1 Introduction

2 Univariate Extreme-Value Theory
   Maxima
   Threshold exceedances
   Point processes

3 Representations of dependent extremes using maxima and point processes
   Introduction to dependent extremes
   Componentwise maxima
   Point processes
   Spectral construction of max-stable processes

4 Representations of dependent extremes using threshold exceedances
   Extremal dependence summaries based on threshold exceedances
   Multivariate and functional threshold exceedances
   Application example: spatial temperature extremes in France

5 Perspectives
1 Introduction

2 Univariate Extreme-Value Theory
   Maxima
   Threshold exceedances
   Point processes

3 Representations of dependent extremes using maxima and point processes
   Introduction to dependent extremes
   Componentwise maxima
   Point processes
   Spectral construction of max-stable processes

4 Representations of dependent extremes using threshold exceedances
   Extremal dependence summaries based on threshold exceedances
   Multivariate and functional threshold exceedances
   Application example: spatial temperature extremes in France

5 Perspectives
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 \(\chi\) symmetric with respect to \(X_1\) and \(X_2\) and is known as \(\chi\)-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

\[ \chi = \lim_{z \to \infty} \Pr(X_2^* > z \mid X_1^* > z) = \lim_{z \to \infty} \frac{\Pr(X_1^* > z, X_2^* > z)}{\Pr(X_1^* > z)} \] (\#)

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

\[ F^{(X_1^*, X_2^*)}(nz, nz)^n \to G^*(z, z), \quad z > 0, \]

is equivalent to

\[ 1 - F^{(X_1^*, X_2^*)}(nz, nz) \approx -\log G^*(nz, nz), \quad \text{for large } n. \]

By using

\[ \Pr(X_1^* > z, X_2^* > z) = (1 - F_{X_1^*}(z)) + (1 - F_{X_2^*}(z)) - (1 - F^{(X_1^*, X_2^*)}(z, z)), \]

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

\[ \chi = 2 - V^*(1, 1). \]

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

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

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

**Tail dependence coefficient (minima)**

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

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

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

- Summary measures for co-occurrences of threshold exceedances
- Focus on minima and maxima of the components of $X^*$

Next...

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

2 Univariate Extreme-Value Theory
   - Maxima
   - Threshold exceedances
   - Point processes

3 Representations of dependent extremes using maxima and point processes
   - Introduction to dependent extremes
   - Componentwise maxima
   - Point processes
   - Spectral construction of max-stable processes

4 Representations of dependent extremes using threshold exceedances
   - Extremal dependence summaries based on threshold exceedances
   - Multivariate and functional threshold exceedances
   - Application example: spatial temperature extremes in France

5 Perspectives
Multivariate and functional threshold exceedances

Consider \( x \in \mathbb{R}^D \) for a compact domain \( D \subseteq \mathbb{R}^k \) with \(|D| > 1\).

**Note:** for a vector \( x = (x_1, \ldots, x_d) \), we can set \( D = \{1, \ldots, d\} \).

No unique definition of threshold exceedances \( \Rightarrow \) Use a risk functional \( r \)

**Extreme event occurs if** \( r(x) > u \) **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

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

and we apply \( r \) on the simple scale.

We further assume continuous realizations: \( x \in \mathcal{C}(D) \).

There is also notation \( \ell \) (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 \( x_1 = \ldots = x_d \) ⇒ 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 \( X_j^* \) satisfying

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

Two common choices

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

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

\[
\Pr(\bar{X}_j > X_j \mid X_j) \approx \frac{1}{X_j^*} \quad \text{for relatively large } X_j
\]

(Note: If \( \Pr(A) = 1/T \), then the event \( A \) has a return period of \( T \) time units)
Limits conditional to risk exceedances $r(X) > u$

**$r$-Pareto limit processes (Dombry & Ribatet 2015)**

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

- If we have the following (weak) convergence in $C(D)$,

$$
\frac{X^*}{u} \mid (r(X^*) > u) \rightarrow Y_r, \quad u \rightarrow \infty,
$$

then $Y_r$ is an $r$-Pareto process, satisfying Peaks-Over-Threshold stability:

$$
\frac{Y_r}{u} \mid (r(Y_r) > u) \overset{d}{=} Y_r, \quad \text{for any } u > 1.
$$

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

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

$\Rightarrow$ Above high thresholds $u$, scale $r(X^*)$ and profile $X^*/r(X^*)$ become independent!
Link to other limits

- Trinity of limits:
  
  Convergence of componentwise maxima
  \[ \iff \]
  Point-process convergence
  \[ \iff \]
  \( r \)-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 \( Y_r \) is
  \[
  Y_r \sim \frac{\Lambda^*(\cdot \cap A_r)}{\Lambda^*(A_r)} \quad \text{with} \quad A_r = \{ y \in C(D) \mid r(y) \geq 1 \}
  \]

- Consider the simple point-process limit \( \{ P_i^*, \; i \in \mathbb{N} \} \)
  \( \Rightarrow \) Construction of \( r \)-Pareto processes \( \hat{=} \) Extraction of \( r \)-exceedances:
  \[
  P_i^* \mid (r(P_i^*) > 1) \quad \overset{d}{=} \quad Y_r
  \]
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 $\log((\cdot)^*)$-scale (Gumbel scale)

$r = \text{Value at fixed location}$  
$r = \text{Geometric Average}$  
$r = \text{Arithmetic Average}$
Example: Geometric average risk for Brown–Resnick models

The popular Huesler–Reiss and Brown–Resnick models have log-Gaussian profile processes $\mathbf{V}$ for $r$ chosen as the geometric average. This is very convenient for statistical methods!

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

Log-Gaussian profile processes for $r = \text{Geometric average}$

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

$$\log \mathbf{V}(s) \overset{d}{=} \tilde{\mathbf{W}}(s) - \overline{\mathbf{W}} - \text{const}(s; \Gamma)$$

with

- a centered Gaussian process $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}(s), s \in D\}$ and its spatial average $\overline{\mathbf{W}}$,

- a constant $\text{const}(s; \Gamma)$, explicit in terms of the semivariogram matrix

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

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

(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

\[ \log V(s) = \tilde{W}(s) - \bar{W} - \text{const}(s; \Gamma) \]

Same semivariograms of the log-profile \( \log V \) and the original Gaussian process \( \tilde{W} \)!

\[ \gamma_{\log V}(s_1, s_2) = \frac{1}{2} \gamma [\log V(s_2) - \log V(s_1)] = \frac{1}{2} \gamma [\tilde{W}(s_2) - \tilde{W}(s_1)] = \gamma_{\tilde{W}}(s_1, s_2) \]

\[ \Rightarrow \text{Classical variogram analysis becomes possible for } \log V! \]
Introduction

Univariate Extreme-Value Theory
  Maxima
  Threshold exceedances
  Point processes

Representations of dependent extremes using maxima and point processes
  Introduction to dependent extremes
  Componentwise maxima
  Point processes
  Spectral construction of max-stable processes

Representations of dependent extremes using threshold exceedances
  Extremal dependence summaries based on threshold exceedances
  Multivariate and functional threshold exceedances
  Application example: spatial temperature extremes in France

Perspectives
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 $r(X_t^*)$
- **Maximum likelihood** using log $V$ with a stable covariance function in $\tilde{W}$
Study domain: 22 French administrative regions

<table>
<thead>
<tr>
<th>ID_Region</th>
<th>Region Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>NA</td>
</tr>
</tbody>
</table>
Results: Marginal GPD parameters

Scale $\sigma_{GP}(s)$

Shape $\xi(s)$
Results: Estimated extremal variograms

Based on the stable covariance function

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

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

and maximum likelihood estimation using observations of log $V$
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 $\text{const}(s; \Gamma)$
- Generally satisfactory fit
- During extreme heat days, stronger spatial variability in the Southern regions
Introduction

Univariate Extreme-Value Theory
  Maxima
  Threshold exceedances
  Point processes

Representations of dependent extremes using maxima and point processes
  Introduction to dependent extremes
  Componentwise maxima
  Point processes
  Spectral construction of max-stable processes

Representations of dependent extremes using threshold exceedances
  Extremal dependence summaries based on threshold exceedances
  Multivariate and functional threshold exceedances
  Application example: spatial temperature extremes in France

Perspectives
Statistical aspects of extreme-value analysis

In practice, we typically have observations of a sample $X_1, \ldots, X_n$ 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 \( X \) and not to standardized \( X^* \)

- 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:

Statistical modeling:
  • Coles (2001). An introduction to statistical modeling of extreme values.

Mix of both:

A review of available software (R-based):