Inférence statistique des valeurs extrêmes spatiales

Modèle asymptotique classique ou modèle tendant vers l'indépendance asymptotique?

Thomas OPITZ

BioSP, INRA Avignon

Recontres Statistiques Avignon-Marseille
24/06/2016

Why do we need statistics of spatial extremes ?

- extreme events
  - **rare by definition, but often high impact**
  - physical dynamics and stochastic behavior often different under stress
  - spatial processes : take into account **spatial dependence**

  $\leadsto$ use models with a sound foundation in **extreme value theory**, avoid plain Gaussian models appropriate only for central tendencies
- emphasis is on the tail of the distribution $\leadsto$ **extrapolation**
- **objective : model**, **quantify, predict** extreme risks

Goals of this talk

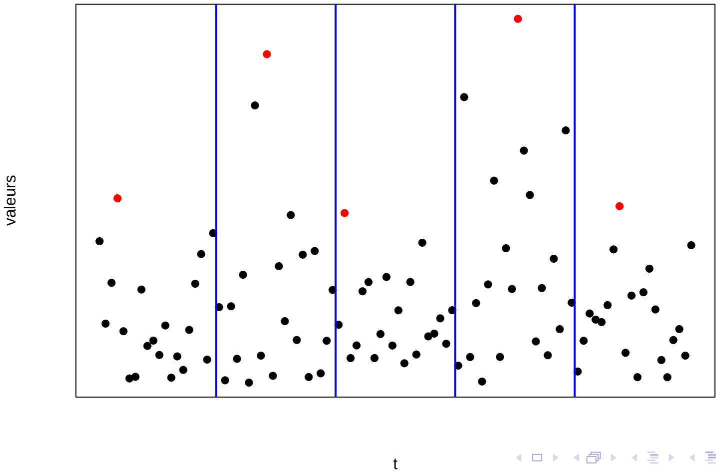
1. present asymptotic theory for **spatial extremes** and the resulting **asymptotic models**, well suited to capture **asymptotic dependence**

2. show that many real data have a tendency towards **asymptotic independence**

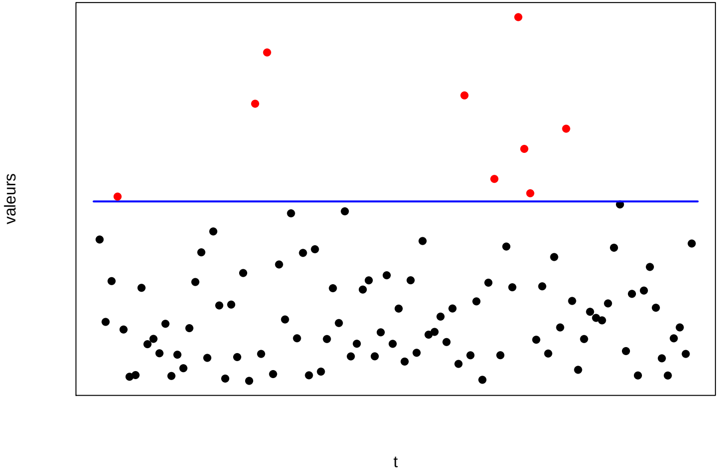
3. review some models suitable for **asymptotically independent data**

What observations are "extreme" ? – Maxima

We obtain a sample of maxima by considering block maxima over blocks of size n.



What observations are "extreme" ? – Threshold exceedances



Max-stable limit processes for pointwise maxima

Let be given i.i.d. copies $\mathbf{X_i}$ of a stochastic process $\mathbf{X} = \{X(s)\}$.

Maximum domain of attraction

If sequences $\sigma_n(s) > 0$, $\mu_n(s)$ exist such that

$$\left\{ \max_{i=1,\ldots,n} \sigma_n(s)^{-1} \left[X_i(s) - \mu_n(s) \right] \right\} \overset{fdd}{\to} \{Z(s)\}, \quad n \to \infty \tag{1}$$

with non-degenerate limit process $\boldsymbol{Z} = \{Z(s)\}$,
then $\boldsymbol{Z}$ is a max-stable process.
We say that $\boldsymbol{X}$ is in the max-domain of attraction of $\boldsymbol{Z}$.

Max-stability of $\boldsymbol{Z}$ means that (1) holds exactly for $\boldsymbol{Z}$.

Univariate limit distributions

The univariate limit distributions are **extreme-value** distributions $Z(s) \sim \mathrm{EV}_{\xi(s)}$.

We can choose $\sigma_n(s), \mu_n(s)$ such that

$$\mathrm{EV}_{\xi(s)}(z) = \exp\left(-T_{\xi(s)}(z)\right), \quad \mathrm{EV}_{\xi(s)}(z) = \exp\left(-T_{\xi(s)}(z)\right), \quad T_{\xi(s)}(z) = (1+\xi(s)z)_+^{-1/\xi(s)}$$

with shape parameter $\xi(s) \in \mathbb{R}$ that determines the tail decay rate

Tail dependence summaries

Multivariate extremal coefficients

Th extremal coefficient $\theta \in [1, D]$ of max-stable random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_D)$ with $Z_j \overset{d}{=} Z_1$ is defined through

$$\mathrm{P}(Z_1 \leq x, \ldots, Z_D \leq x) = \mathrm{P}(Z_1 \leq x)^{\theta}. \quad (2)$$

Interpretation : D/θ is average "cluster size" of extreme events.

For stochastic processes, we can consider the **extremal coefficient function** $\theta(s_1, s_2) \in [1, 2]$ of bivariate extremal coefficients.

The **tail correlation function** $\lambda(s_1, s_2) = 2 - \theta(s_1, s_2) \in [0, 1]$ is a particular type of correlation function.

Asymptotic dependence and asymptotic independence

Asymptotic independence of $X(s_1)$ and $X(s_2)$ corresponds to **independence of** $Z(s_1)$ **and** $Z(s_2)$, *i.e.*, to $\theta(s_1, s_2) = 2$ and $\lambda(s_1, s_2) = 0$.

Otherwise ($\lambda > 0$), we observe **asymptotic dependence**.

More generally (even if a max-stable limit does not exist),

for standardized variables
$\boldsymbol{X}^\star = (X_1^\star, X_2^\star) = (1/(1 - F_{X(s_1)}(X(s_1))), 1/(1 - F_{X(s_2)}(X(s_2)))),$

we can define the tail correlation coefficient $\lambda(s_1, s_2)$ as

$$\lambda(s_1, s_2) = \lim_{x \to \infty} \mathrm{P}(X_1^\star > x \mid X_2^\star > x).$$

In spatial data, asymptotic independence means that the most extreme events become **more and more isolated in space**.

Note : Gaussian processes are **asymptotically independent** [Sibuya, 1960].

The spectral construction of max-stable processes

[de Haan, 1984]

A max-stable process $\boldsymbol{Z}^\star$ can be represented through a **spectral construction**

$$Z^\star(s) = \max_{i=1,2,\dots} \varepsilon_i(s)/U_i, \quad U_i \sim \mathrm{PPP}(\mathrm{d}u) \text{ on } [0,\infty),$$

- here : unit Fréchet marginal distribution, $\mathrm{pr}(Z^\star(s) \leq z) = \exp(-1/z)$, $z > 0$
- **profile processes** $\varepsilon_i(s)$, $i = 1, 2, \dots$ are i.i.d. with $\mathbb{E}\,\varepsilon_i(s)_+ = 1$

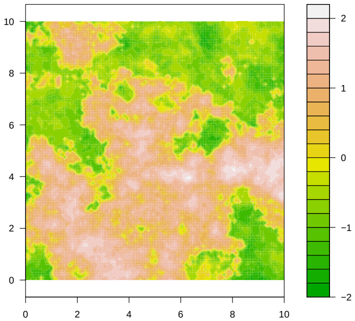
The spectral construction can be used to **construct spatial max-stable models** :

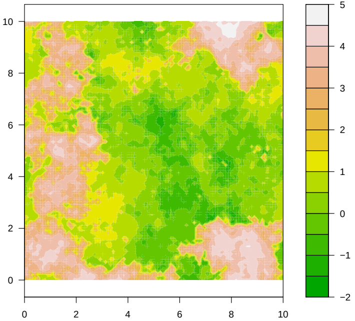
- centered Gaussian profile process : $\varepsilon_1(s) = W(s)_+^{\mathrm{df}}$, $\mathrm{df} > 0$
  $\rightsquigarrow$ **extremal-t process** $Z^\star(s)$ [Opitz, 2013, Thibaud and Opitz, 2016]
  $\mathrm{df} = 1$: **Schlather process** [Schlather, 2002]
- log-Gaussian profile process :
  $\varepsilon_1(s) = \exp(W(s) - \sigma^2(s)/2)$ where $\sigma^2(s) = \mathrm{Var}(W(s))$
  $\rightsquigarrow$ **Brown–Resnick type process** $Z^*(s)$ [Kabluchko et al., 2009]

Example 1 : Schlather process

[Schlather, 2002]

- extremal-t process with $\mathrm{df} = 1$
- simulation on $[0, 10] \times [0, 10]$
- exponential correlation function with range 3 in Gaussian profile process $W(s)_+^{\mathrm{df}}$
- plot of $\log(Z^\star(s))$

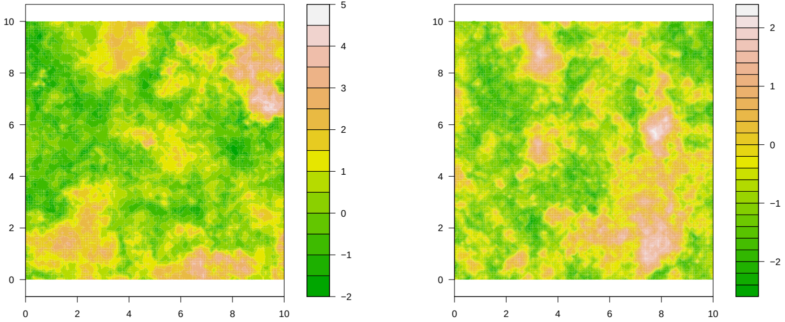




Example 2 : Extremal-t

[Opitz, 2013]

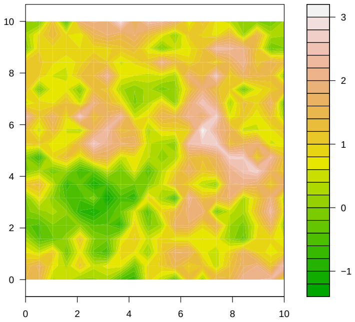
- extremal-t process with $\mathrm{df} = 4$
- simulation on $[0, 10] \times [0, 10]$
- exponential correlation function with range 3 in profile process $W(s)_+^{\mathrm{df}}$
- plot of $\log(Z^\star(s))$

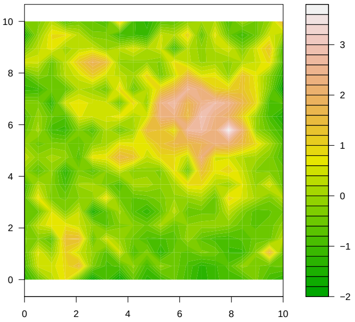


Example 3 : Brown–Resnick process

[Brown and Resnick, 1977, Kabluchko et al., 2009]

- simulation on $[0, 10] \times [0, 10]$
- profile process $\exp(W(s) - \sigma^2(s)/2)$ with fractional Brownian motion $W(s)$ (Hurst index 1/4)
- plot of $\log(Z^\star(s))$





Statistical inference

Idea : use max-stable limit distributions for observed maxima

- **cdf** of max-stable vector is

$$G(\boldsymbol{z}) = \exp\left(-V(T_{\xi_1}^{-1}(z_1), \ldots, T_{\xi_D}^{-1}(z_D)\right),$$

  where $V(\boldsymbol{z})$ with $V(t\boldsymbol{z}) = t^{-1}V(\boldsymbol{z})$ is the **exponent function**
- full likelihood inference impossible in high dimension since combinatorial explosion of terms when deriving cdf
- common likelihood-based inference through **pairwise likelihood approach**
- in **spatial modeling**, we need :
  - a marginal model for $\xi(s), \sigma(s), \mu(s)$
  - a dependence model (Schlather, Brown–Resnick, extremal-t, ...)
- **simulation** of max-stable processes is based on the spectral construction, conditional simulation is tricky but possible
- **extrapolation :**
  - if $\{Z(s)\} = \left\{\mu(s) + \sigma(s)T_{\xi(s)}^{-1}(Z^\star(s))\right\}$ models annual maxima, then $\left\{\mu(s) + \sigma(s)T_{\xi(s)}^{-1}(nZ^\star(s))\right\}$ models n-year maxima
  - often Monte–Carlo based calculation of complicated functionals $f(\boldsymbol{Z})$

From maxima to threshold exceedances

- **difficulties with maxima-based modeling**
  - pointwise maximum process $\boldsymbol{M}_n$ can be composed of components of different events
    $\rightsquigarrow$ not straightforward to derive results for individual extreme events
  - cannot capture non-i.i.d. behavior within blocks
- **equivalent limit relations exist for threshold exceedances**
  - in practice, can fix the **threshold value to balance bias and variance in estimation**
  - however, there is **no natural ordering relation for multivariate and spatial data**, making the definition of "extreme events" ambiguous

Genereralized Pareto limits for threshold exceedances

- the univariate max-domain of attraction condition holds **iff** threshold exceedances converge to a **generalized Pareto distribution** :

$$\sigma_u(s)^{-1}(X(s) - \mu_u(s)) \mid (X(s) > u) \to Y(s) \sim \mathrm{GP}_{\xi(s)}, \quad u \uparrow F_{X(s)}^{-1}(1^-)$$

  where $\mathrm{GP}_{\xi(s)}(y) = 1 - T_{\xi(s)}(y)/T_{\xi(s)}(u)$ with $T_{\xi(s)}(y) = (1+\xi(s)y)_+^{-1/\xi(s)}$

- for characterizing dependence and for estimating models, it is useful to **treat separately marginal and dependence behavior**
  - standardize margins : $X^\star(s) = 1/(1 - F_{X(s)}(X(s)))$
  - $X^\star(s)$ is standard Pareto distributed if $X(s)$ has continuous distribution
- define **extreme events** as **exceedances of a homogeneous risk functional** ℓ
  - $\boldsymbol{X}^\star$ is an extreme event if $\ell(\boldsymbol{X}^\star) > u$ with treshold $u > 0$
  - need **homogeneity** ($\ell(t\boldsymbol{x}) = t\ell(\boldsymbol{x})$) and **positivity** ($\ell(\boldsymbol{x}) > 0$ if $\boldsymbol{x} > \boldsymbol{0}$) for convergence
  - for instance, $\ell(\boldsymbol{x})$ given as $x(s)$, $\max_{s\in K} x(s)$, $\min_{s\in K} x(s)$ or $\mathrm{mean}_{s\in K} x(s_i)$

Limit processes for ℓ-exceedances

[Ferreira and De Haan, 2014, Dombry and Ribatet, 2015, Thibaud and Opitz, 2016]
If $\boldsymbol{X}$ is a the max-domain of attraction, we get

$$u^{-1}\boldsymbol{X}^{\star} \mid (\ell(\boldsymbol{X}^{\star}) > u) \to \boldsymbol{Y}^{\star} \sim \mathrm{GP}_{\ell}, \quad u \to \infty, \tag{3}$$

with an ℓ-Pareto process $\boldsymbol{Y}^{\star}$.

- if (3) holds for $\ell(\boldsymbol{x}) = \max_s x(s)$, the max-domain of attraction condition is satisfied.
- if $\ell(\boldsymbol{x}) = \max(x(s_1), \ldots, x(s_D))$, then $\mathrm{pr}(Y^{\star}(\boldsymbol{s}) \leq \boldsymbol{y}) = 1 - V(\boldsymbol{y})/V(\boldsymbol{u})$ for $\boldsymbol{y} > \boldsymbol{u}$
- we can retransform to the original scale of data,

  $$Y(s) = \mu(s) + \sigma(s)(Y^{\star}(s)^{\xi} - 1)/\xi,$$

  yielding a **generalized ℓ-Pareto process**

Censored likelihood inference for threshold exceedances

Assume data $\boldsymbol{x}_i = x_i(s_j)$, $i = 1, \ldots, n$ have been observed on sites $s_1, \ldots, s_D$.

Here we consider $\ell(\boldsymbol{x}) = \max(x^\star(s_1), \ldots, x^\star(s_D))$.

- using the ℓ-Pareto model amounts to $\mathrm{pr}(\boldsymbol{X}^* \not\leq \boldsymbol{x}^\star) = V(\boldsymbol{x}^\star)$ for $\boldsymbol{x}^\star > (u, \ldots, u)$
- non-extreme components $x_i^\star(s_j) < u$ are **censored**
- likelihood contribution of $\boldsymbol{x}_i^\star$:
  - when none of the components exceeds its threshold : $1 - V(u, \ldots, u)$
  - when w.l.o.g. components $x_i^\star(s_1)$, ..., $x_i^\star(s_{j_0})$ are exceedances :

$$-\frac{\partial^{j_0}}{\partial x_1 \times \ldots \times \partial x_{j_0}} V(x_i^\star(s_1), \ldots, x_i^\star(s_{j_0}), u, \ldots, u)$$

Modeling precipitation extremes around Zurich

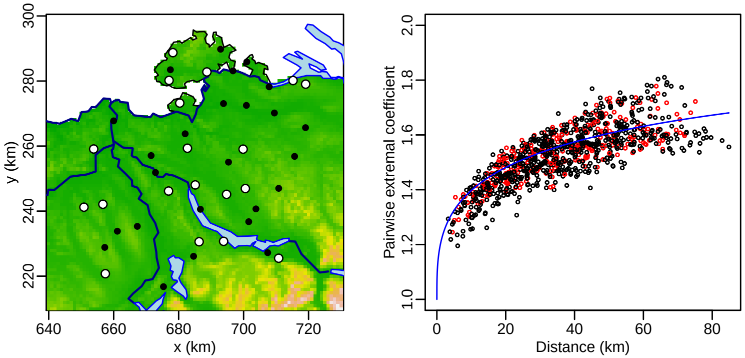
[Thibaud and Opitz, 2016]

- daily summer precipitation data (1962-2012) for 44 sites (25 for estimation, 19 for validation)
- ℓ-**Pareto model** with $\ell(\boldsymbol{x}) = \max(x(s_1), \ldots, x(s_D))$
  - marginal model :
    $\sigma(s) = \beta_{\sigma,0} + \beta_{\sigma,1}\mathtt{LAT} + \beta_{\sigma,2}\mathtt{LON} + \beta_{\sigma,3}\mathtt{ALT}$
    $\mu(s) = \beta_{\mu,0} + \beta_{\mu,1}\mathtt{LAT} + \beta_{\mu,2}\mathtt{LON} + \beta_{\mu,3}\mathtt{ALT}$
    $\xi(s) \equiv \xi_0$
  - extremal-t dependence with stable correlation function
- estimation :
  - threshold $u = 20$ (marginal 95%-quantile)
  - **two-step estimation :**
    independence likelihood for marginal parameters,
    full likelihood on standardized data for dependence parameters
  - $\hat{\mathrm{df}} = 6$
  - $\mathrm{Cor}(s_1, s_2) = \exp\left[-(\|h\|/\hat{\beta})^{\hat{\kappa}}\right]$ with $\hat{\beta} = 483(35)$ km, $\hat{\kappa} = 0.64(0.01)$

Goodness-of-fit of the extremal coefficient function

- good coverage of confidence intervals for conditional distributions on validation sites
- AIC-based model selection : extremal-t outperforms Brown–Resnick

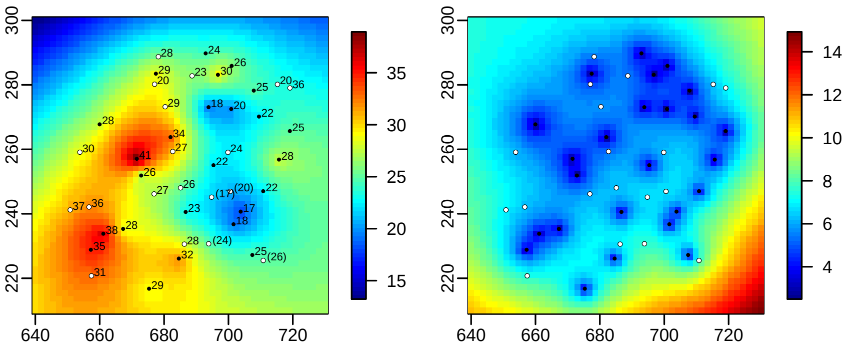
Extremal coefficient function : empirical vs. fitted



Spatial prediction with the fitted model (1 June 1962)

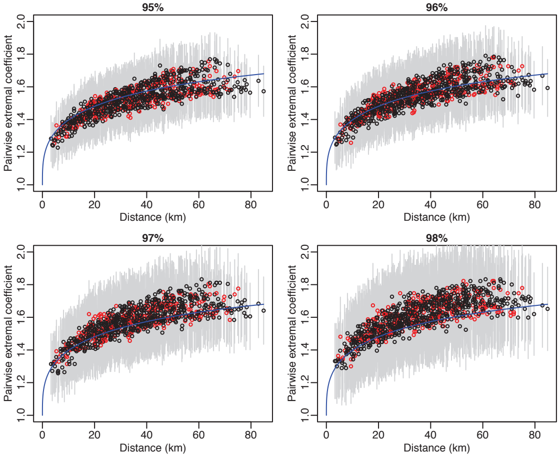
- condition on values at observed sites
- conditional process has (transformed) finite-dimensional *t*-distributions

Left : conditional mean ; right : standard deviation



Zurich data – have we reached asymptotics ?

Fitted extremal coefficient function **vs. Empirical** extremal coefficients (empirical coefficients calculated for different thresholds)

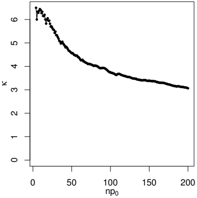


When going far into the tail of data, empirical coefficients tend to increase towards 2, the value for **asymptotic independence**.

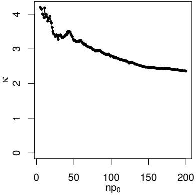
Other examples : asymptotic independence ?

Empirical estimates of extremal coefficients θ with respect to number of exceedances

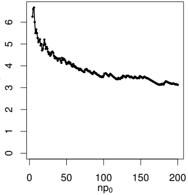
precipitation (Cévennes, $D = 21$)



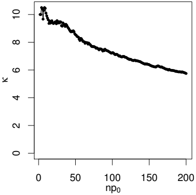
precipitation (Drenthe, $D = 15$)



wind gusts (Netherlands, $D = 24$)



temperature (France, $D = 33$)



Again, we observe a strong increase in estimates when going farther into the tail.

Can we handle asymptotic independence in spatial data ?

- **max-stable models** only allow **either asymptotic dependence or classical independence**, but nothing in between

- empirical evidence of many spatial data sets suggests that an asymptotic independent model would provide a better fit

- asymptotically dependent data : $\mathrm{pr}(X_1^\star > x, X_2^\star > x) \sim \lambda x^{-1}$ with $0 < \lambda \leq 1$

- **faster joint tail decay in asymptotically independent data**, *e.g.*

$$\mathrm{pr}(X_1^\star > x, X_2^\star > x) \sim h(x)x^{-\eta} \quad (4)$$

with $\eta > 1$ and slowly varying h, $h(tx)/h(t) \to 1$ $(t \to \infty)$

Can we have flexible spatial models for (4) **?**

Inverted max-stable processes

[Wadsworth and Tawn, 2012]

Given a max-stable process $Z^\star(s)$, the corresponding **inverted max-stable process**

$$X(s) = 1/Z^\star(s)$$

has exponential marginal distributions. Then $X^\star(s) = \exp(X(s))$.

Its **joint tail decay is**

$$\mathrm{pr}(X_1^\star > x, \ldots, X_D^\star > x) = x^{-D/\theta},$$

where $1 < \theta \leq D$ is the extremal coefficient of $(Z(s_1), \ldots, Z(s_D))$.

Since $D/\theta > 1$, the dependence of the original max-stable process determines the joint tail decay rate, leading to a **flexible framework for modeling asymptotic independence**.

We can estimate this tail dependence model by using threshold exceedances in data, *i.e.*, by censoring non-exceeding data values in the pairwise likelihood.

Gaussian scale mixture processes

[Opitz, 2016] + work in progress (joint with R. Huser, E. Thibaud)

Gaussian processes $W(s)$ are well-studied and well-tractable for inference, but they lack flexibility in the tail. We get more flexible tail behavior by **embedding a random variable R^2 with $R \geq 0$ for the variance** :

$$X(s) = RW(s)$$

- if R has power law tail, then $\boldsymbol{X}$ is asymptotically dependent with extremal-t limit
- if R has Weibull-type tail, $\mathrm{pr}(R > r) \sim r^{-c_1} \exp(-c_2 r^{-\alpha})$, then $\boldsymbol{X}$ is asymptotically independent with

  $$\mathrm{pr}(X_1^\star > x, X_2^\star > x) \sim h(x) x^{-[2/(1-\rho)]^{\alpha/(\alpha+2)}},$$

  yielding **flexible joint tail behavior** strongly determined by α
- can interpolate smoothly between Gaussian dependence and asymptotic dependence
- [Opitz, 2016] : **Laplace model** with $R^2 \sim \mathrm{Exp}$ has nice properties

Conclusion

- **Gaussian models are not flexible enough** to capture extreme value behavior
- spatial extreme value theory provides exploratory tools and asymptotic models
  - **max-stable models** are well studied and understood
  - **generalized Pareto models** are better suited for threshold exceedances
  - classical full likelihood inference is always tricky but possible in some cases, although computationally heavy
- **classical asymptotic models for spatial extremes** are well adapted to capture asymptotic dependence
- when convergence is not observed in data, we may need **subasymptotic models**, in particular for **asymptotic independence** to **avoid overestimation of extreme joint risks**

☺ – Merci – Danke – Thank you – ☺

Brown, B. M. and Resnick, S. I. (1977).
Extreme values of independent stochastic processes.
J. Appl. Probab., pages 732–739.

Coles, S. (2001).
An introduction to statistical modeling of extreme values.
Springer.

de Haan, L. (1984).
A spectral representation for max-stable processes.
Ann. Probab., 12(4) :1194–1204.

Dey, D. K. and Yan, J. (2015).
Extreme Value Modeling and Risk Analysis : Methods and Applications.
Chapman and Hall/CRC.

Dombry, C. and Ribatet, M. (2015).
Functional regular variations, Pareto processes and peaks over threshold.
Statistics and Its Interface, 8(1) :9–17.

Ferreira, A. and De Haan, L. (2014).
The generalized Pareto process ; with a view towards application and simulation.
Bernoulli, 20(4) :1717–1737.

Kabluchko, Z., Schlather, M., and De Haan, L. (2009).
Stationary max-stable fields associated to negative definite functions.
Ann. Probab., 37(5) :2042–2065.

Opitz, T. (2013).
Extremal t processes : Elliptical domain of attraction and a spectral representation.
Journal of Multivariate Analysis, 122 :409–413.

Opitz, T. (2016).
Modeling asymptotically independent spatial extremes based on Laplace random fields.
Spatial Statistics, 16 :1–18.

Schlather, M. (2002).
Models for stationary max-stable random fields.
Extremes, 5(1) :33–44.

Sibuya, M. (1960).
Bivariate extreme statistics, I.
Ann. Inst. Statist. Math., 11 :195–210.

Thibaud, E. and Opitz, T. (2016).
Efficient inference and simulation for elliptical Pareto processes.
Biometrika.

Wadsworth, J. L. and Tawn, J. A. (2012).
Dependence modelling for spatial extremes.
Biometrika, 99(2) :253–272.