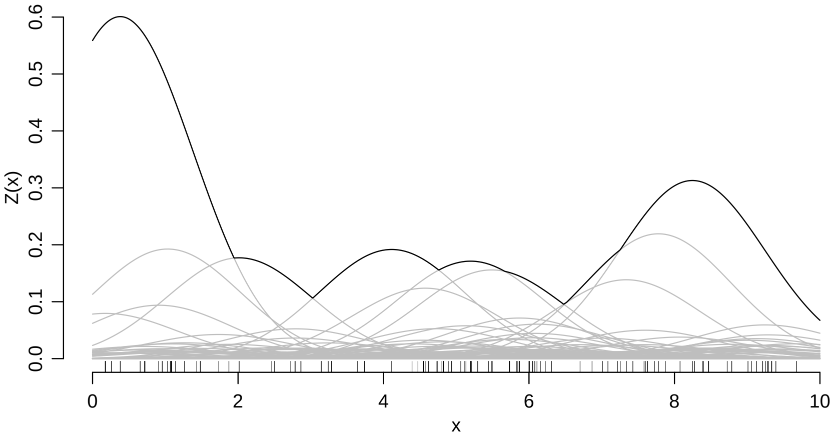


http://stat.epfl.ch CLADAG—Milano – slide 25

Making a Smith process



Making a Smith process

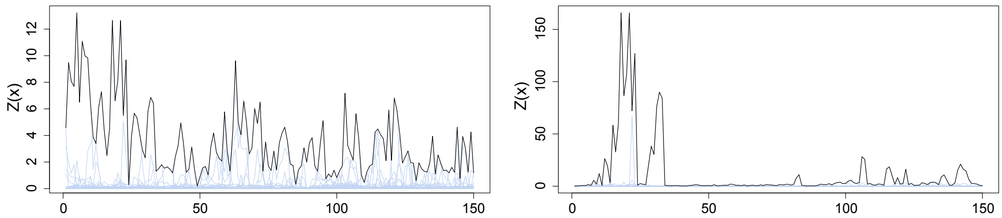


Comments

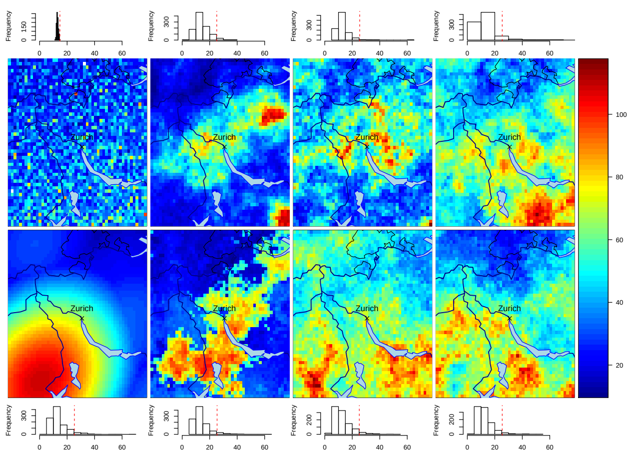
- ☐ Numerous max-stable models now exist, some more ‘realistic’ than others
- ☐ Particularly flexible example is the Brown–Resnick process, which takes

$$W(x) = \exp\{\varepsilon(x) - \gamma(x)\},$$

where $\varepsilon(x)$ is a stationary or intrinsically stationary Gaussian process with semi-variance or semivariogram $\gamma(x)$—can use panoply of functions γ from spatial statistics, or can invent your own.



Realisations from spatial models



Top: results from the latent variable, Student t copula, Hüsler–Reiss copula and extremal-t copula models. Bottom: results from the 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.

Extremal coefficient

- For any set $\mathcal{D} \subset \mathcal{X}$, homogeneity of V means that a max-stable model satisfies

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

and the **extremal coefficient**

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

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

- In particular, the pairwise version,

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

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

$$\text{(total dependence)} \quad 1 \leq \theta(x, x') \leq 2 \quad \text{(independence)},$$

and the interpretation

$$P\left\{Z(x') > z \mid Z(x) > z\right\} \sim 2 - \theta(x, x'), \quad z \to \infty.$$

- θ can be estimated nonparametrically, either as a basis for model checking, or for simple estimation of parameters.

Likelihood inference

- Suppose we have independent (annual) maxima observed at $\mathcal{D} = \{x_1, \ldots, x_D\} \subset \mathcal{X}$ for n years, so the data for each year have joint distribution

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

- The formulation of the model using its CDF means that to compute the likelihood function we must differentiate e^{-V} with respect to $z_1, \ldots, z_D$, leading to combinatorial explosion:

$$-V_1 e^{-V}, \quad (V_1 V_2 - V_{12})e^{-V}, \quad (-V_1 V_2 V_3 + V_{12} V_3[3] - V_{123})e^{-V}, \quad \ldots,$$

with about 10^5 terms for $D = 10$. Clearly this is infeasible for realistic applications, so we need to avoid this, by
  - using a composite (usually a pairwise) likelihood; or
  - using the timing of events to chose the term of the partition in the likelihood;
  - using threshold exceedances.
- In any case we must compute (many) derivatives of V, and sometimes integrate them ... can be painful.

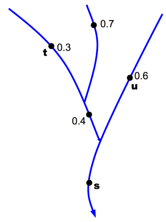
Extremal dependence on river network

- Sources of dependence between data at locations x_1 and x_2 on the network $\mathcal{X}$:
  - **flow-dependence**; x_2 is downstream of x_1, or vice versa
  - **'geo'-dependence**: the same events may impact nearby watersheds
- Overall semi-variogram

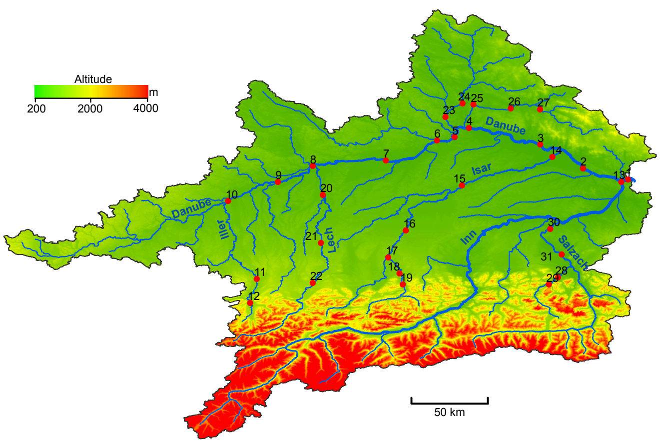
$$\gamma(x_1, x_2) = \lambda_{\mathrm{RIV}} \{1 - C_{\mathrm{RIV}}(x_1, x_2)\} + \lambda_{\mathrm{GEO}}\, \gamma_{\mathrm{GEO}}(x_1, x_2), \quad x_1, x_2 \in \mathcal{X},$$

  where $\lambda_{\mathrm{RIV}}, \lambda_{\mathrm{EUC}} > 0$.
- Flow-dependence in terms of shortest river distance $d(\cdot, \cdot)$:

$$\begin{aligned} C_{\mathrm{RIV}}(s, u) &= C_1\{d(s, u)\} \times \sqrt{0.6}, \\ C_{\mathrm{RIV}}(s, t) &= C_1\{d(s, t)\} \times \sqrt{0.4 \times 0.3}, \\ C_{\mathrm{RIV}}(u, t) &= 0, \\ C_1(h) &= \exp(-h/\theta), \quad \theta > 0. \end{aligned}$$



Upper Danube Basin

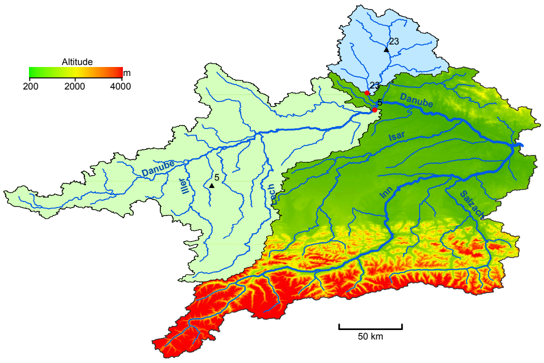


Asadi, Davison, Engelke (2016) *Annals of Applied Statistics*

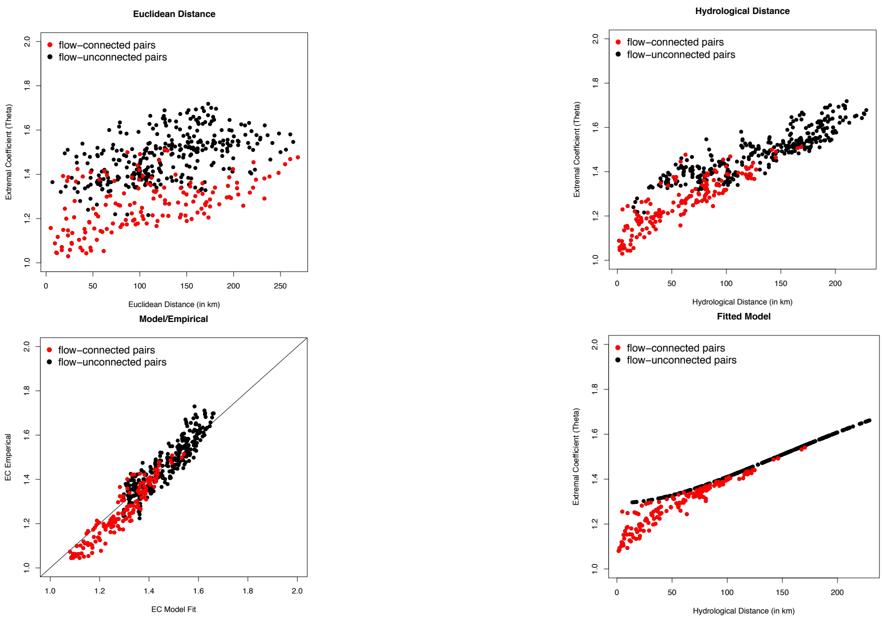
Extremal dependence on river network

☐ Introduce **hydrological location** of each station, as $h(x) \in \mathbb{R}^2$ as centroid of its sub-catchment, and define dependence measure

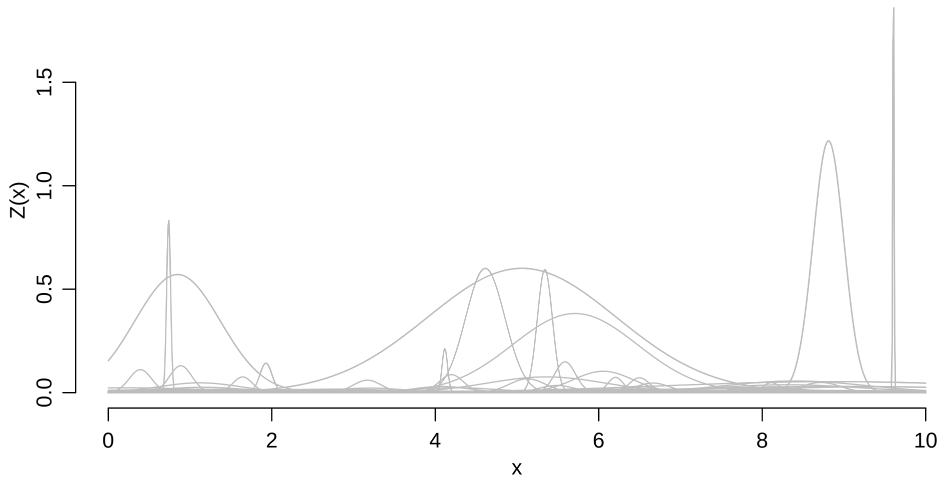
$$\gamma_{\text{EUC}}(x_1, x_2) = \|h(x_1) - h(x_2)\|^{\alpha}, \quad \alpha \in (0, 2].$$



Extremal coefficients



Individual events



Exceedances and risk functions

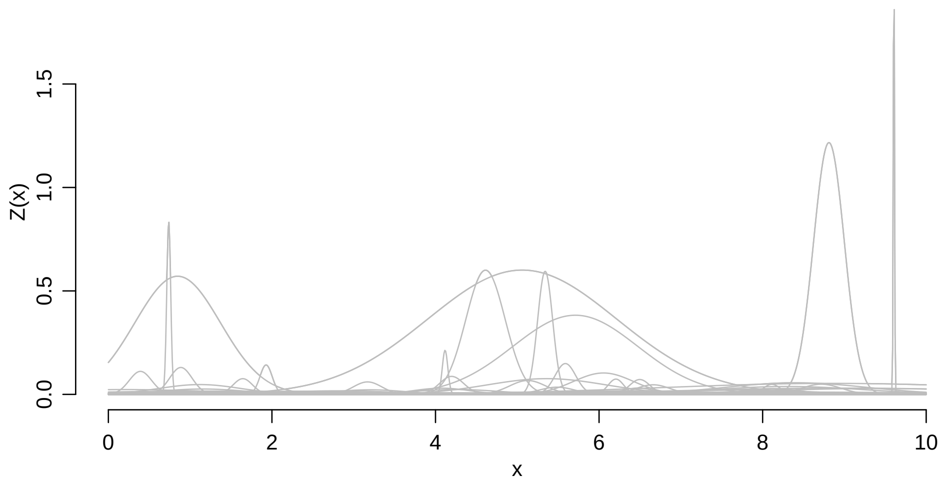
- ☐ Modelling threshold exceedances is widely used in (scalar) practice:
  - more flexible than using maxima
  - statistically more efficient, makes better use of data
- ☐ For scalar data, choosing rare events is easy: either they're big or they're small.
- ☐ For multivariate data, we need to say what 'direction' is extreme
- ☐ Do this via a scalar **risk function** f applied to the individual events $Q_j(x) = R_j W_j(x)$ of the max-stable process
  - Choose those events Q_j for which $f(Q_j)$ exceeds a threshold u
  - **Red**: extremes on $[0, 2]$, selected using risk function

$$f(Q) = \int_0^2 Q(x)\,dx$$

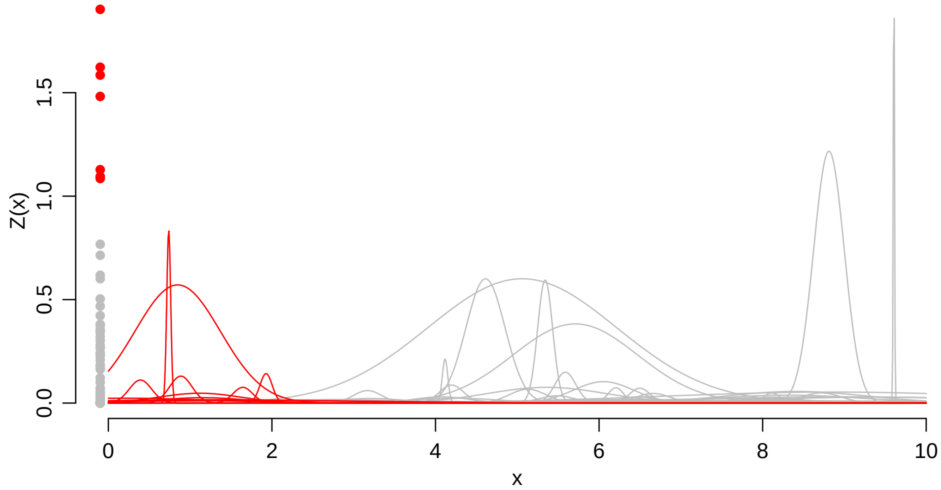
  - **Blue**: most intense events, selected using risk function

$$f(Q) = \max Q(x)$$

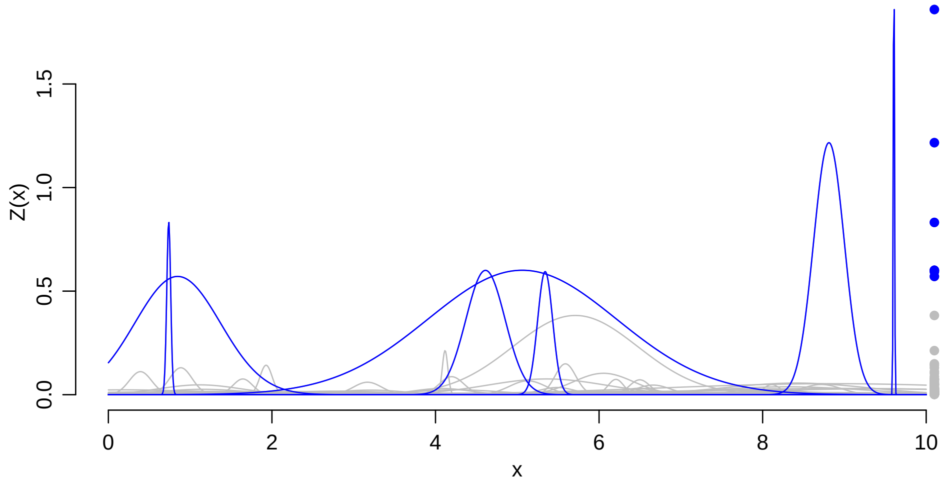
Individual events



Extremes in $[0, 2]$



Highest peaks anywhere

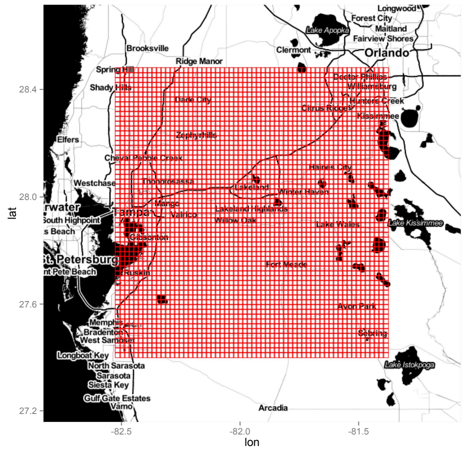


Inference

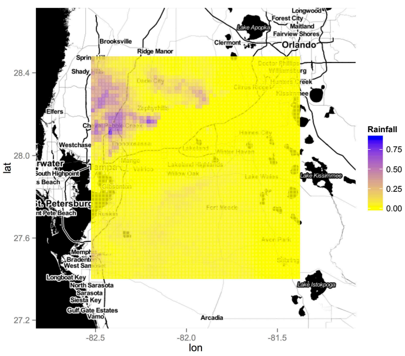
- Fitting for the 'exceedances' Q_j is (in principle) much easier than for the max-stable process $Z(x)$:
  - likelihoods can be constructed, at least for Gaussian-based processes $W(x)$ but
  - they still involve lots of burdensome integrals to compute norming constants.
- Fixes
  - estimate the integrals using quasi-Monte Carlo or other methods,
  - avoid likelihood inference, using the gradient score to dodge computing the norming constants.
- Big problems ($D \approx 1000$s) feasible with the gradient score, smaller ones ($D \approx 100$s) with quasi-Monte Carlo approximation.

Extreme rainfall over Florida

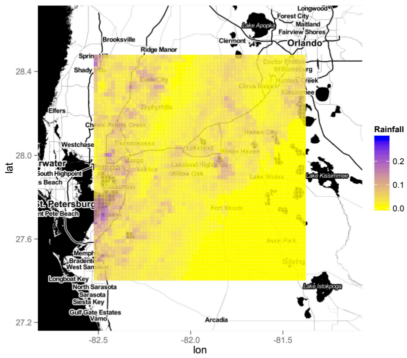
- 15-minute radar rainfall measurements over Florida from 1994–2010
- We focus on a 120 km× 120 km square south-west of Orlando and on the wet season, i.e., June to September.



Florida rainfall



Local rainfall



Spatially dispersed rainfall

Risk functionals

- We define two risk functionals

$$f_{\max}(X^*) = \left[\sum_{i=1}^{\ell} \{X^*(s_i)\}^{20}\right]^{1/20}, \quad f_{\text{sum}}(X^*) = \left[\sum_{i=1}^{\ell} \{X^*(s_i)\}^{\xi_0}\right]^{1/\xi_0},$$

where $\ell = 3600$ is the number of grid cells.

- Here
  - $f_{\max}$ is a continuous and differentiable approximation of $\max_{i=1,\ldots,\ell} X^*(s_i)$ which satisfies the requirements for the gradient score,
  - f_{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.

Spatial model and parameter estimates

- Non-separable semi-variogram model

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

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

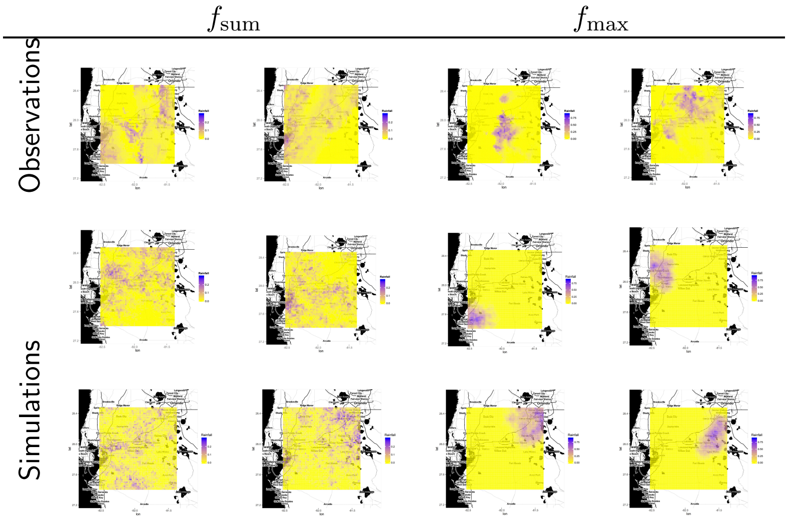
$$\Omega = \begin{bmatrix} \cos\eta & -\sin\eta \\ a\sin\eta & a\cos\eta \end{bmatrix}, \quad \eta \in \left(-\tfrac{\pi}{2}; \tfrac{\pi}{2}\right], \quad a > 1.$$

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

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

  - $f_{\max}$ estimates are quite smooth with a small scale, they capture high quantiles and induce a model similar to that in earlier work.
  - For f_{sum}, the semi-variogram is rougher but with a much larger scale, which is consistent with large-scale events.
  - Anisotropy does not seem significant.

Simulated extreme rainfall



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

Closing

Closing

- ☐ Basic ideas on maxima and point processes extend to spatial and space-time settings.
- ☐ Max-stable processes give asymptotic dependence models—asymptotic independence can be bothersome in practice, but models exist to account for it.
- ☐ Can fit such models using
  - pairwise likelihood (can be inefficient),
  - full likelihood (needs additional information, difficult with large D),
  - Bayesian methods, or
  - gradient score methods.
- ☐ Model-checking possible, using simulation from fitted models and other techniques—but difficult to validate far into tails, because of lack of data.
- ☐ Current 'hot' research area: lots going on (e.g., threshold models, non-stationarity, gridded data, non-Euclidean spaces, . . .).

Some reading

- ☐ Coles (2001), *Introduction to the Statistical Modeling of Extreme Values*, Springer
- ☐ de Haan and Ferreira (2006) *Extreme Value Theory: An Introduction*, Springer
- ☐ Davison and Huser (2015) *Annual Review of Statistics and its Applications*

- ☐ de Haan (1984) *Annals of Probability*
- ☐ Smith (1990) unpublished
- ☐ Davison, Padoan and Ribatet (2012) *Statistical Science*
- ☐ Wadsworth and Tawn (2012) *Biometrika*
- ☐ Thibaud, Mutzner and Davison (2013), *Water Resources Research*
- ☐ Huser and Davison (2014) *J. R. Statist. Soc., series B*
- ☐ Wadsworth and Tawn (2015) *Biometrika*
- ☐ Thibaud and Opitz (2015) *Biometrika*
- ☐ Asadi, Davison and Engelke (2016) *Annals of Applied Statistics*
- ☐ de Fondeville and Davison (2016) arXiv