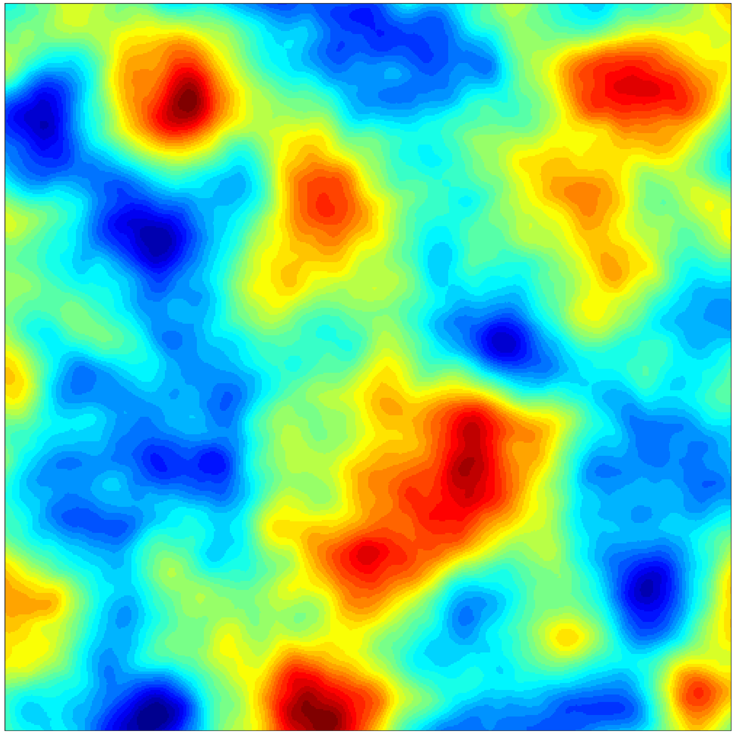
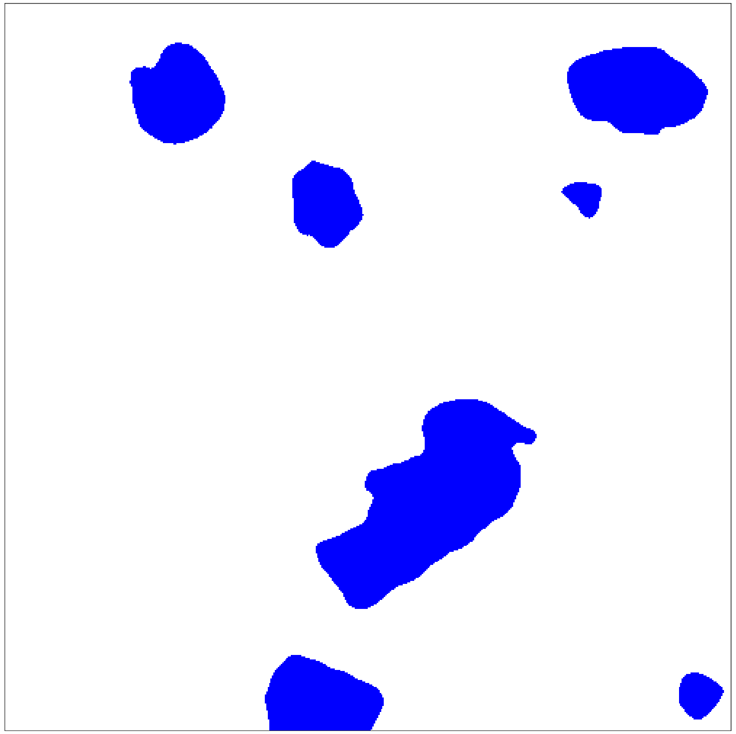
Illustration: Spatial co-occurrence of exceedances

Original spatial field



Excursion set above a high threshold



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

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

By using

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

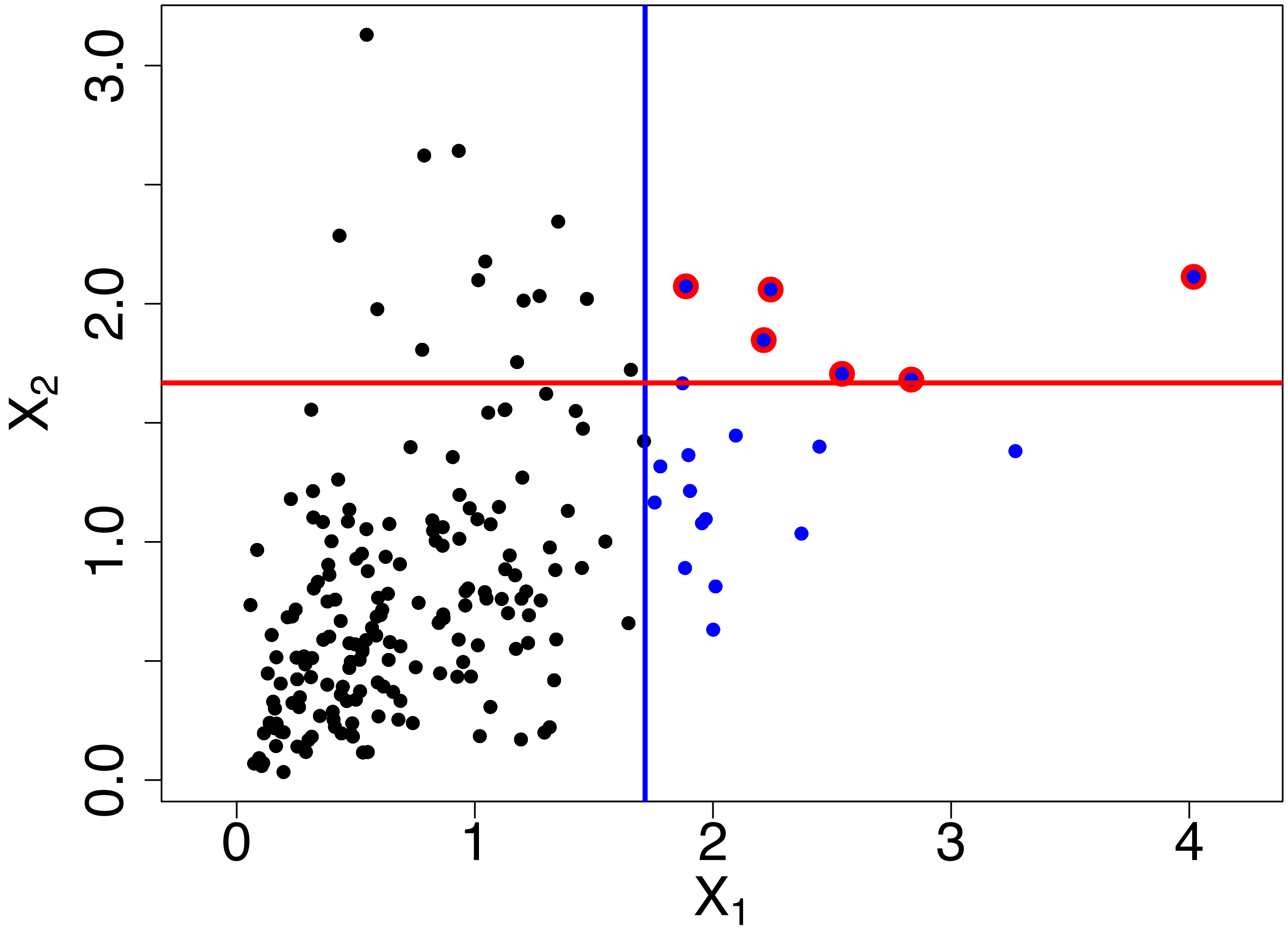
and $-\log G^\star(nz, nz) = \frac{V^\star(1,1)}{nz}$ and $1 - G_j^\star(nz) \approx 1/(nz)$ in $(\star)$, we obtain

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

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 $u \in (0, 1)$.

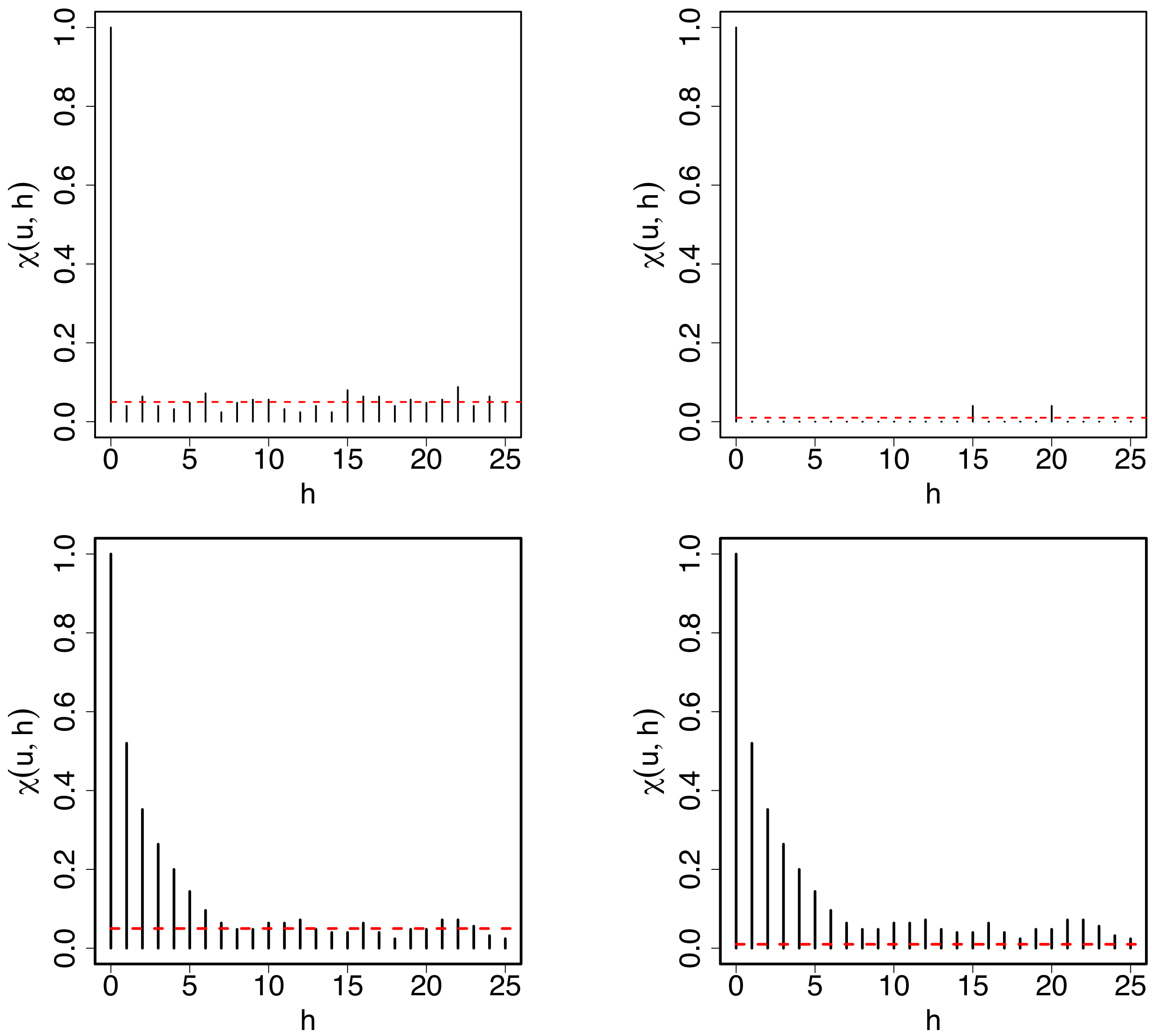
We define the **tail autocorrelation function** as the limit (if it exists)

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

- 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 $X_1, X_2, \ldots, X_d$ with $d \geq 2$ and $X_j \sim F_j$.

Extremal coefficient (maxima)

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

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

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

Tail dependence coefficient (minima)

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

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

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

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

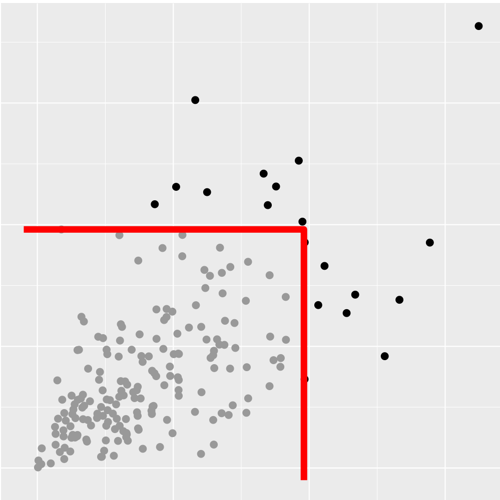
Consider $\boldsymbol{x} \in \mathbb{R}^D$ for a compact domain $D \subset \mathbb{R}^k$ with $|D| > 1$.
Note: for a vector $\boldsymbol{x} = (x_1, \ldots, x_d)$, we can set $D = \{1, \ldots, d\}$.

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

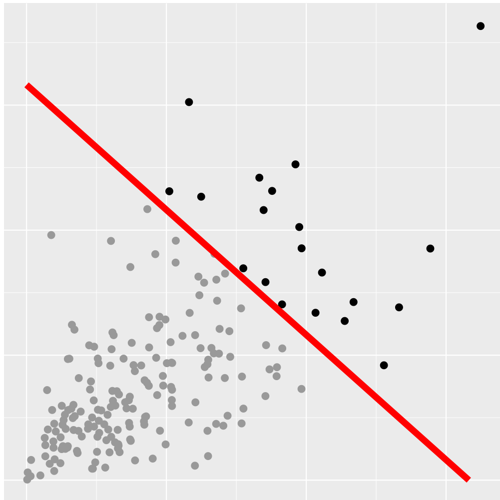
Extreme event occurs if $r(\boldsymbol{x}) > u$ with high threshold u

Bivariate illustrations:

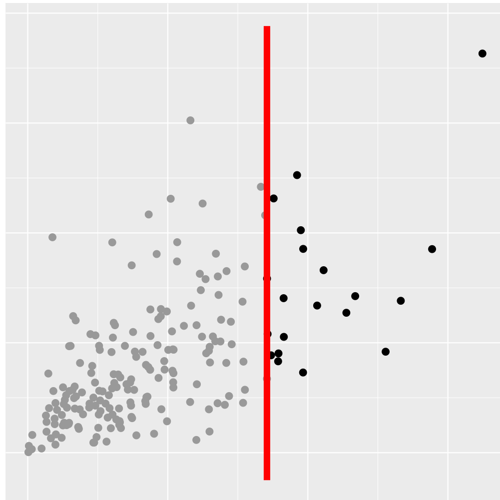
Maximum
$r(x_1, x_2) = \max(x_1, x_2)$



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 $\mathcal{C}(D)$,

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

  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 $\{\boldsymbol{P}_i^{\star},\ i \in \mathbb{N}\}$

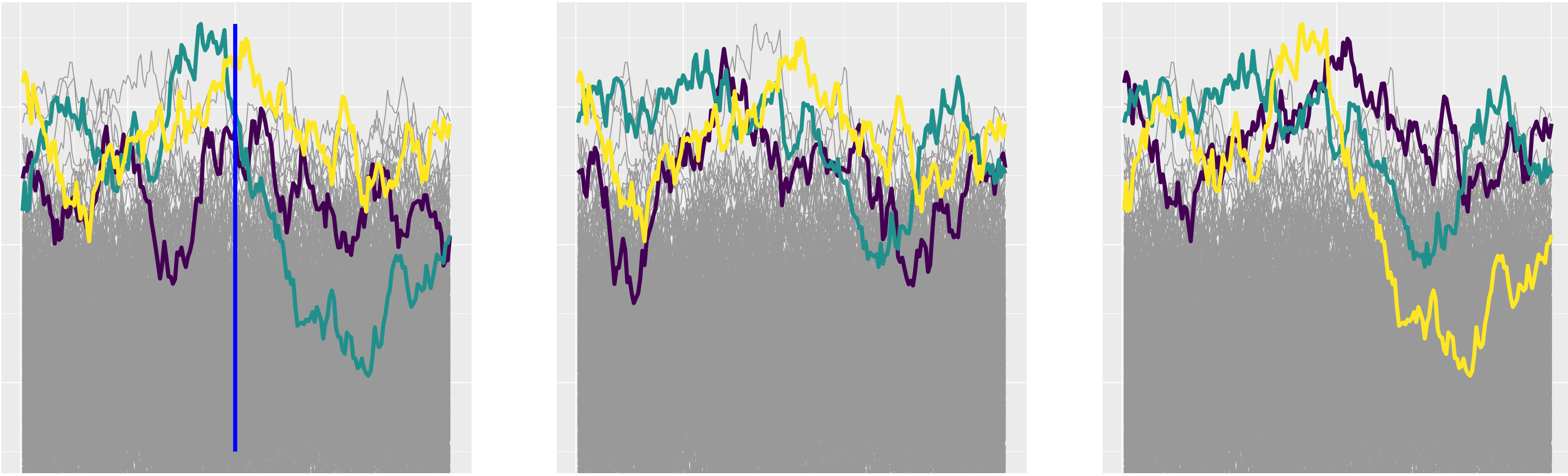
  $\Rightarrow$ **Construction of** r**-Pareto processes** $\hat{=}$ **Extraction of** r**-exceedances:**

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

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 $\quad$ r = Geometric Average $\quad$ 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 $\log \boldsymbol{V}$!**

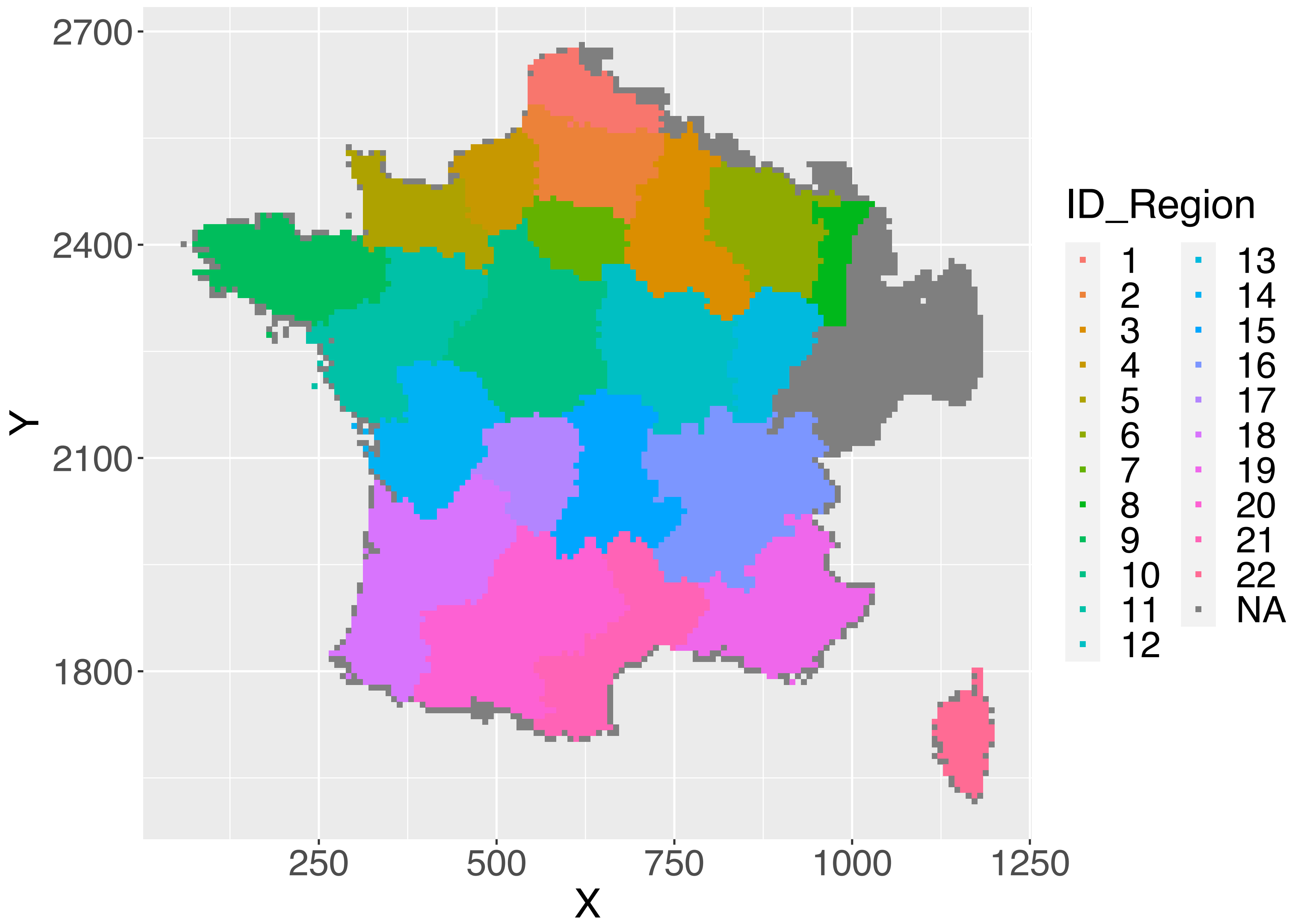
Gridded temperature reanalysis data

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

Modeling approach

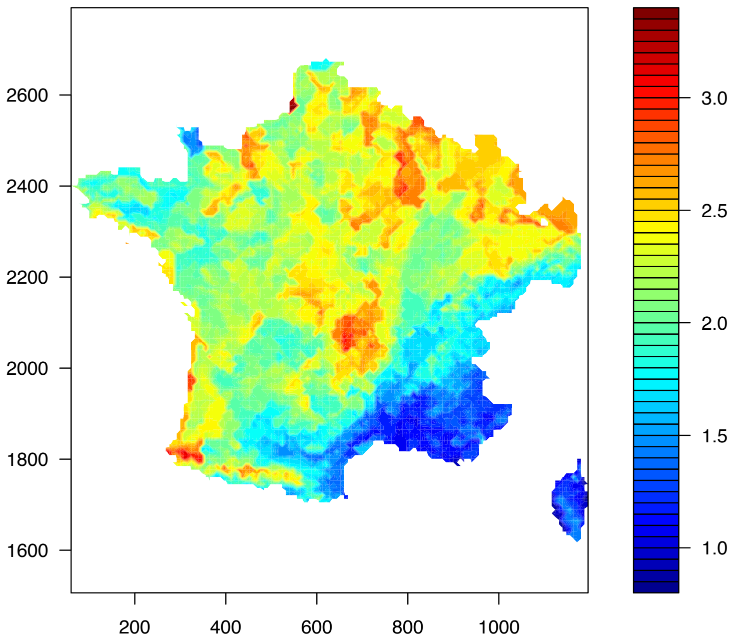
- Marginal transformation to **standard Pareto**
- We fit separate r-Pareto models for separate **administrative regions**
- Daily risk exceedances using **Geometric Average of return periods**
- **Temporal declustering** with runs method for the risk series $r(\boldsymbol{X}_t^\star)$
- **Maximum likelihood**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

Study domain: 22 French administrative regions

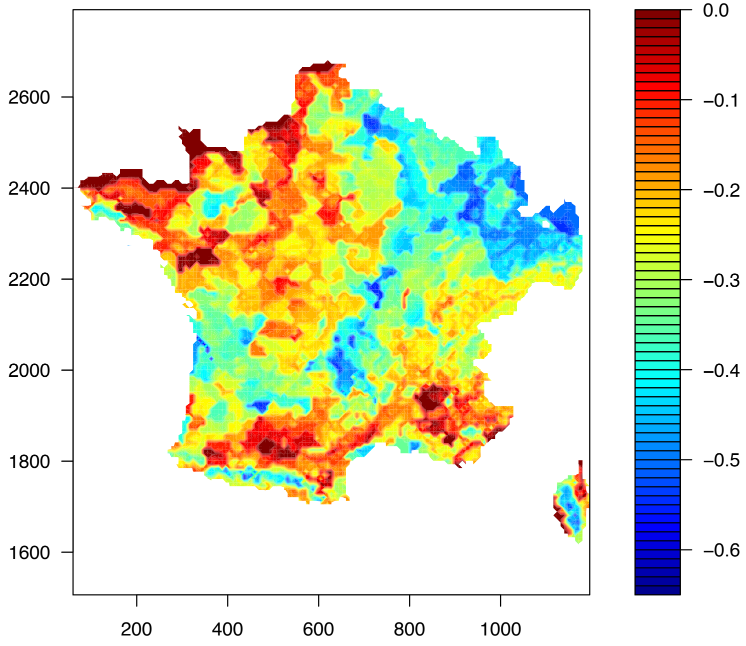


Results: Marginal GPD parameters

Scale $\sigma_{GP}(s)$



Shape $\xi(s)$



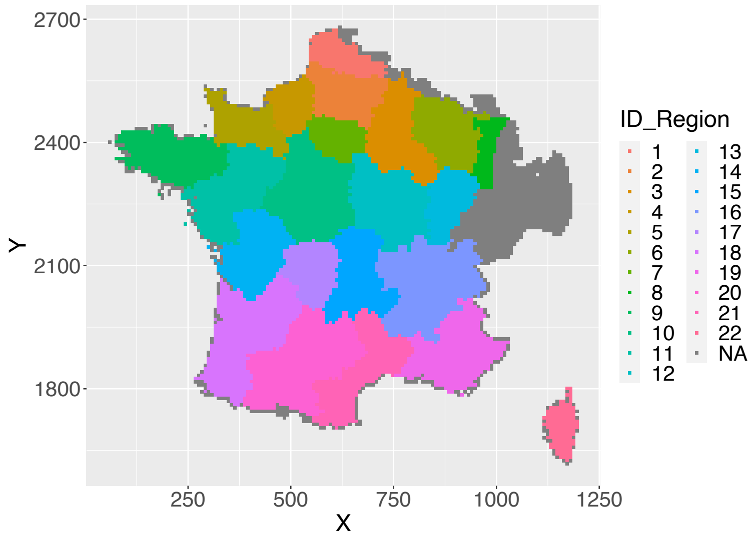
Results: Estimated extremal variograms

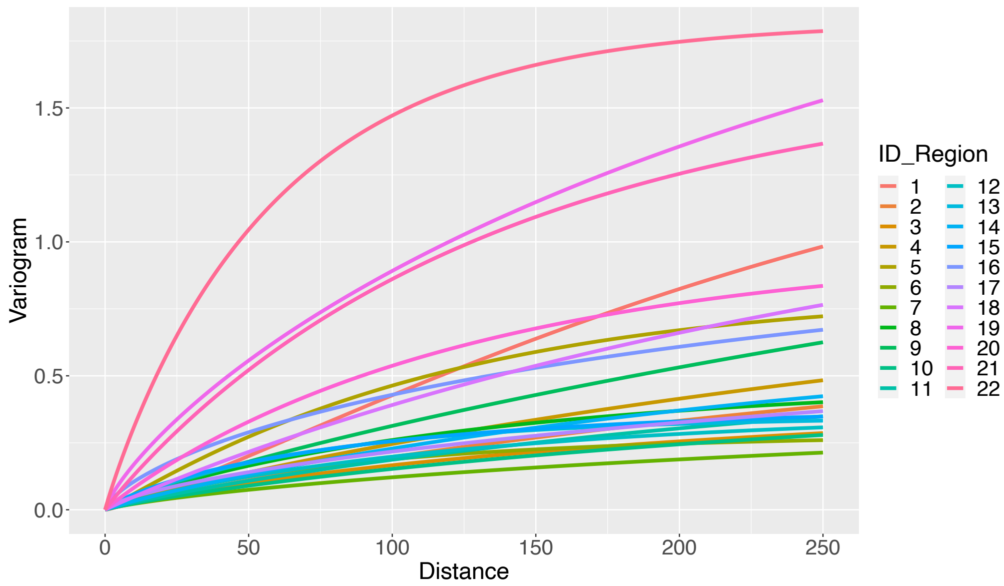
Based on the **stable covariance function**

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

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

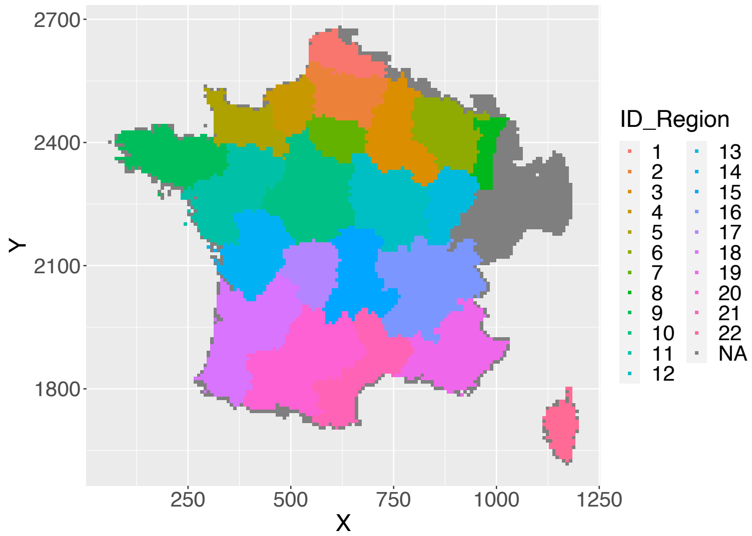
and maximum likelihood estimation using observations of $\log \boldsymbol{V}$

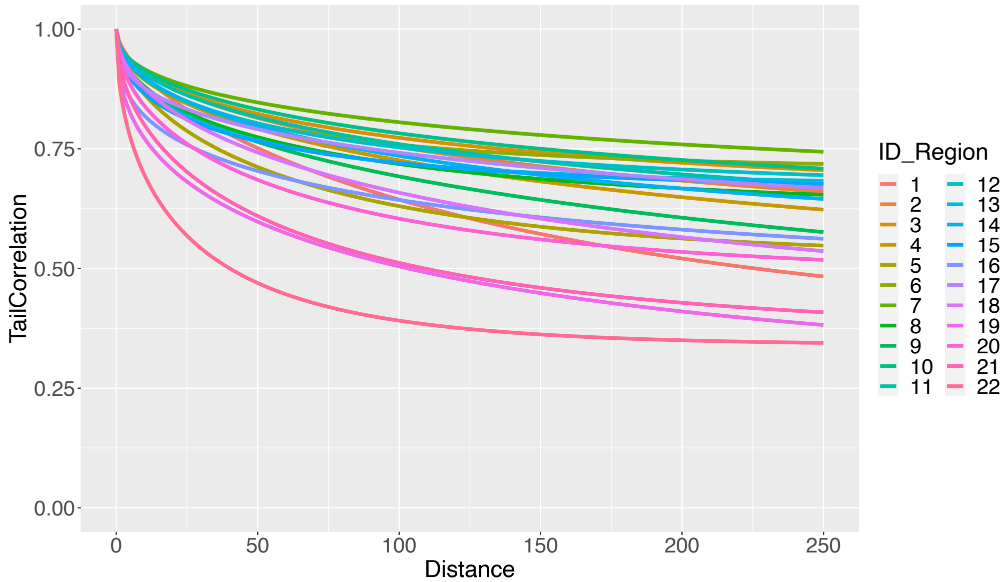




Results: Estimated tail correlations

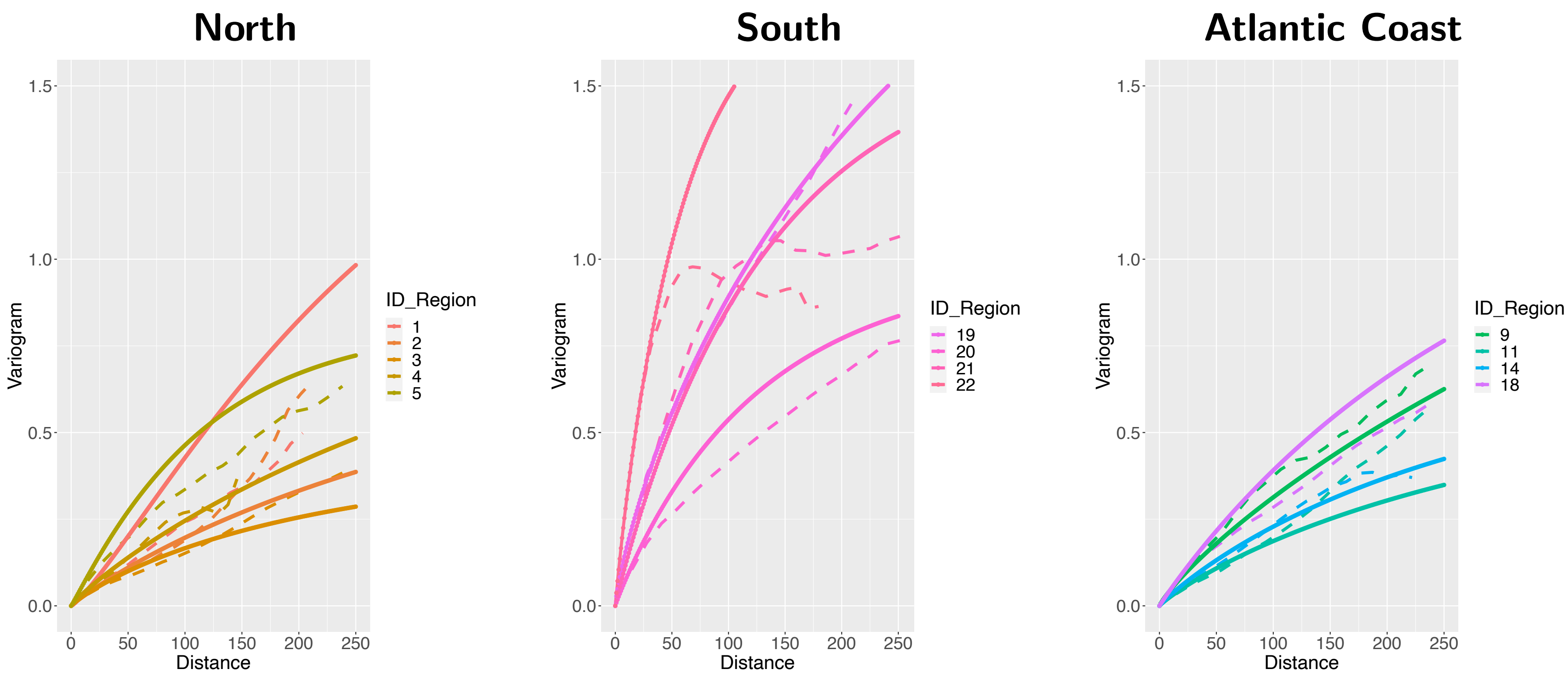
$$\chi(s, s+\Delta s) = \lim_{u \to \infty} \Pr(X^P(s+\Delta s) > u \mid X^P(s) > u) = 2\left(1 - \Phi\left(\sqrt{(\gamma(s, s+\Delta s)}\right)\right)$$





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** $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ **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`