An introduction to extreme-value theory

Thomas Opitz

BioSP, INRAE, Avignon

Short course at the Stochastic Geometry Days

Dijon, 12-13 June, 2023

Plan for this course

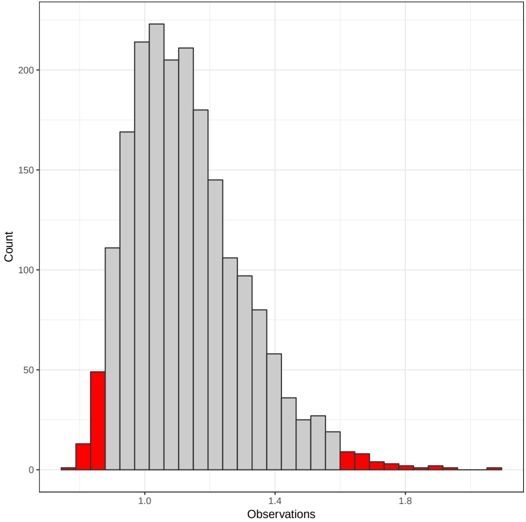
Three sessions of two hours, with (very roughly) the following main topics:

- Theory for univariate extremes
- Theory for dependent extremes based on maxima and point processes
- Theory for dependent extremes based on threshold exceedances

The origins of Extreme-Value Theory (EVT)

- A **probabilistic theory** with its origins in the first half of the 20th century:
  - Fréchet (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique*.
  - Fisher, Tippett (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*.
  - von Mises (1936). La distribution de la plus grande de n valeurs. *Revue Mathématique de l'Union Interbalcanique*
  - Gnedenko (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*.
- Strong development of **multivariate and process theory** since the 1970s
- **Statistical methods and applications**
  - Often at the origin of theoretical developments
    (for example, Tippett's work for the cotton industry)
  - Seminal monograph *Statistics of Extremes* (1958) of Gumbel
  - Numerous applications since the 1980s
  - Today, strong use for finance/insurance and climate/environment
  - Typical goals:
    - Estimate and extrapolate extreme-event probabilities
    - Stochastically generate new extreme-event scenarios

Extreme events are located in the upper or lower **tail of the distribution**:



Without loss of generality, we focus on the extremes in the upper tail.

Classical asymptotic frameworks: Averages / Extremes

Consider independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$

Averages $\overline{S}_n = \frac{1}{n}\sum_{i=1}^n X_i$

Central Limit Theorem

$$\frac{\overline{S}_n - \mu}{\sigma_n} \to Z \sim \mathcal{N}(0,1)$$

Gaussian limit distribution
(Sum-stability)

Spatial extension:

Gaussian processes

Geostatistics

Extremes (maxima) $M_n = \max_{i=1}^n X_i$

Fisher–Tippett–Gnedenko Theorem

$$\frac{M_n - a_n}{b_n} \to Z \sim \mathrm{GEV}(\xi) \text{ (tail index } \xi \in \mathbb{R})$$

Extreme-value limit distribution
(Max-stability)

Spatial extension:

Max-stable processes

Spatial Extreme-Value Theory

The trinity of the three fundamental approaches

Three asymptotic approaches to study extreme events in an i.i.d. sample $\{X_i\}$:

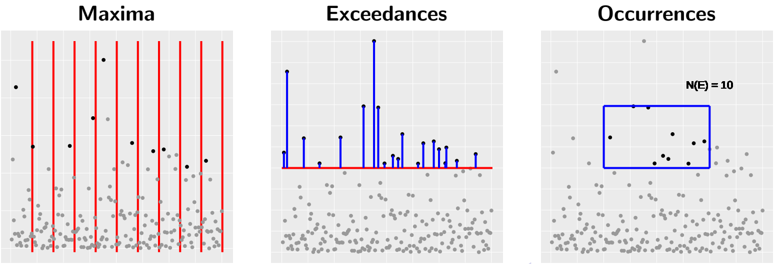
1. **Block maxima**: $M_n = \max_{i=1}^n X_i$ using blocks of size n
2. **Threshold exceedances** above a high threshold u: $(X_i - u) \mid X_i \geq u$
3. **Occurrence counts**: $N(E) = |\{X_i \in E,\ i = 1, \ldots, n\}|$ for extreme events E

Asymptotic theory

For

- increasing block size n,
- for increasing threshold u, and
- for more and more extreme event sets E,

we obtain **coherent theoretical representations across the three approaches**.



The maximum of a sample

For a series of **independent and identically distributed (iid) random variables**

$$X_i \sim F, \quad i = 1, 2, \ldots$$

we consider the **maximum**

$$M_n = \max_{i=1}^{n} X_i \sim F^n,$$

where

$$F^n(x) = (F(x))^n.$$

The fundamental extreme-value limit theorem

Fisher–Tippett–Gnedenko Theorem

Let X_i, $i = 1, 2, \ldots$ iid. If deterministic normalizing sequences a_n (location) and $b_n > 0$ (scale) exist such that

$$\frac{M_n - a_n}{b_n} \xrightarrow{d} Z \sim G, \quad n \to \infty, \quad (\star)$$

with a nondegenerate limit distribution G, then G is of one of the **three types of extreme-value distributions**:

- **(Reverse) Weibull**: $\tilde{G}(z) = \exp(-(-x)_+^{-\alpha})$ with $\alpha > 0$ (with support $(-\infty, 0)$)
- **Gumbel**: $\tilde{G}(z) = \exp(-\exp(-x))$ (with support $\mathbb{R}$)
- **Fréchet**: $\tilde{G}(z) = \exp(-x_+^{\alpha})$ with $\alpha > 0$ (with support $(0, \infty)$)

Remarks:

- Being of a certain type means being equal up to a location-scale transformation: $G(z) = \tilde{G}(a + bz)$ with some $b > 0$, $a \in \mathbb{R}$. We can always choose a_n, b_n such that $G = \tilde{G}$.
- If convergence $(\star)$ holds, we say that **F is in the maximum domain of attraction (MDA) of G**.
- Equivalently to $(\star)$, we have $F^n(a_n + b_n z) \to G(z)$, $n \to \infty$, $z \in \mathbb{R}$.

Sketch of the proof (1)

A key ingredient is the **Extremal-Types Theorem**, here copied from Embrechts, Klueppelberg, Mikosch (1996). An early proof is due to Gnedenko & Kolmogorov (1954).

Extremal-Types Theorem

Let $A, B, A_1, A_2, \ldots$ be random variables and $b_n > 0$, $\beta_n > 0$ and $a_n, \alpha_n \in \mathbb{R}$ be deterministic sequences. If the following convergence holds,

$$\frac{A_n - a_n}{b_n} \xrightarrow{d} A, \quad n \to \infty,$$

then the alternative convergence

$$\frac{A_n - \alpha_n}{\beta_n} \xrightarrow{d} B, \quad n \to \infty, \tag{1}$$

holds if and only if

$$\frac{b_n}{\beta_n} \to b \in [0, \infty), \quad \frac{a_n - \alpha_n}{\beta_n} \to a \in \mathbb{R}, \quad n \to \infty.$$

If (1) holds, then $B \stackrel{d}{=} bA + a$ with a, b being uniquely determined. Moreover, A is nondegenerate if and only if $b > 0$, and the A and B are said to belong to the same type.

Sketch of the proof (2)

In the following, all convergences are understood for $n \to \infty$.

1. If the convergence $F^n(a_n + b_n z) \to G(z)$ holds, then for any $t > 0$,
$$F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor} + b_{\lfloor nt \rfloor} z) \to G(z), \quad z \in \mathbb{R}. \tag{2}$$
2. Observe that
$$F^{\lfloor nt \rfloor}(a_n + b_n z) = (F^n(a_n + b_n z))^{\lfloor nt \rfloor / n} \to G^t(z). \tag{3}$$
3. Using the Extremal-Types Theorem, there exist deterministic functions $\gamma(t) > 0$ and $\delta(t)$ such that
$$\frac{b_n}{b_{\lfloor nt \rfloor}} \to \gamma(t), \quad \frac{a_n - a_{\lfloor nt \rfloor}}{b_{\lfloor nt \rfloor}} \to \delta(t), \quad t > 0.$$
By considering (2) and (3), we get
$$G^t(z) = G(\delta(t) + \gamma(t) z), \quad t > 0.$$
4. A consequence of the last equality is that for $s, t > 0$,
$$\gamma(st) = \gamma(s)\gamma(t), \quad \delta(st) = \gamma(t)\delta(s) + \delta(t).$$
5. The solutions of this functional equation are given by the three distribution functions of the reverse Weibull, Gumbel and Fréchet type.

Generalized Extreme-Value distribution (GEV)

The **Generalized Extreme-Value distributions (GEV)** uses threes parameter to jointly represent all possible limit distributions G:

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

- **Shape parameter** (or **tail index**) $\xi \in \mathbb{R}$, determining the extremal type:
  - Reverse-Weibull MDA for $\xi < 0$
  - Gumbel MDA for $\xi = 0$
  - Fréchet MDA for $\xi > 0$
- Location parameter $\mu \in \mathbb{R}$
- Scale parameter $\sigma > 0$

For $\xi = 0$, $(\star\star)$ is the limit for $\xi \to 0$: $G(z) = \exp(-\exp(-(z-\mu)/\sigma))$, $z \in \mathbb{R}$.

The $(...)_+$-operator in $(\star\star)$ means that the distribution G has positive density dG/dz for values z satisfying $1 + \xi \frac{z-\mu}{\sigma} > 0$

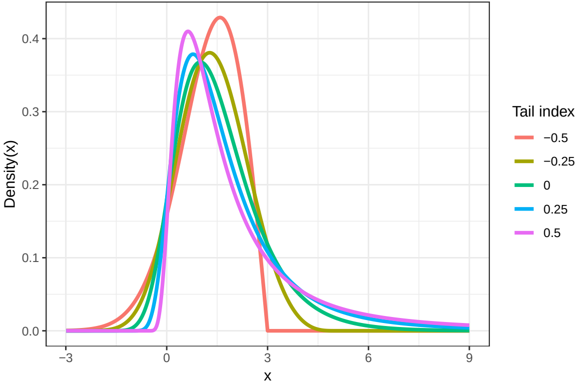
$\Rightarrow$ **Support of the GEV:** $A_{\xi,\sigma,\mu} = \begin{cases} (-\infty, \mu - \sigma/\xi), & \xi < 0, \\ (-\infty, \infty), & \xi = 0, \\ (\mu - \sigma/\xi, \infty), & \xi > 0. \end{cases}$

Illustration: GEV densities

In the MDA convergence ($\star$), we can always choose the normalizing sequences a_n, b_n such that $\mu = 0$, $\sigma = 1$, as for the probability densities shown below.

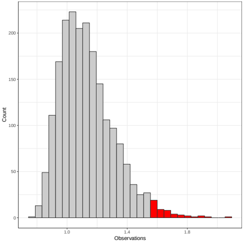
The three types have very different upper tail structure:

- Reverse-Weibull for $\xi < 0$: light tails with finite upper endpoint (GEV finite upper endpoint is $\mu - \sigma/\xi$)
- Gumbel for $\xi = 0$: exponential tail
- Fréchet for $\xi > 0$: power-law tails, i.e., heavy tails

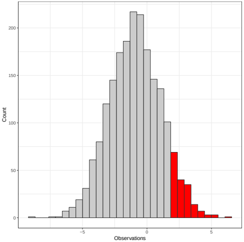


Histograms of i.i.d. samples X_i, $i = 1, 2, \ldots, n$, with different tail index ξ.

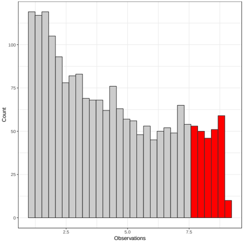
Heavy tail
$\xi > 0$



Exponential tail
$\xi = 0$



Bounded tail
$\xi < 0$



Examples of MDAs of common distributions:

- $\xi > 0$: Pareto ($\xi = 1/\text{shape}$), student's t ($\xi = \text{shape}$)
- $\xi = 0$: Normal, Exponential, Gamma, Lognormal
- $\xi < 0$: Uniform ($\xi = -1$), Beta

Example: GEV limit of the exponential distribution

Consider the standard exponential distribution with cdf $F(x) = 1 - \exp(-x)$, $x > 0$. The distribution F^n of the maximum $M_n = \max_{i=1}^n X_i$, where $X_i \overset{iid}{\sim} F$, $i = 1, \ldots, n$, is

$$F^n(x) = (1 - exp(-x))^n .$$

Can we find a_n and b_n such that $\lim_{n\to\infty} F^n(a_n + b_n x)$ exists and is nondegenerate?

For $x > -\log n$,

$$F^n(\log n + x) = (1 - \exp(-(\log n + x))^n = \left(1 - \frac{\exp(-x)}{n}\right)^n$$
$$\to \exp(-\exp(-x)), \quad n \to \infty$$

Conclusion:

- Using $a_n = \log(n)$ and $b_n = 1$, we obtain $\lim_{n\to\infty} F^n(a_n + b_n x) = \exp(-\exp(-x))$ for any $x \in \mathbb{R}$.
- The exponential distribution is in the **maximum domain of attraction of the standard Gumbel distribution**, i.e., the GEV with $\xi = 0$, $\mu = 0$, $\sigma = 1$.

A key theoretical characterisation of extreme-value limit distributions is as follows:

Class of extreme-value limit distributions G = Class of max-stable distributions

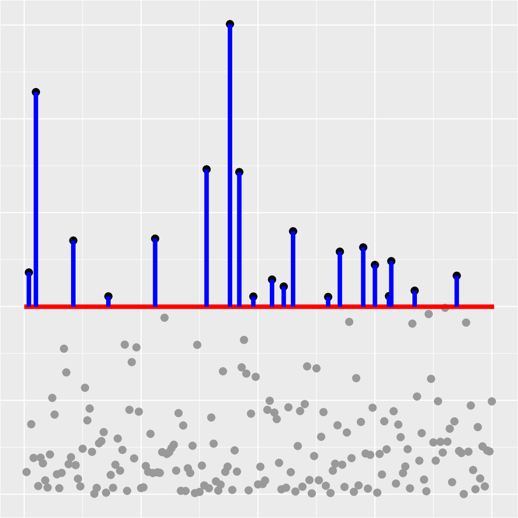
Max-stable distribution

A probability distribution G is called max-stable if for any $n \in \mathbb{N}$ there exist appropriate choices of deterministic normalizing sequences α_n and $\beta_n > 0$ such that

$$G^n(\alpha_n + \beta_n z) = G(z), \quad \text{for any} n \in \mathbb{N}.$$

This also means that the MDA limit $(\star)$ is exact (and not asymptotic) if F is max-stable.

Threshold exceedances in a univariate sample



What are possible limits for threshold excesses

$$X - u \quad \text{given} \quad X > u \quad ?$$

Generalized Pareto limits for threshold exceedances

Consider iid $X, X_1, X_2, \ldots$ where $X \sim F$
with upper endpoint $x^\star = \sup\{x \in \mathbb{R} : F(x) < 1\} \in (-\infty, \infty]$.

Pickands–Balkema–de-Haan Theorem

Suppose that $M_n = \max(X_1, \ldots, X_n)$ converges to a $\mathrm{GEV}(\xi, \mu, \sigma)$ distribution according to the Fisher–Tippett–Gnedenko theorem.
Equivalently, there exists a scaling function $\sigma(u) > 0$ such that

$$(X - u)/\sigma(u) \mid (X > u) \quad \to \quad Y, \quad u \to x^\star,$$

and Y follows the **Generalized Pareto Distribution** $\mathrm{GPD}(\xi, \sigma_{GPD})$ given as

$$\mathrm{GPD}(y; \xi, \sigma_{GPD}) = \mathrm{Pr}(Y \leq y) = 1 - (1 + \xi y/\sigma_{GPD})_+^{-1/\xi} \quad y > 0,$$

with scale parameter $\sigma_{GPD} > 0$.

- This result dates back to the 1970s.
- As before, the case $\xi = 0$ is interpreted as the limit for $\xi \to 0$:

$$\mathrm{GPD}(y; 0, \sigma_{GPD}) = 1 - \exp(-y/\sigma_{GPD}), \quad y > 0$$

(= Exponential distribution).

Sketch of the proof

We here sketch the proof of "$\Rightarrow$"
(Convergence of maxima leads to convergence of threshold excesses).

1. Set $u_n = a_n + b_n \tilde{u}$ for $\tilde{u}$ chosen in the support of the $\mathrm{GEV}(\xi, \mu, \sigma)$. Then,

$$\Pr((X - u_n)/b_n > y \mid X > u_n) = \frac{1 - F(a_n + b_n(y + \tilde{u}))}{1 - F(a_n + b_n \tilde{u})}. \tag{4}$$

2. On the one hand, the MDA condition $F^n(a_n + b_n z) \to G(z)$ implies

$$\log F(a_n + b_n z) \approx \frac{1}{n} \log G(z), \quad \text{for large } n.$$

   On the other hand, since $F(a_n + b_n z) \approx 1$ as n increases, we can use the first-order approximation $\log(1 + x) \approx x$ for small $|x|$, such that

$$\log F(a_n + b_n z) \approx F(a_n + b_n z) - 1.$$

   Combining the two yields

$$1 - F(a_n + b_n z) \approx -\frac{1}{n} \log G(z). \tag{5}$$

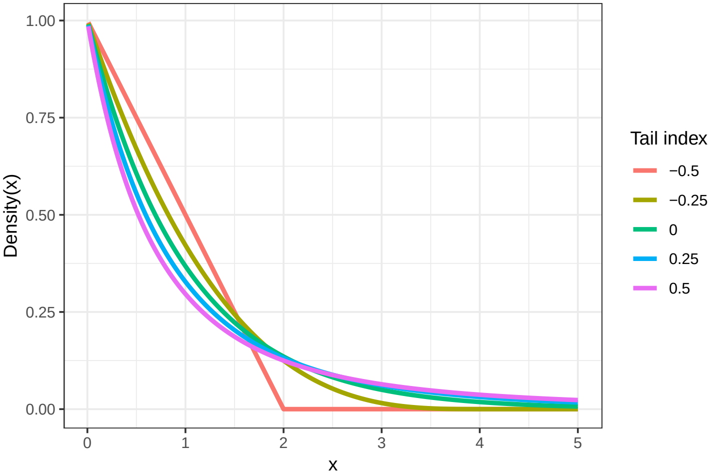
3. By using the approximation (5) for the numerator and denominator of (4), we get

$$\Pr((X - u_n)/b_n > y \mid X > u_n) \to \frac{\log G(\tilde{u} + y)}{\log G(\tilde{u})} = 1 - \mathrm{GPD}(y; \xi, \sigma_{GPD}), \quad n \to \infty;$$

   with $\sigma_{GPD} = \sigma + \xi(\tilde{u} - \mu) > 0$, and we can set $\sigma(u_n) = b_n$.

Illustration: GPD densities

The value of the tail index ξ characterizes the shape of the distribution. Here, σ_{GPD} is fixed to 1.



Peaks-over-threshold stability

By analogy with **max-stability** of GEV limit distributions for maxima, we have **Peaks-Over-Threshold (POT) stability** for limit distributions of threshold exceedances.

Peaks-Over-Threshold stability of the GPD

Suppose that $Y \sim \mathrm{GPD}(\xi, \sigma_{GPD})$. Consider a new, higher threshold $\tilde{u} > 0$ such that $\mathrm{GPD}(\tilde{u}; \xi, \sigma_{GPD}) < 1$. Then

$$Y - \tilde{u} \mid (Y > \tilde{u}) \sim \mathrm{GPD}(\xi, \tilde{\sigma}_{GPD}), \quad \tilde{\sigma}_{GPD} = \sigma_{GPD} + \xi \tilde{u}.$$

Exercice: Prove this using pencil + paper by showing

$$\frac{1 - \mathrm{GPD}(\tilde{u} + y; \xi, \sigma_{GPD})}{1 - \mathrm{GPD}(\tilde{u}; \xi, \sigma_{GPD})} = 1 - \mathrm{GPD}(y; \xi, \tilde{\sigma}_{GPD})$$

$\Rightarrow$ Application of the POT approach to a GPD yields again a GPD!

For $\xi = 0$, where the GPD is the exponential distribution, the POT stability is also known as the **lack-of-memory property**.

The **trinity** of univariate extreme-value limits is completed by point patterns.

Theorem (Point-process convergence)

For i.i.d. copies $X_1, X_2, \ldots$ of $X \sim F$, the following two statements are equivalent:

1. The distribution F is in the maximum domain of attraction of the max-stable distribution G with support $A_{\xi,\sigma,\mu}$ for the normalizing sequences $a_n \in \mathbb{R}$ and $b_n > 0$.
2. For the normalizing sequences $a_n \in \mathbb{R}$ and $b_n > 0$, we have the following point-process convergence with a locally finite Poisson-process process limit:

$$\left\{ \left(\frac{i}{n}, \frac{X_i - a_n}{b_n} \right), \; i = 1, \ldots, n \right\} \to \{(t_i, P_i), \; i \in \mathbb{N}\} \sim \mathrm{PPP}(\lambda_1 \times \Lambda), \quad n \to \infty,$$

   with intensity measure $\lambda_1 \times \Lambda$ where λ_1 is the Lebesgue measure on $(0,1)$.

If 1) and 2) hold, then $G(z) = \exp(-\Lambda[z;\infty))$, and the **exponent measure Λ** defined on $A_{\xi,\sigma,\mu}$ is characterized by its tail measure

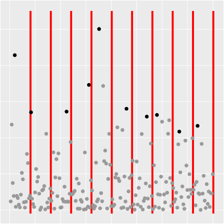
$$\Lambda[z,\infty) = -\log G(z) = \begin{cases} \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ \exp\left(\frac{z-\mu}{\sigma}\right), & \xi = 0 \end{cases}, \qquad \mu \in \mathbb{R}, \; \sigma > 0.$$

Remark: Λ is singular at $\inf A_{\xi,\sigma,\mu}$.

Summary: The extreme-value trinity

We allow for affine-linear rescaling $\tilde{X}_i = \frac{X_i - b_n}{a_n}$ of the iid sample X_i, $i = 1, \dots, n$.

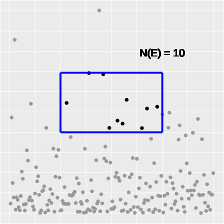
Maxima



$\Pr(\max_{i=1}^{n} \tilde{X}_i \le z)$
$\to \exp\left(-\Lambda[z, \infty)\right)$

Max-stable distr. (GEV)

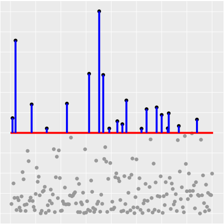
Occurrence counts



$\Pr(N(E) = k) \to$
$\exp(-(\lambda_1 \times \Lambda)(E)) \frac{(\lambda_1 \times \Lambda)(E)^k}{k!}$

Poisson process

Threshold exceedances



$\Pr(\tilde{X}_i - u > y \mid \tilde{X}_i > u)$
$\to \Lambda[y, \infty)/\Lambda[u, \infty)$

Gen. Pareto distr. (GPD)

Exponent measure Λ **possessing asymptotic stability:**
for any event E and $c > 0$, there are constants $\alpha(c) \in \mathbb{R}$, $\beta(c) > 0$ such that

$$c \times \Lambda(E) = \Lambda\left(\frac{E - \alpha(c)}{\beta(c)}\right)$$

Practical motivation for dependent extremes

Often, several variables are stochastically dependent,
for example in environmental and climatic data.

Examples:

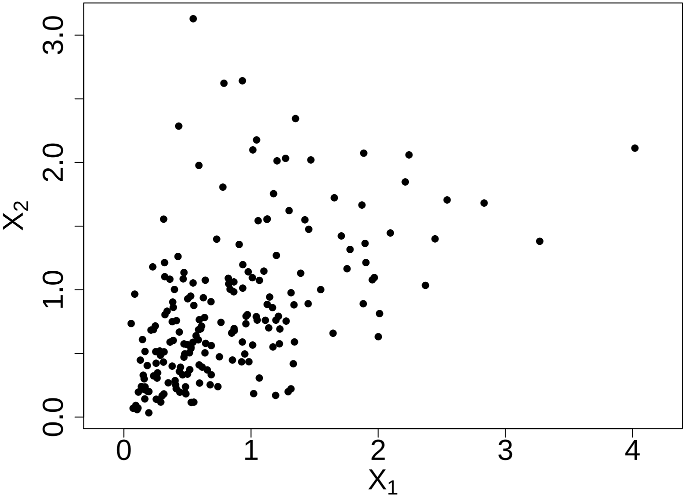
- Different physical variables observed at the same location, such as minimum temperature, maximum temperature, precipitation, wind speed.
- The same physical variable observed at different locations, such as precipitation at different locations of a river catchment.

Many interesting aspects of dependent extremes:

- **Aggregation** of extreme observations in several components (example: cumulated precipitation $\Rightarrow$ flood risk)
- **Spatial extent** and **temporal duration** of environmental extreme events
- **Reliability:** simultaneous failure of several critical components

Illustration: a bivariate sample with dependence

Scatterplot of an iid bivariate sample $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$, $i = 1, 2, \ldots, n$.



A note on notations (multivariate / process)

Representations for extremes of random vectors and stochastic processes are structurally quite similar.

For indexing the variables of interest,

- we can either put focus on the multivariate aspect and use indices $1, \dots, d$ for the d components of a random vector

$$(X_1, \dots, X_d)$$

  (and we can write $D = \{1, \dots, d\}$ for the domain),
- or we put focus on the process aspect (for example, when working with a random field on a nonempty domain $D \subset \mathbb{R}^k$) and use notation such as

$$\{X(s),\ s \in D\}$$

  for the whole process, or

$$(X(s_1), \dots, X(s_d))$$

  for the multivariate vector of variables observed at d locations $s_1, \dots, s_d \subset \mathbb{R}^k$.

When the distinction is important, we point it out explicitly (for example, for "functional convergence" in a space of functions with continuous sample paths defined over a compact domain D).

Componentwise maxima of random vectors

Consider a sequence of iid random vectors

$$\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d}) \stackrel{d}{=} \boldsymbol{X} \sim F_{\boldsymbol{X}},$$

where $F_{\boldsymbol{X}}$ is the joint distribution of the components of $\boldsymbol{X}$:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(x_1, \ldots, x_d) = \Pr(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

The **componentwise maximum**

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d}) = \left(\max_{i=1}^{n} X_{i,1}, \ldots, \max_{i=1}^{n} X_{i,d} \right)$$

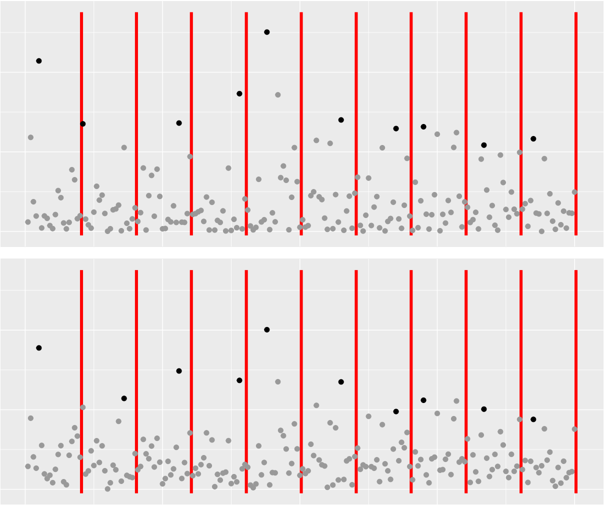
has distribution $F_{\boldsymbol{X}}^n$, that is, for $\boldsymbol{x} = (x_1, \ldots, x_d)$,

$$F_{\boldsymbol{X}}^n(\boldsymbol{x}) = (F_{\boldsymbol{X}}(\boldsymbol{x}))^n = \Pr(X_{i,1} \leq x_1, \ldots, X_{i,d} \leq x_d, \; i = 1, \ldots, n)$$

⚠ The componentwise maximum $\boldsymbol{M}_n$ can be composed of values $X_{i,j}$ with different indices i.

Illustration: bivariate componentwise block maxima

A bivariate series $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$ (with strong cross-correlation) and its componentwise maxima within the blocks separated by red lines. Most but not all of the maxima occur at the same time in the two series.



Definition: max-stable distribution; max-stable process

A **multivariate (d-dimensional) distribution** G is called **max-stable** if there exist deterministic vector sequences $\boldsymbol{\alpha}_n = (\alpha_{n,1}, \ldots, \alpha_{n,d})$ and $\boldsymbol{\beta}_n = (\beta_{n,1}, \ldots, \beta_{n,d}) > \mathbf{0}$, $n \in \mathbb{N}$, such that

$$G^n(\boldsymbol{\alpha}_n + \boldsymbol{\beta}_n \boldsymbol{z}) = G(\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^d.$$

If all finite-dimensional distributions of a stochastic process $\boldsymbol{Z} = \{Z(s), s \in D \subset \mathbb{R}^k\}$ are max-stable, we call $\boldsymbol{Z}$ a **max-stable process**.

Equivalently, if $\boldsymbol{X}_1 \sim G$, then the componentwise maximum over n iid copies of $\boldsymbol{X}_1$ satisfies

$$\frac{\boldsymbol{M}_n - \boldsymbol{\alpha}_n}{\boldsymbol{\beta}_n} \overset{d}{=} \boldsymbol{X}_1, \quad n \in \mathbb{N}.$$

⚠ Multivariate max-stability is stronger than max-stability of the univariate marginal distributions.

- If $\boldsymbol{Z} = (Z_1, \ldots, Z_d) \sim G$ with $Z_j \sim G_j$, then the univariate marginal distributions G_j are max-stable:

$$Gj(z_j) = \mathrm{GEV}(z_j; \xi_j, \mu_j, \sigma_j) = \Pr(Z_j \leq z_j) = G(\infty, \ldots, \infty, z_j, \infty, \ldots, \infty).$$

- Additionally, max-stability of G implies a stability property for the dependence structure.

Multivariate Maximum-Domain-of-Attraction theorem

Theorem: Multivariate Maximum Domain of Attraction

If there exist deterministic normalizing vector sequences $\boldsymbol{a}_n = (a_{n,1}, \ldots, a_{n,d})$ and $\boldsymbol{b}_n = (b_{n,1}, \ldots, b_{n,d}) > 0$, $n \in \mathbb{N}$, such that the following convergence holds,

$$\frac{\boldsymbol{M}_n - \boldsymbol{a}_n}{\boldsymbol{b}_n} \to \boldsymbol{Z} = (Z_1, \ldots, Z_d) \sim G, \quad n \to \infty,$$

where $\boldsymbol{Z}$ has non-degenerate marginal distributions, then G is a **multivariate extreme-value distribution**, that is, a **multivariate max-stable distribution**.

If all finite-dimensional distributions of a process $\boldsymbol{X} = \{X(s),\, s \in D \subset \mathbb{R}^k\}$ satisfy the above convergence, then $\boldsymbol{Z} = \{Z(s), s \in D \subset \mathbb{R}^k\}$ is a **max-stable process**.

(see, for instance, Resnick (1987) for the proof)

Remark: For stochastic processes, we here define convergence in terms of finite-dimensional distributions. There also exist results for convergence in spaces of continuous functions over a compact domain D.

Formulation using standardized marginal distributions

To focus on the extremal dependence structure, it is useful to **standardize the marginal distributions** F_j of X_j and G_j of Z_j.

- Often, the **unit Fréchet marginal distribution** is used:

$$G_j^\star(z) = \mathrm{GEV}(z;\, \xi = 1, \mu = 1, \sigma = 1) = \exp\left(-\frac{1}{z}\right), \quad z > 0.$$

- We can transform any continuous random variable $X \sim F$ towards a variable with unit Fréchet distribution as follows: $X^\star = -\frac{1}{\log F(X)} \sim G^\star$.

- If $X_j \sim \mathrm{GEV}(\xi, \mu, \sigma)$, then $X_j^\star = \left(1 + \xi \frac{X-\mu}{\sigma}\right)^{1/\xi} \sim G_j^\star$.

- If G is a multivariate max-stable distribution, we write $G^\star$ for the corresponding max-stable distribution with unit Fréchet margins. We call $G^\star$ a **simple max-stable distribution**.

We call representations simple if they are based on the marginal $\star$-scale.

Simple Maximum Domain of Attraction

We use the following notation: $T_{\xi,\mu,\sigma}(z) = \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{1/\xi}$.

Maximum Domain of Attraction using standardized marginal distributions

Consider a random vector $\boldsymbol{X} \sim F_{\boldsymbol{X}}$. The following two statements are equivalent:

1. The distribution $F_{\boldsymbol{X}}$ is in the MDA of a multivariate max-stable distribution G.
2. The following two properties hold jointly:
   1. **Marginal convergence:** Each component X_j is in the univariate MDA of a $\mathrm{GEV}(\xi_j, \mu_j, \sigma_j)$ distribution.
   2. **Convergence on the standardized scale:** The distribution of the marginally standardized random vector

      $$\boldsymbol{X}^\star = (X_1^\star, \dots, X_d^\star) \sim F_{\boldsymbol{X}^\star}$$

      satisfies

      $$F^n_{\boldsymbol{X}^\star}(n\,\boldsymbol{z}) \to G^\star(\boldsymbol{z}), \quad n \to \infty,$$

      i.e., $F_{\boldsymbol{X}^\star}$ is in the MDA of $G^\star$, where

      $$G(z_1, \dots, z_d) = G^\star(T_{\xi_1,\mu_1,\sigma_1}(z_1), \dots, T_{\xi_d,\mu_d,\sigma_d}(z_d)).$$

With standardized marginal distributions, we can choose normalizing vector sequences $\boldsymbol{a}_n^\star = (0, \dots, 0)$ and $\boldsymbol{b}_n^\star = (n, \dots, n)$.

Some remarks about max-stable dependence

- There is no exhaustive representation of all simple max-stable distributions $G^\star$ using a finite number of parameters.

- We can write $G^\star$ using the **exponent function $V^\star$**,

$$G^\star(\boldsymbol{z}) = \exp(-V^\star(\boldsymbol{z})), \quad \boldsymbol{z} > \boldsymbol{0},$$

  where $t \times V^\star(t\boldsymbol{z}) = V^\star(\boldsymbol{z})$ (**(-1)-homogeneity**).

- We say that two variables X_1 and X_2 are **asymptotically independent** if

$$G(z_1, z_2) = G_1(z_1) \times G_2(z_2),$$

  and in this case

$$G^\star(z_1, z_2) = \exp(-(1/z_1 + 1/z_2)) = \exp(-1/z_1) \times \exp(-1/z_2), \quad z_1, z_2 > 0.$$

Example: multivariate logistic distribution

A large variety of **parametric multivariate max-stable distribution** has been proposed.

The **multivariate logistic model** was introduced by Emil J. Gumbel in 1960 and can be defined through its exponent function

$$V^\star(\boldsymbol{z}) = \left(z_1^{-1/\alpha} + \ldots + z_d^{-1/\alpha}\right)^\alpha, \quad \boldsymbol{z} > \boldsymbol{0},$$

such that

$$G^\star(z_1, \ldots, z_d) = \exp\left(-\left(z_1^{-1/\alpha} + \ldots + z_d^{-1/\alpha}\right)^\alpha\right), \quad \boldsymbol{z} > \boldsymbol{0}$$

with parameter $0 < \alpha \leq 1$ and

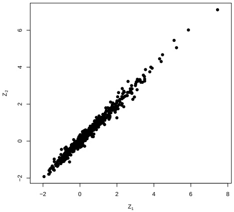
- perfect dependence for $\alpha \to 0$;
- independence for $\alpha = 1$.

Example: Simulations of bivariate logistic distribution

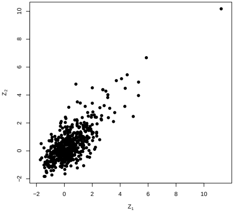
Sample size $n = 500$

Bivariate scatterplots show $\log \boldsymbol{Z}^\star$ (standard Gumbel margins) with $\boldsymbol{Z}^\star \sim G^\star$

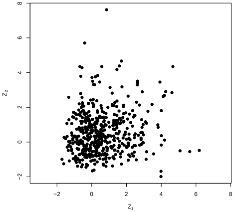
$\alpha = 0.1$



$\alpha = 0.5$



$\alpha = 0.9$



Example: Huesler–Reiss distribution

Huesler–Reiss distributions are related to multivariate Gaussian distributions. Consider a multivariate Gaussian vector $\tilde{\boldsymbol{Y}}$.

Bivariate case: the simple max-stable distribution has parameter $\gamma_{12} = \mathrm{Var}(\tilde{Y}_2 - \tilde{Y}_1) > 0$ and for $z_1, z_2 > 0$,

$$G^\star(z_1, z_2) = \exp\left(-\frac{1}{z_1}\Phi\left(\frac{\sqrt{\gamma_{12}}}{2} + \frac{1}{\sqrt{\gamma_{12}}}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{\sqrt{\gamma_{12}}}{2} + \frac{1}{\sqrt{\gamma_{12}}}\log\frac{z_1}{z_2}\right)\right)$$

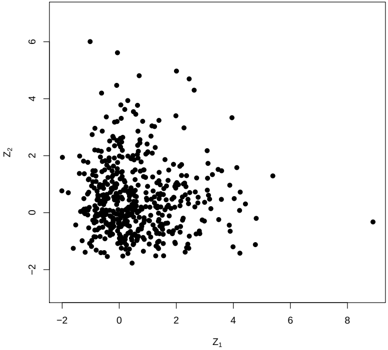
(with standard Gaussian cdf Φ)
$\Rightarrow$ independence for $\gamma_{12} \to \infty$, perfect dependence for $\gamma_{12} \to 0$

The general multivariate distribution $G^\star$ is parametrized by $d(d-1)/2$ variogram values $\gamma_{j_1,j_2} = \mathrm{Var}(\tilde{Y}_{j_2} - \tilde{Y}_{j_1})$ for $1 \le j_1 < j_2 \le d$.

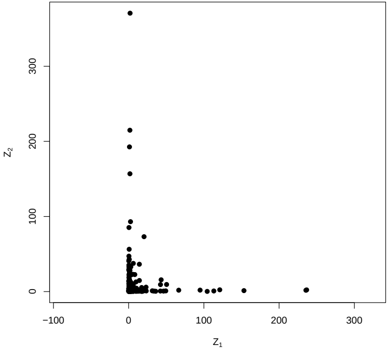
Example: Simulations of the Huesler–Reiss distribution

Sample size $n = 500$
Relatively weak dependence

log $\boldsymbol{Z}^\star$ (Gumbel margins)



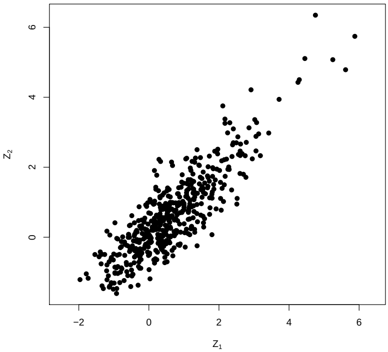
$\boldsymbol{Z}^\star$ (Fréchet margins)



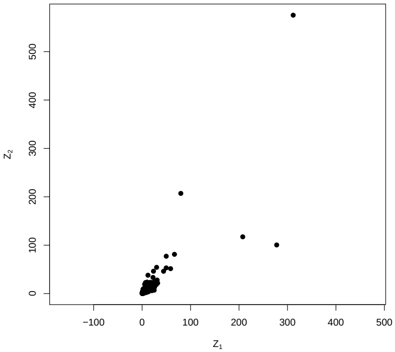
Example, cont'd

Sample size $n = 500$
Relatively strong dependence

log $\boldsymbol{Z}^\star$ (Gumbel margins)



$\boldsymbol{Z}^\star$ (Fréchet margins)



Theorem (Point-process convergence)

For i.i.d. copies $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d) \sim F$, the following two statements are equivalent:

1. The distribution F is in the multivariate MDA of the max-stable distribution G for the normalizing sequences $\boldsymbol{a}_n \in \mathbb{R}^d$ and $\boldsymbol{b}_n > \boldsymbol{0}$.
2. For the normalizing sequences $\boldsymbol{a}_n \in \mathbb{R}^d$ and $\boldsymbol{b}_n > \boldsymbol{0}$, we have the following point-process convergence with a locally finite Poisson point process limit:

$$\left\{ \frac{\boldsymbol{X}_i - \boldsymbol{a}_n}{\boldsymbol{b}_n},\ i = 1, \ldots, n \right\} \to \{\boldsymbol{P}_i,\ i \in \mathbb{N}\} \sim \mathrm{PPP}(\Lambda), \quad n \to \infty,$$

   with intensity measure Λ.

If 1) and 2) hold, then $G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z}))$ with

$$V(\boldsymbol{z}) = \Lambda\left((-\boldsymbol{\infty}, \boldsymbol{z}]^C \right),$$

where the **exponent measure Λ** is defined on $A_\Lambda = \left(\overline{A}_{\xi_1,\mu_1,\sigma_1} \times \ldots \times \overline{A}_{\xi_d,\mu_d,\sigma_d} \right) \setminus \boldsymbol{u}_\star$, with the marginal GEV parameters ξ_j, μ_j, σ_j, $j = 1, \ldots, d$, where the lower endpoint

$$\boldsymbol{u}_\star = \left(\inf A_{\xi_1,\mu_1,\sigma_1}, \ldots, \inf A_{\xi_d,\mu_d,\sigma_d} \right)$$

is excluded.

Simple representation with standardized margins

Specifically, the convergence of componentwise maxima and of point patterns is equivalent on the simple scale using standardized marginal distributions in $\boldsymbol{X}^\star$.

Recall: Standardized marginal scale

- $X_j^\star = -1/\log F_j(X_j)$ (or any other probability integral transform ensuring $X_j^\star \geq 0$ and $x \times \Pr(X_j^\star > x) \to 1$ as $x \to \infty$)
- Normalizing sequences on standardized scale are $\boldsymbol{a}_n = \boldsymbol{0}$ and $\boldsymbol{b}_n = (n, \ldots, n)$
- GEV margins of $G^\star$ are unit Fréchet $G_j^\star(z_j) = \exp(-1/z_j)$, $z_j > 0$ ($\xi_j = 1$, $\mu_j = 1$, $\sigma_j = 1$).

Simple exponent measure and homogeneity (asymptotic stability)

For any Borel set $B \subset A_\Lambda$, the **simple exponent measure $\Lambda^\star$** satisfies

$$\Lambda(B) = \Lambda^\star(B_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma}})$$

where $B_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma}} = \left\{\left(T_{\xi_1,\mu_1,\sigma_1}(x_1), \ldots, T_{\xi_d,\mu_d,\sigma_d}(x_d)\right) \mid (x_1, \ldots, x_d) \in B\right\}$. The simple measure $\Lambda^\star$ is defined on $A_{\Lambda^\star} = [0,\infty)^d \setminus \boldsymbol{0}$ and is **(-1)-homogeneous**, that is, for any Borel set $B \subset A_{\Lambda^\star}$, we have

$$t \times \Lambda^\star(tB) = \Lambda^\star(B), \quad t > 0.$$

Bivariate illustration of asymptotic stability ($D = \{1, 2\}$)

Simple scale

$(\boldsymbol{\xi} = (1,1), \boldsymbol{\mu} = (1,1), \boldsymbol{\sigma} = (1,1))$

$\boldsymbol{\alpha}_n = (n, n)$, $\boldsymbol{\beta}_n = (0, 0)$

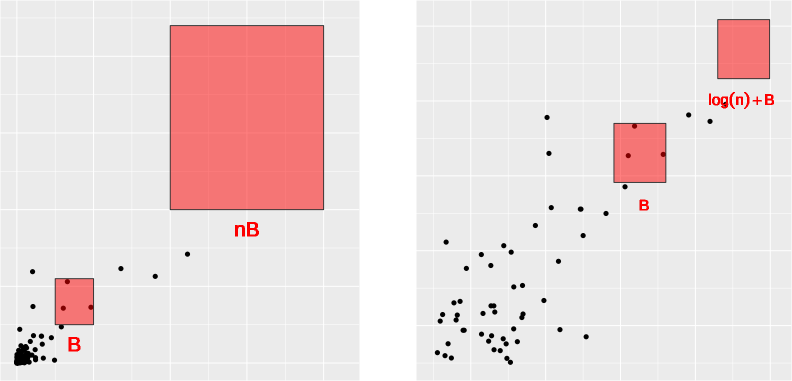
$n \times \Lambda^\star(nB) = \Lambda^\star(B)$

Standard exponential scale

$(\boldsymbol{\xi} = (0,0), \boldsymbol{\mu} = (0,0), \boldsymbol{\sigma} = (1,1))$

$\boldsymbol{\alpha}_n = (1, 1)$, $\boldsymbol{\beta}_n = (\log n, \log n)$

$n \times \Lambda(\log(n) + B) = \Lambda(B)$



The trinity: three classical multivariate formulations

- The trinity of the three classical limit also holds in the multivariate setting.
- For **threshold exceedances**, a standard approach is to condition on an exceedance in at least one of the d components.
- To avoid technical notation, we focus on the **simple setting**.

Theorem

The following three convergences are equivalent:

- Point-process convergence:

$$\left\{\frac{\boldsymbol{X}_i^\star}{n},\ i=1,\dots,n\right\} \to \{\boldsymbol{P}_i^\star,\ i\in\mathbb{N}\} \sim \mathrm{PPP}(\Lambda^\star), \quad n\to\infty.$$

- Convergence of componentwise maxima:

$$\frac{\boldsymbol{M}_n^\star}{n} \to \boldsymbol{Z}^\star \sim G^\star, \quad n\to\infty,$$

with $G^\star(\boldsymbol{z}) = \exp(-V^\star(\boldsymbol{z}))$ where $V^\star(\boldsymbol{z}) = \Lambda^\star\left([\boldsymbol{0},\boldsymbol{z}]^C\right)$.

- Peaks-Over-Threshold convergence:

$$\frac{\boldsymbol{X}^\star}{u} \mid \left(\max_{j=1}^{d} X_j^\star > u\right) \to \boldsymbol{Y}^\star \sim \frac{\Lambda^\star(\,\cdot \cap [\boldsymbol{0},\boldsymbol{1}]^C)}{\Lambda^\star\left([\boldsymbol{0},\boldsymbol{1}]^C\right)}, \quad u\to\infty.$$

Remarks about the functional setting

- The trinity of limits also holds in the functional setting (e.g., Dombry & Ribatet, 2016).
- Usually one considers $\boldsymbol{X} \in \mathcal{C}(D)$ with compact domain D.
- One has to appropriately define weak convergence in a Banach function space.

The spectral construction of simple processes

Spectral representation of simple point processes

Any Poisson point process $\{\boldsymbol{P}_i^\star,\ i \in \mathbb{N}\}$ with simple ((-1)-homogeneous) intensity measure $\boldsymbol{\Lambda}^\star$ can be constructed as follows:

$$\{P_i^\star(s),\ i \in \mathbb{N}\} = \{R_i W_i(s),\ i \in \mathbb{N}\}$$

where $R_i = 1/U_i$ and

- $0 < U_1 < U_2 < ....$ are the points of a unit-rate Poisson process on $[0, \infty)$, and
- $\boldsymbol{W}_i = \{W_i(s)\}$ are iid nonnegative random functions, independent of $\{U_i\}$, with $\mathbb{E}W_i(s) = 1$ and $\mathbb{E}W_i(s)^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$.

A consequence of this is the spectral representation of simple max-stable processes.

Spectral representation of the simple max-stable processes

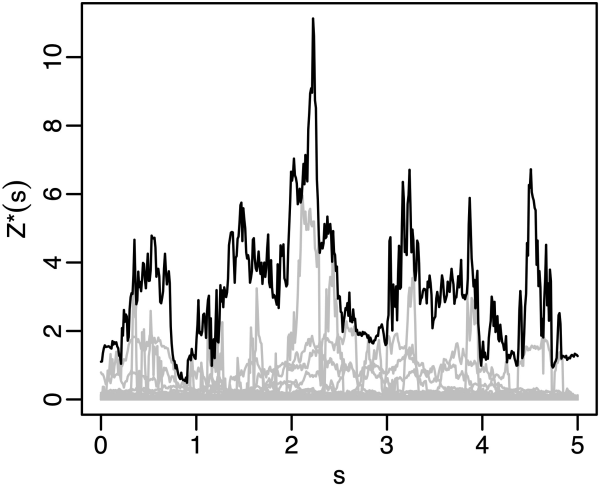
With notations as above, any simple max-stable process $\boldsymbol{Z}^\star$ can be constructed as

$$Z^\star(s) = \max_{i \in \mathbb{N}} R_i W_i(s),$$

and any such construction is a simple max-stable process.

Illustration: simple max-stable construction

- In gray, "points" $\boldsymbol{P}_i^\star$ of the Poisson process on $D = [0, 5]$
- Max-stable process is the componentwise maximum (in black)



Simulation based on the spectral representation

If it is simple to simulate from the distribution $F_{\boldsymbol{W}}$ of the spectral process $\boldsymbol{W}$, we can draw samples from the simple max-stable process $\boldsymbol{Z}^\star$.

Exact simulation

If $P(W_j \leq w_0) = 1$ for some threshold value $0 < w_0 < \infty$, $j = 1, \ldots, d$, then we can perform **exact simulation of $\boldsymbol{Z}^\star$** (even if $Z_j^\star = \max_{i \in \mathbb{N}} R_i W_{ij}$ is defined as a maximum over an infinite number of components):

1. set $m = 1$
2. generate $E_m \sim \mathrm{Exp}(1)$
3. generate $\boldsymbol{W}_m = (W_{m1}, \ldots, W_{md})^T \sim F_{\boldsymbol{W}}$
4. set $\boldsymbol{Z}^\star = (Z_1^\star, \ldots, Z_d^\star)^T$ with $Z_j^\star = \max_{i=1,\ldots,m} \frac{W_{ij}}{\sum_{k=1}^{i} E_k}$ for $j = 1, \ldots, d$
5. IF $\frac{w_0}{\sum_{k=1}^{m} E_k} \leq \min_{j=1,\ldots,d} Z_j^\star$ RETURN $\boldsymbol{Z}^\star$
   ELSE set $m = m + 1$ and GO TO 2

Remarks:

- If the distribution of W_j is not finitely bounded, we can fix w_0 such that $P(W_j > y_0)$ becomes very small and perform approximation simulation.
- Even with unbounded W_j, exact simulation remains possible for many models using different algorithms (see the review of Oesting, Strokorb, 2022).

Example: Log-Gaussian spectral processes

A possible construction uses a **centered Gaussian process** $\tilde{W}(s)$ with variance function $\sigma^2(s)$ and sets

$$W(s) = \exp(\tilde{W}(s) - \sigma^2(s)/2)$$

$\Rightarrow$ **A class of popular max-stable models:**

- Multivariate: **Huesler–Reiss distributions**
- Spatial: **Brown–Resnick processes**

Remark: The distribution of the simple max-stable process $\boldsymbol{Z}^\star = \{Z^\star(s), \, s \in D\}$ depends only on the variogram

$$\gamma(s_1, s_2) = \mathrm{Var}(\tilde{W}(s_2) - \tilde{W}(s_1)), \quad s_1, s_2 \in D.$$

Illustration: Simulation of Brown–Resnick processes

Two realisation of a spatial Brown-Resnick process
(obtained using the `rmaxstab` function of the `SpatialExtremes` package)
Simulation on a grid 20×20 (such that $d = 400$) in the square $[0, 10]^2$.

Illustration: process $\log(Z^\star(s))$ (with standard Gumbel margins)

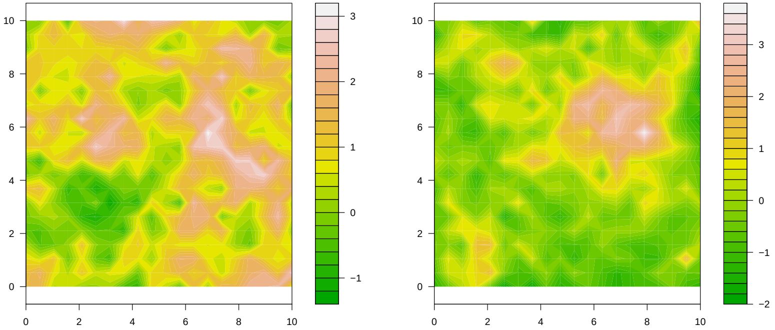
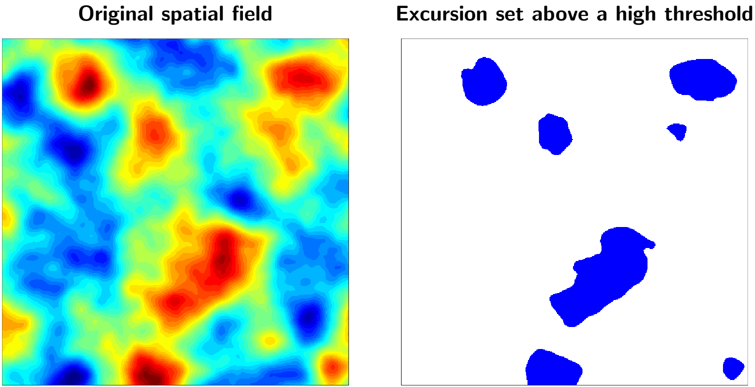


Illustration: Spatial co-occurrence of exceedances



Assessing co-occurrences of threshold exceedances

Threshold exceedances can occur simultaneously,

- in different variables,
- at nearby locations,
- at close time steps.

Do co-occurrences happen by chance (independence), or are they correlated in some way?

A simple and flexible exploratory approach

Idea: Study **pairwise conditional co-occurrence probabilities** given as

$$\Pr(X_2 > u \mid X_1 > u) = \frac{\Pr(X_1 > u,\, X_2 > u)}{\Pr(X_1 > u)},$$

and assess how they change with increasing u and for different pairs, for instance with respect to temporal lag or spatial distance.

Tail correlation coefficient

Consider a bivariate random vector (X_1, X_2) with $X_1 \sim F_1$ and $X_2 \sim F_2$.

Tail correlation

Consider the conditional probability

$$\chi(u) = \Pr(F_2(X_2) > u \mid F_1(X_1) > u) = \frac{\Pr(F_2(X_2) > u, F_1(X_1) > u)}{\Pr(F_1(X_1) > u)}, \quad u \in (0,1).$$

We define the following limit (if it exists):

$$\chi = \lim_{u \to 1} \chi(u) \in [0,1]$$

The coefficient χ symmetric with respect to X_1 and X_2 and is known as **χ-measure** or **tail correlation**. We say that

- X_1 and X_2 are **asymptotically dependent** if $\chi > 0$;
- X_1 and X_2 are **asymptotically independent** if $\chi = 0$.

Link between tail correlation and max-stability

We have

$$\chi = \lim_{z\to\infty} \Pr(X_2^\star > z \mid X_1^\star > z) = \lim_{z\to\infty} \frac{\Pr(X_1^\star > z, X_2^\star > z)}{\Pr(X_1^\star > z)} \quad (\star)$$

Assume that (X_1, X_2) is in the MDA of G. The bivariate max-stable convergence

$$F_{(X_1^\star, X_2^\star)}(nz, nz)^n \to G^\star(z, z), \quad z > 0,$$

is equivalent to

$$1 - F_{(X_1^\star, X_2^\star)}(nz, nz) \approx -\log G^\star(nz, nz), \qquad \text{for large } n.$$

By using

$$\Pr(X_1^\star > z, X_2^\star > z) = (1 - F_{X_1^\star}(z)) + (1 - F_{X_2^\star}(z)) - (1 - F_{(X_1^\star, X_2^\star)}(z, z)),$$

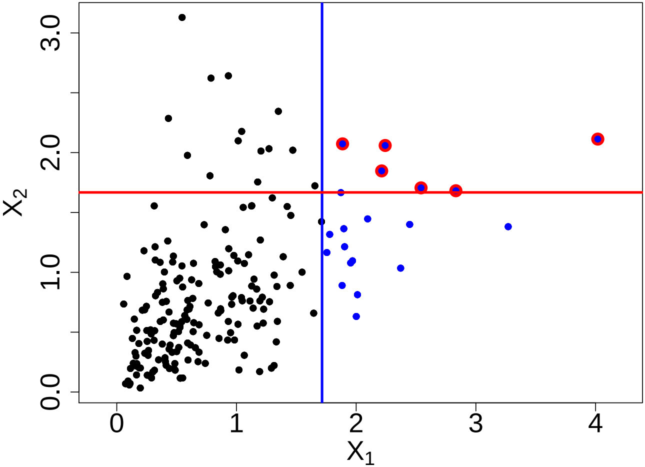
and $-\log G^\star(nz, nz) = \frac{V^\star(1,1)}{nz}$ and $1 - G_j^\star(nz) \approx 1/(nz)$ in $(\star)$, we obtain

$$\chi = 2 - V^\star(1, 1).$$

Remark: asymptotic independence corresponds to classical independence in the max-stable limit distribution. We have $\chi = 0$ if and only if $V^\star(1,1) = 2$, and in this case $V^\star(z_1, z_2) = 1/z_1 + 1/z_2$ for $z_1, z_2 > 0$, and $G^\star(z_1, z_2) = G_1^\star(z_1) \times G_2^\star(z_2)$.

Illustration: empirical tail correlation

Data setting: $n = 200$, $u = 0.9$.
Blue points: exceedances of empirical distribution function $\hat{F}_1(X_1)$ above u.
Red points: exceedances of $\hat{F}_2(X_2)$ above u given that $\hat{F}_1(X_1)$ is above u.
Empirical tail correlation: $\hat{\chi}(u) = \frac{6}{20} = 0.3$.



Tail autocorrelation function (Extremogram)

Consider $X(s) \sim F$ with index $s \in \mathbb{R}^k$.

What is the tail correlation at a given distance $h = \Delta s \geq 0$?

For $h \geq 0$, we consider the conditional exceedance probability

$$\chi(h; u) = \Pr(F(X(s+h)) > u \mid F(X(s)) > u) = \frac{\Pr(F(X(s+h)) > u, F(X(s)) > u)}{\Pr(F(X(s)) > u)},$$

for $u \in (0, 1)$.

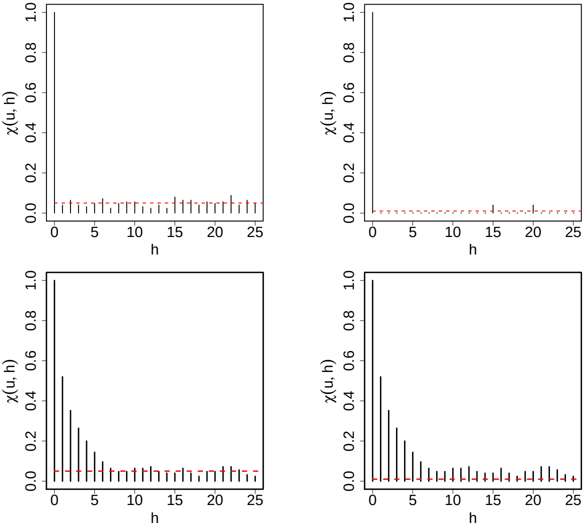
We define the **tail autocorrelation function** as the limit (if it exists)

$$\chi(h) = \lim_{u \to 1} \chi(h; u) \in [0, 1].$$

- By definition, $\chi(0) = 1$.
- Usually, $\chi(h)$ decreases as $\|h\|$ increases.
- $\chi(h)$ is also called **auto-tail dependence function** or **extremogram**.

Illustration: Empirical (temporal) extremogram

Top row: temporal independence in $X(t)$; bottom row: asymptotic dependence
Left column: $u = 0.95$; right column: $u = 0.99$
Dashed red line corresponds to theoretical $\chi(h; u)$ for independence.



Summary measures for more than two variables

Consider d random variables $X_1, X_2, \dots, X_d$ with $d \geq 2$ and $X_j \sim F_j$.

Extremal coefficient (maxima)

The following limit (if it exists) is called **extremal coefficient**:

$$\theta_d = \lim_{u \to \infty} u \times \Pr\left(\max_{j=1,\dots,d} X_j^\star > u \right)$$

- $\theta_d = V(1, \dots, 1)$
- $\theta_2 = 2 - \chi$.
- **Interpretation:** d/θ_d = **average cluster size** of jointly extreme events
- With MDA convergence, we have $G^\star(z^\star, \dots, z^\star) = \exp(-\theta_d / z^\star)$, $z^\star > 0$.

Tail dependence coefficient (minima)

The following limit (if it exists) is called **tail dependence coefficient**:

$$\lambda_d = \lim_{u \to \infty} \Pr\left(\min_{j=1,\dots,d} X_j^\star > u \mid X_1^\star > u \right) = \lim_{u \to \infty} u \times \Pr\left(\min_{j=1,\dots,d} X_j^\star > u \right)$$

- For $d = 2$, we have $\lambda_2 = \chi$.
- Extremal coefficients and tail dependence coefficients are linked through inclusion-exclusion formulas using coefficients for $\tilde{d} = 2, \dots, d$.

So far...

- Summary measures for co-occurrences of threshold exceedances
- Focus on minima and maxima of the components of $\boldsymbol{X}^\star$

Next...

- More flexibility through more general risk functionals
- Generative and parametric models, not only summaries

Multivariate and functional threshold exceedances

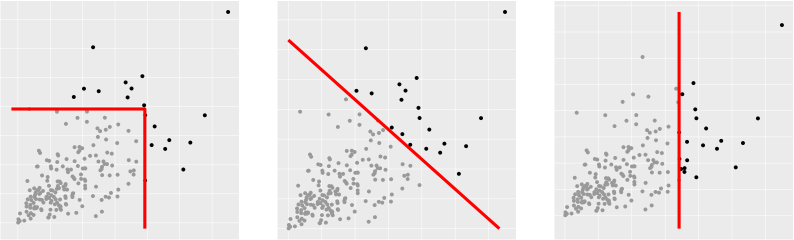
Consider $\boldsymbol{x} \in \mathbb{R}^D$ for a compact domain $D \subset \mathbb{R}^k$ with $|D| > 1$.
Note: for a vector $\boldsymbol{x} = (x_1, \dots, x_d)$, we can set $D = \{1, \dots, d\}$.

No unique definition of threshold exceedances $\Rightarrow$ Use a **risk functional** r

Extreme event occurs if $r(\boldsymbol{x}) > u$ with high threshold u

Bivariate illustrations:

Maximum $r(x_1, x_2) = \max(x_1, x_2)$ | **Average** $r(x_1, x_2) = x_1 + x_2$ | **Fixed component** $r(x_1, x_2) = x_1$



Many relevant choices for risk functionals

To formulate asymptotic theory,
we use continuous homogeneous risk functionals

$$r : [0, \infty)^D \to [0, \infty), \quad r(t \times \boldsymbol{x}) = t \times r(\boldsymbol{x})$$

and **we apply r on the simple scale.**
We further assume continuous realizations: $\boldsymbol{x} \in \mathcal{C}(D)$.
There is also notation ℓ (for *loss*) instead of r (for *risk*).

Examples for $D = \{1, 2, \ldots, d\}$

- Minimum: $r(x_1, \ldots, x_d) = \min_{j=1}^d x_j$
- Maximum: $r(x_1, \ldots, x_d) = \max_{j=1}^d x_j$
- k^{th} order statistics: $r(x_1, \ldots, x_d) = k^{th}$ smallest value among $x_1, \ldots, x_d$
- Specific component: $r(x_1, \ldots, x_d) = x_{j_0}$
- Arithmetic average: $r(x_1, \ldots, x_d) = \frac{1}{d} \sum_{j=1}^d x_j$
- Geometric average $r(x_1, \ldots, x_d) = \left(\prod_{j=1}^d x_j\right)^{1/d}$
- Any norm, such as $r(x_1, \ldots, x_d) = \left(\sum_{j=1}^d x_j^p\right)^{1/p}$

Comparison of arithmetic and geometric average

Arithmetic average:

$$r(x_1, \ldots, x_d) = \frac{1}{d} \sum_{j=1}^{d} x_j$$

Geometric average:

$$r(x_1, \ldots, x_d) = \left(\prod_{j=1}^{d} x_j \right)^{1/d}$$

- Constant values $x_1 = \ldots = x_d \Rightarrow$ Geometric $=$ Arithmetic average
- Stronger variability in values x_j leads to relatively lower Geometric average

How to standardize marginal distributions (recall + extension)

Given $X_j \sim F_j$ with continuous distribution function F_j, we apply a probability integral transform to a standardized scale $X_j^\star$ satisfying

- $X_j^\star \geq 0$, and
- $x \times \Pr(X_j^\star > x) \to 1$ as $x \to \infty$, which means $\Pr(X_j^\star > x) \approx 1/x$ for large x

Two common choices

- **Unit Fréchet scale:** $X_j^\star = -\frac{1}{\log} F(_j(X_j))$
  (makes sense when working with maxima since the unit Fréchet is a GEV)
- **Standard Pareto scale:** $X_j^\star = 1/(1 - F_j(X_j))$
  (makes sense when working with exceedances since the standard Pareto is a GPD)

Interpretation of $X_j^\star$ as the (approximate) return period of X_j:
for an independent copy $\overline{X}_j$ of X_j, we get

$$\Pr(\overline{X}_j > X_j \mid X_j) \approx \frac{1}{X_j^\star} \quad \text{for relatively large } X_j$$

(**Note:** If $\Pr(A) = 1/T$, then the event A has a return period of T time units)

Limits conditional to risk exceedances $r(\boldsymbol{X}) > u$

r-Pareto limit processes (Dombry & Ribatet 2015)

Consider a random element $\boldsymbol{X} = \{X(s), s \in D\} \subset \mathcal{C}(D)$ with compact domain D.

- If we have the following (weak) convergence in $\mathcal{C}(D)$,

$$\frac{\boldsymbol{X}^\star}{u} \mid (r(\boldsymbol{X}^\star) > u) \quad \to \quad \boldsymbol{Y}_r, \quad u \to \infty,$$

  then $\boldsymbol{Y}_r$ is an **r-Pareto process**,
  satisfying **Peaks-Over-Threshold stability**:

$$\frac{\boldsymbol{Y}_r}{u} \mid (r(\boldsymbol{Y}_r) > u) \quad \stackrel{d}{=} \quad \boldsymbol{Y}_r, \quad \text{for any } u > 1.$$

- r-Pareto processes are characterized by a **scale-profile decomposition:**

$$\boldsymbol{Y}_r = R \times \boldsymbol{V}, \quad R = r(\boldsymbol{Y}_r) \sim \text{standard Pareto}, \quad \boldsymbol{V} = \frac{\boldsymbol{Y}_r}{r(\boldsymbol{Y}_r)}, \quad R \perp \boldsymbol{V}$$

$\Rightarrow$ **Above high thresholds u, scale $r(\boldsymbol{X}^\star)$ and profile $\boldsymbol{X}^\star / r(\boldsymbol{X}^\star)$ become independent!**

Link to other limits

- **Trinity of limits:**

Convergence of componentwise maxima

$\Leftrightarrow$

Point-process convergence

$\Leftrightarrow$

r**-Pareto convergence for** $r = \sup$

- r-Pareto convergence for sup $\Rightarrow$ r-Pareto convergence for all r

- The **probability measure of the** r**-Pareto process** $\boldsymbol{Y}_r$ is

$$\boldsymbol{Y}_r \sim \frac{\Lambda^\star(\,\cdot \cap A_r)}{\Lambda^\star(A_r)} \quad \text{with } A_r = \{\boldsymbol{y} \in \mathcal{C}(D) \mid r(\boldsymbol{y}) \geq 1\}$$

- Consider the simple point-process limit $\{\boldsymbol{P}_i^\star,\ i \in \mathbb{N}\}$

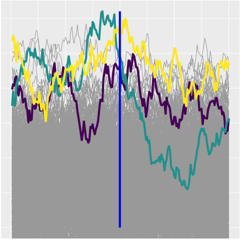
  $\Rightarrow$ **Construction of** r**-Pareto processes** $\hat{=}$ **Extraction of** r**-exceedances:**

$$\boldsymbol{P}_i^\star \mid (r(\boldsymbol{P}_i^\star) > 1) \quad \stackrel{d}{=} \quad \boldsymbol{Y}_r$$

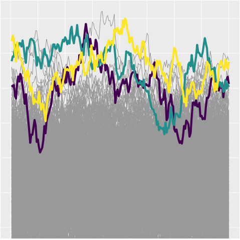
Illustration: Simulation of r-Pareto processes

- Same realizations of the Poisson point process in all three displays
- Colors correspond to 3 most extreme risks for **different risk functionals** r
- Illustrations are on the $\log((\cdot)^{\star})$-scale (Gumbel scale)

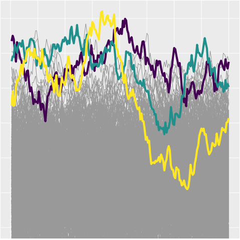
r = Value at fixed location



r = Geometric Average



r = Arithmetic Average



Example: Geometric average risk for Brown–Resnick models

The popular **Huesler–Reiss** and **Brown–Resnick models** have log-Gaussian profile processes $\boldsymbol{V}$ for r chosen as the geometric average.

This is very convenient for statistical methods!

Recall: Poisson process has construction $\{P_i^\star(s)\} = \{R_i \exp(\tilde{W}_i(s) - \sigma^2(s))\}$ with a centered Gaussian process $\tilde{\boldsymbol{W}}$ with variance function $\sigma^2(s)$

Log-Gaussian profile processes for r = Geometric average

Given the Pareto process $\boldsymbol{Y}_r = R \times \boldsymbol{V}$, we have

$$\log V(s) \quad \stackrel{d}{=} \quad \tilde{W}(s) - \overline{W} - \mathrm{const}(s;\Gamma)$$

with

- a centered Gaussian process $\tilde{\boldsymbol{W}} = \{\tilde{W}(s),\, s \in D\}$ and its spatial average $\overline{W}$,
- a constant $\mathrm{const}(s;\Gamma)$, explicit in terms of the semivariogram matrix

$$\Gamma = \{\gamma(s_1, s_2),\ s_1, s_2 \in D\},$$

of $\tilde{\boldsymbol{W}}$.

(Result follows from Engelke et al. 2012; Dombry et al. 2016; Engelke et al. 2019)

Semivariograms for log-Gaussian profile processes

Recall: The log-profile process is

$$\log V(s) = \tilde{W}(s) - \overline{W} - \mathrm{const}(s; \Gamma)$$

Same semivariograms of the log-profile $\log \boldsymbol{V}$ **and the original Gaussian process** $\tilde{\boldsymbol{W}}$**!**

$$\gamma_{\log \boldsymbol{V}}(s_1, s_2) = \frac{1}{2}\mathbb{V}\left[\log V(s_2) - \log V(s_1)\right] = \frac{1}{2}\mathbb{V}\left[\tilde{W}(s_2) - \tilde{W}(s_1)\right] = \gamma_{\tilde{\boldsymbol{W}}}(s_1, s_2)$$

$\Rightarrow$ **Classical variogram analysis becomes possible for** $\log \boldsymbol{V}$**!**

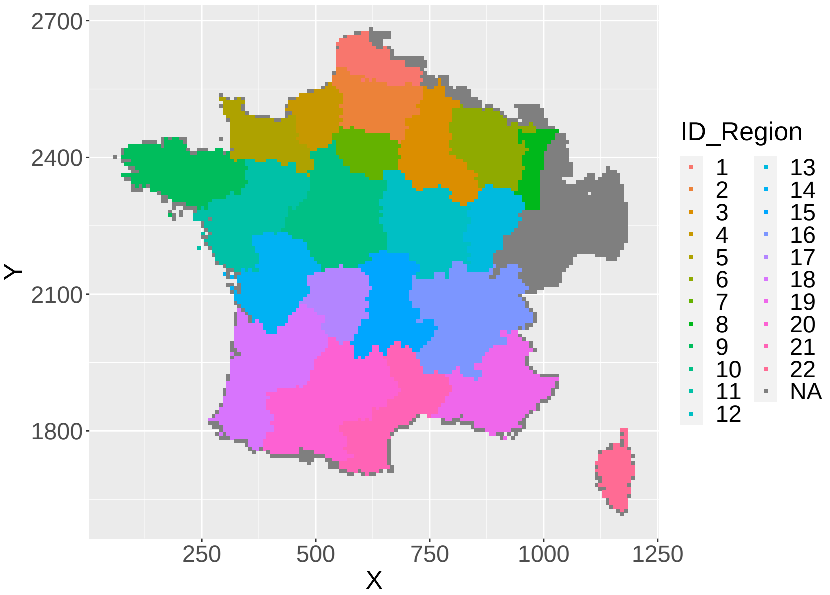
Gridded temperature reanalysis data

- Daily average temperature reanalysis of Météo France (SAFRAN model at $8km$ resolution)
- Study period 1991–2020
- Focus on summer temperatures (June-September)

Modeling approach

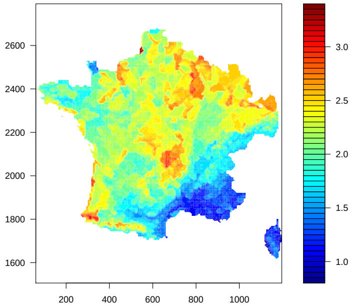
- Marginal transformation to **standard Pareto**
- We fit separate r-Pareto models for separate **administrative regions**
- Daily risk exceedances using **Geometric Average of return periods**
- **Temporal declustering** with runs method for the risk series $r(\boldsymbol{X}_t^\star)$
- **Maximum likelihood** using $\log \boldsymbol{V}$ with a stable covariance function in $\tilde{\boldsymbol{W}}$

Study domain: 22 French administrative regions

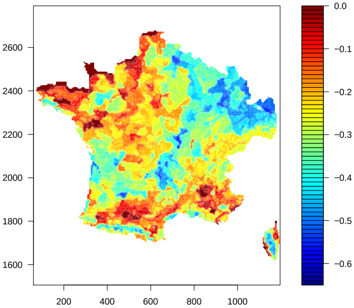


Results: Marginal GPD parameters

Scale $\sigma_{GP}(s)$



Shape $\xi(s)$



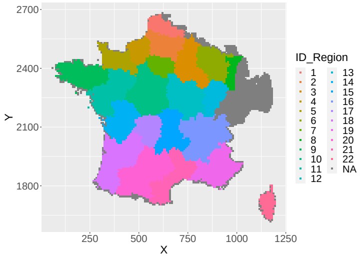
Results: Estimated extremal variograms

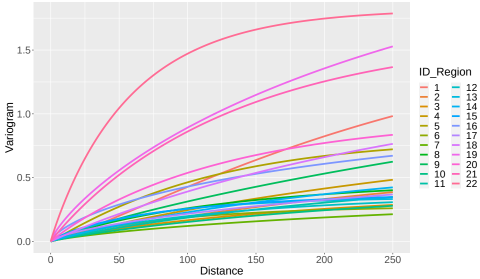
Based on the **stable covariance function**

$$\mathrm{Cov}(\mathsf{Distance}) = \mathsf{SD}^2 \times \exp\left(-(\mathsf{Distance}/\mathsf{Scale})^{\mathsf{Shape}}\right)$$

(for Distance $= \|\Delta s\| = \|s_2 - s_1\|$)

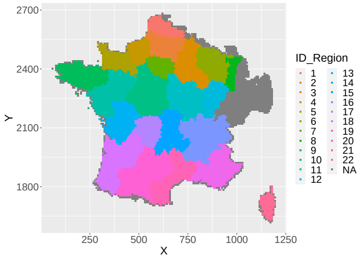
and maximum likelihood estimation using observations of log $\boldsymbol{V}$

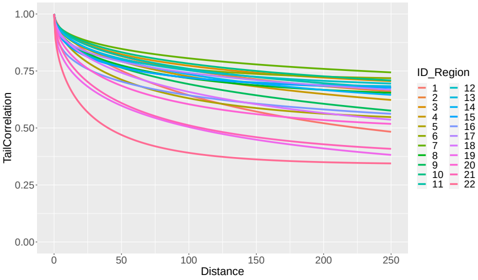




Results: Estimated tail correlations

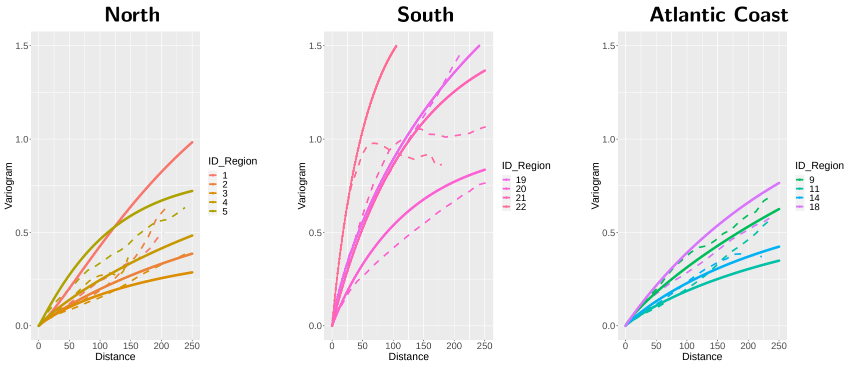
$$\chi(s, s+\Delta s) = \lim_{u \to \infty} \Pr(X^P(s+\Delta s) > u \mid X^P(s) > u) = 2\left(1 - \Phi\left(\sqrt{(\gamma(s, s+\Delta s)}\right)\right)$$





Results: Empirical/parametric extremal variograms

- Empirical (in dashed lines) and fitted parametric variograms
  ⚠ Parametric estimates exploit also the Gaussian mean $\mathrm{const}(s;\Gamma)$
- Generally satisfactory fit
- During extreme heat days, stronger spatial variability in the Southern regions



Statistical aspects of extreme-value analysis

In practice, we typically have observations of a **sample** $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ **with** n **fixed**.

- Most approaches exploit one of three classical representations: block maxima; threshold exceedances; point patterns.
- Peaks-over-threshold methods offer high flexibility, especially by using risk functionals for dependent extremes.
- We assume that extreme-value limits provide a good approximation for large n or high threshold u.
- **Bias-variance tradeoff** in statistical estimation:

  Higher threshold or Larger block $\Leftrightarrow$ Less bias but higher variance

- Rough distinction between likelihood-based (parametric) approaches and other "semi-parametric" approaches
- Likelihood approaches for dependent extremes usually require calculating $\Lambda(A_r)$ for some risk region A_r, which can be computationally very costly, or even prohibitive if $|D|$ is large.

Important topics in current extreme-value research

Types of extreme events:

- Application of risk functionals r directly to $\boldsymbol{X}$ and not to standardized $\boldsymbol{X}^\star$
- Improved analysis of **nonstationary extremes**, especially for applications to **climate change**
- **Compound extremes** (in the climate and risk literature)
  - Aggregation of not necessarily extreme components leads to extreme impacts
  - Example: Persistence of relatively high temperatures and low precipitation leads to extreme drought conditions
- **Subasymptotic extremal dependence** that is not stable at observed levels

  $\Rightarrow$ Non-asymptotic representations and statistical garantuees?

Methods and algorithms:

- **Machine Learning** for extreme events
- **Scalability** of algorithms to large datasets, such as climate-model simulations

Some literature for further reading

Theory and probabilistic foundation:

- Resnick (1987). Extreme Values, Regular Variation and Point Processes.

Statistical modeling:

- Coles (2001). An introduction to statistical modeling of extreme values.

Mix of both:

- Embrechts, Klüppelberg, Mikosch (1997). Modelling extremal events: for insurance and finance.
- de Haan, Ferreira (2006). Extreme-value theory: an introduction.

A review of available software (R-based):

- Belzile, Dutang, Northrop, Opitz (2023+). A modeler's guide to extreme-value software. `https://arxiv.org/abs/2205.07714`