Extremal Dependence

R. de Fondeville

Federal Office of Statistics
Athens week, Paris

March 15th, 2022

Motivation

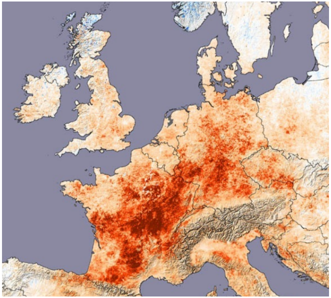
> "Human influence has likely increased the chance of compound extreme events since the 1950s."
>
> — IPCC 6th Assessment Report on Climate Change 2021

Compound extremes

Definition

Compound extremes, referred to as *simultaneous, concurrent, or coincident* extremes, may lead to larger *impacts* to human society and the environment than individual extremes alone.

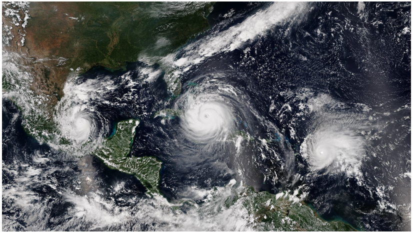
Example of compound extremes



2003 Heat wave in the South of France

⇒ **Impact**: the larger the number of cities suffered from the heatwaves the higher the death toll is likely to be.

Example of compound extremes



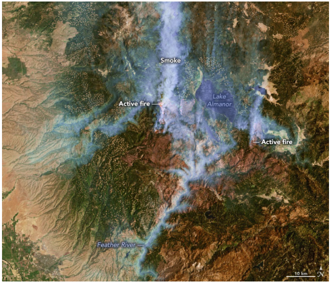
Cyclones Irma, Jose and Katia.

$\Rightarrow$ **Impact**: the larger the number of infrastructures suffers from severe winds, the higher the potential of damages (and induced insurance losses).

Example of compound extremes

$\Rightarrow$ **Impact**: directional alignment of wind speed and waves heights are likely to cause platform destruction / sink tankers yielding high risk of pollution ("black tides").

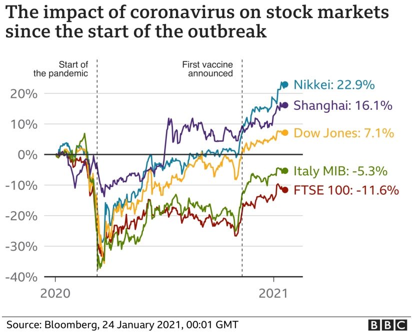
Example of compound extremes



Wildfire "Dixie" in July 2021 in California.

$\Rightarrow$ **Impact**: combination of severe drought and high temperatures increase wildfire risk causing forests and properties destructions.

Example of compound extremes



Global Financial crisis.

$\Rightarrow$ **Impact**: increased poverty and unemployment rates.

Notion of Dependence

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Statistical dependence

Dependence refers to any statistical association between two variables. Most of the time, it is replaced by the notion correlation, the degree to which a pair of variables are *linearly* related.

Notion of Spatial Dependence

Tobler's First Law of Geography

Everything is related to everything else, but near things are more related than distant things.

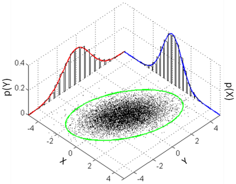
Covariance

- Classical definition of pairwise dependence:

$$\mathrm{Cov}(X, Y) = \mathrm{E}\{(X - \mu_X)(Y - \mu_Y)\} = \mathrm{E}(XY) - \mu_X \mu_Y,$$

where X and Y are two random variables with finite mean μ_X and μ_Y and satisfying

$$\mathrm{E}(X^2) < \infty, \quad \mathrm{E}(Y^2) < \infty.$$



- Covariance is very popular as it simple and fully specifies multivariate normal / Gaussian distributions.

Limits of the covariance measure

- Covariance quantifies association for bulk of the distribution.
- For random variables with sufficiently heavy tails ($\xi > 1/2$) covariance does not exist!
- Outside of the family of Gaussian distribution, pairwise summaries does not fully characterize dependence.

$\Rightarrow$ We need alternative measures of dependence tailored for extremes events.

Recap on univariate EVT: Problem definition

Using observations $X_1, \ldots, X_n \sim F$, for some set $A \subset \mathbb{R}$, we want to estimate

$$\Pr(X \in A).$$

- Use:
  - empirical estimate if A is in the body of the distribution;
  - parametric model if A contains no, or limited, data.
- If A is far from data, bias due to the fit being dominated by the body of the data.
- To compensate for the lack of data in the tails, we need a asymptotically motivated parametric model.

Recap on univariate EVT: Block maxima

For $X_1, \ldots, X_n \sim F$, the distribution of block-maxima

$$M_n = \max(X_1, \ldots, X_n),$$

converges as $n \to \infty$ to

$$\Pr\left(\frac{M_n - b_n}{a_n} < x\right) = \{F(a_n x + b_n)\}^n \to G(x),$$

where G is non-degenerate.

Recap on univariate EVT: Block maxima

Then, the distribution of G is

$$G(x) = GEV(\mu, \sigma, \xi)(x) = \exp\left[-\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-1/\xi}\right\},$$

with $x_+ = \max(x, 0)$ and

- Location parameter μ;
- Scale parameter $\sigma < 0$;
- Shape parameter / Tail index ξ.

For large n, we then assume that $M_n \sim GEV(\mu, \sigma, \xi)$.

Recap on univariate EVT: Peaks-over-threshold

From

$$\{F(a_n x + b_n)\}^n \to G(x),$$

for large n, logarithm and Taylor expansion yields

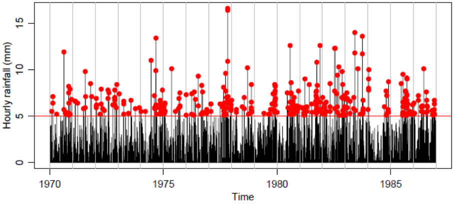
$$n\{1 - F(a_n x + b_n)\} \to -\log G(x).$$

For $x > 0$, as $\to \infty$,

$$\begin{aligned}
\Pr(X > a_n x + b_n | X > b_n) &= \frac{1 - F(a_n x + b_n)}{1 - F(b_n)} \\
&= \frac{n\{1 - F(a_n x + b_n)\}}{n\{1 - F(a_n u + b_n)\}} \\
&\to \frac{\log G(x)}{\log G(u)} \\
&\to \left\{1 + \xi \frac{x}{\sigma_u}\right\}_+^{-1/\xi},
\end{aligned}$$

where $\sigma_u = \sigma + \xi(u - \mu)$.

Recap on univariate EVT: Peaks-over-threshold



For large n, the sequence b_n tends to the upper end point of F and motivate the approximation for large threshold $u > 0$

$$\Pr(X > x + u | X > u) \approx GPD(\sigma_u, \xi) = \left(1 + \xi \frac{x}{\sigma_u}\right)_+,$$

where $\sigma_u = \sigma + \xi(u - \mu)$.

Multivariate / Bivariate Definition of Extremes

- We now suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with multivariate distribution function F (density f) and we note

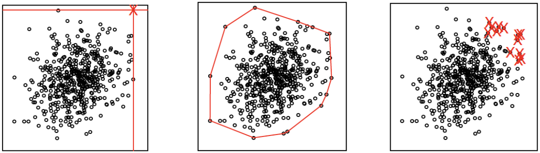
$$X_i = (X_{i,1}, \ldots, X_{i,D}) \in \mathbb{R}^D.$$

- There is no natural "ordering" or unique definition of extremes in a multivariate setting.

When / How can we write $X_1 \geq X_2$ or $X_1 \leq X_2$?

Multivariate / Bivariate Definition of Extremes

- Barnett (1976) proposes several possible orderings:
  - Componentwise maxima:
  $$M_n = (M_{X_1,n}, \ldots, M_{X_D,n});$$
  - Convex hull: $f(X) = c$;
  - Structure variable: extreme of univariate variable $r(X)$ with r chosen depending of the problem, e.g., $r(X) = \sum_{i=1}^{D} X_i$.



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Structure variables: notations

- Using observation $X_1, \ldots, X_n$, we want to estimate

$$\Pr(X \in A)$$

  for some set $A \subset \mathbb{R}^D$.

- For some function $r : \mathbb{R}^D \to \mathbb{R}$ and threshold $u > 0$, we analyze

$$\Pr(X \in A) = \Pr\{r(X) > u\}.$$

- Coles and Tawn (1994) showed that this can be tackeld as a univariate problem.

Examples of Structure Variables

1. Force on offshore structure :

$$r(X) = a_1 X_1^2 + a_2 X_2^2,$$

   where X_1 and X_2 are the wave and wind forces.

2. Water volume over a region :

$$r(X) = \sum_{i=1}^{D} a_i X_i,$$

   where X_i is the rainfall measured at station i and a_i the estimated region covered by the station.

3. Portfolio risk :

$$r(X) = \sum_{i=1}^{D} a_i X_i,$$

   where X_i is return loss of asset i. The aim is here to find a_i such that the risk of losses is minimized.

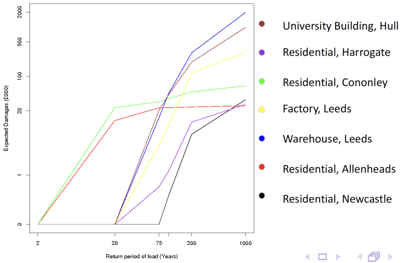
Examples of Structure Variables

4. Financial loss from spatial river flows:

$$r(X) = \sum_{i=1}^{D} r_i(X_i),$$

where X_i is the river flow at gauging station i and r_i is the (highly non-linear) property loss function.

Individual property damage data



Univariate approach for Structure Variables

- Observations $r(X_1), \ldots, r(X_n)$ are univariate, so we apply GEV/GPD approximations.
- Stenghts:
  - very simple and flexible;
  - no need to model dependence.
- Limits:
  - cannot handle missing data;
  - the estimated model is strongly influenced by the choice of r;
  - We have no information on the structure of the underlying event.

Multivariate Approach for Structure Variables

- Using observation $X_1, \ldots, X_n$, we estimate their joint distribution F

$$\Pr(X \in A) = \int_A dF(x)$$

for some set $A \subset \mathbb{R}^D$.

- For some function $r : \mathbb{R}^D \to\to$ and threshold $u > 0$, this is equivalent to

$$\Pr\{r(X) > u\} = \int_{\mathbb{R}^D} 1\{r(x) > u\} dF(x).$$

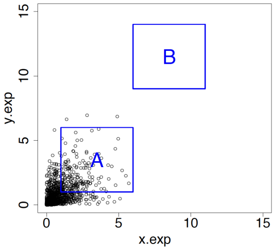
- The corresponding probability might not be available in closed form (explicit formula) but can be evaluated numerically.

Multivariate Extremes: Problem Definition

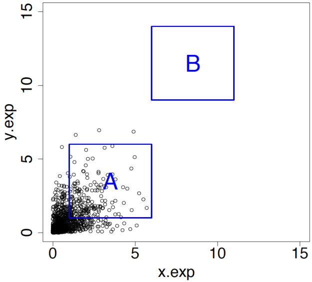
- Aim: estimate $\Pr(X \in B)$ when there is little to no data in B.

- Tool: we use asymptotically motivated models to extrapolate into the multivariate tail.

- The methodology relies on linking $\Pr(X \in B)$ to $\Pr(X \in A)$ by modelling the tail decay.



Multivariate Extremes: Difficulties



- No natural direction of extrapolation.
- Each "direction" can be considered with different underlying hypothesis.
- Multiple models are available with different hypothesis.
- Diagnostics are required to asses the quality of the proposed models.

Copulas to model dependence structure

- Let $X \sim F$, the copula C is given by

$$F(x) = C\{F_1(x_1), \ldots, F_D(x_D)\},$$

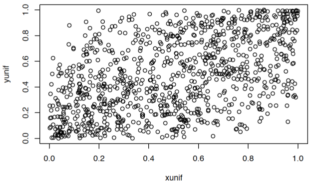
where F_i is the marginal distribution of compoent X_i of X.

- The copula C is unique and determines the dependent structure of X;

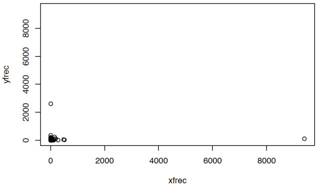
- C defines a multivariate distribution with uniform marginals.

$\Rightarrow$ The Sklar theorem, ensuring the existence and uniqueness of the copula, allows to separate marginal (univariate) distributions and dependence modeling.

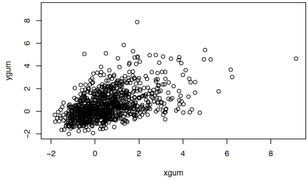
Does the Marginal Choice Matter?



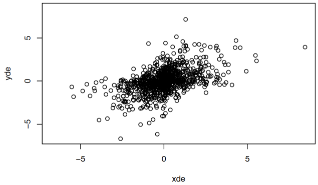
Uniform - $F_i(x) = x$



Fréchet - $F_i(x) = x^{-1}$



Gumbel - $F_i(x) = \exp\{-\exp(-x)\}$



Laplace - $F_i(x) = \exp -|x|$

Copulas with different marginals

- Copulas can have any marginal, the choice is arbitrary:

$$\begin{aligned} F(x) &= C_U\{F_1(x_1), \ldots, F_D(x_D)\}, \text{ Uniform} \\ &= C_F\{-1/\log F_1(x_1), \ldots, -1/\log F_D(x_D)\} \text{ Fréchet} \\ &= C_G[-\log\{-\log F_1(x_1)\}, \ldots, -\log\{-\log F_D(x_D)\}] \text{ Gumbel.} \end{aligned}$$

- There is thus an equivalence between C_U, C_F and C_G, e.g.,

$$C_U(u_1, \ldots, u_D) = C_F(-1/\log u_1, \ldots, -1/\log u_D).$$

- Changing of marginals can thus simply be made by repeated use of probability integral transform.

- The choice of marginal is in general made for convenience.

Bivariate PoT: Marginal standardization

- We focus on bivariate peaks-over-threshold for simplicity, but multivariate can be generalized very similarly.
- Suppose we have observations $X_1, \ldots, X_n \subset \mathbb{R}^2$, that we first normalize to a common marginal distribution to model dependence (Sklar theorem).
- We use standard Pareto margins, i.e.,

$$(Y_1, Y_2) = \left(\frac{1}{1 - F_{X_1}(X_1)}, \frac{1}{1 - F_{X_2}(X_2)} \right),$$

and then

$$F_{Y_1}(y) = F_{Y_2}(y) = 1 - y^{-1}.$$

Bivariate PoT: Marginal standardization

- We need to estimate F_{X_1} and F_{X_2}.
- For inference purposes only, we can simply use

$$F_X(x) = \tilde{F}(x),$$

where $\tilde{F}$ is the empirical distribution function.
- An alternative that can be used also for simulation and extrapolation is, for some high threshold u,

$$F_X(x) = \begin{cases} \tilde{F}(x), & x \leq u, \\ 1 - \{1 - \tilde{F}(u)\} \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-1/\xi}, & x > u, \end{cases}$$

where σ_u and ξ are the parameters of a GP distribution estimated on exceedances above threshold u.

Bivariate PoT: Regular variation

- We now assume that the vector X is regularly varying, i.e., that for any $A \subset (0,\infty)^2$
$$t\Pr\{(Y_1, Y_2) \in tA\} \to \Lambda(A) > 0, \quad t \to \infty, \tag{1}$$
where Λ is a non-degenerate measure on $(0,\infty)^2$.

- We recall that for the univariate case, we had assumed that $X_1 \in MDA(G)$ which translated into
$$n\{1 - F(a_n x_1 + b_n)\} \to -\log G(x_1), \tag{2}$$
but because of the Pareto normalization, $\xi = 1$ and we can choose $b_n \equiv 0$ and $a_n = n$, thus obtain
$$n\Pr\{X_1 > nx_1\} = n\Pr\{X_1 \in n(x_1, \infty)\} \to -\log G(x_1),$$
as $n \to \infty$

$\Rightarrow$ Equation (1) is the direct generalization of univariate EVT (2).

Return on structure variables

- We focus on a specific case of structure variables, i.e., variables for which the function r satisfies an homogeneity property:

$$r(tx) = tr(x), \quad t > 0, \quad x \in (0, \infty)^2.$$

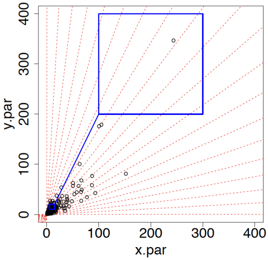
- For example, this includes:
  - $r(x) = x_1$;
  - $r(x) = \max(x_1, x_2)$;
  - $r(x) = (x_1 + x_2)/2$.

Pseudo-polar decomposition and illustration

- Regular variation implies that for some $u > 1$

$$\begin{aligned}\Pr\{Y_1 > ut, Y_2 > tu \mid r(Y) > t\} &= \frac{\Pr\{Y_1 > tu, Y_2 > tu\}}{\Pr\{r(Y) > u\}}, \\ &\to \frac{1}{u} \frac{\Lambda\{y \in (1,\infty)^2\}}{\Lambda\{y \in (0,\infty)^2 : r(y) > 1\}},\end{aligned}$$

as $t \to \infty$.



$\Rightarrow$ Regular variation implies a specific "direction" for extrapolation.

Measure of extremal dependence: the χ coefficient

- We now focus on $r(y_1, y_2) = y_1$, regular variation implies

$$\Pr\{Y_1 > u | Y_2 > u\} = \frac{u\Pr\{Y_1 > u, Y_2 > u\}}{u\Pr\{Y_2 > u\}}$$

$$\chi(u) \to \frac{\Lambda\{y \in (1,\infty) \times (1,\infty)\}}{\Lambda\{y \in (0,\infty) \times (1,\infty)\}}$$

$$\to \chi > 0.$$

- The limit χ is used as a pairwise measure of (asymptotic) extremal dependence.

- This measure is very popular for its simplicity and interpretability.

- χ is linked to the so-called extremal coefficient θ through

$$\chi = 2 - \theta.$$

Asymptotic Independence

- What happens if Y_1 and Y_2 are independent?
- We have simply $F(y_1, y_2) = F(y_1)F(y_2)$, and thus

$$\begin{aligned}\Pr\{Y_1 > u | Y_2 > u\} &= \frac{\Pr\{Y_1 > u, Y_2 > u\}}{\Pr\{Y_2 > u\}} \\ &= \frac{\{1 - F(u)\}\{1 - F(u)\}}{\{1 - F(u)\}}, \\ &= 1 - F(u), \\ \chi &\to 0,\end{aligned}$$

  as $u \to \infty$.
- More generally *asymptotic independence* is defined when $\chi = 0$.
- ⚠ Asymptotic independence does not mean that Y_1 and Y_2 are exactly independent.

Asymptotic independence

- If $\chi = 0$, then regular variation cannot hold and

$$t\Pr(Y_1 > ty_1, Y_2 > ty_2) \to 0.$$

- Once again if Y_1 and Y_2 were exactly independent:

$$\begin{aligned}
\Pr(Y_1 > ty, Y_2 > ty) &= \frac{1}{(ty)^2} \\
&= \frac{1}{t^2}\frac{1}{y^2} \\
&= \frac{1}{t^2}\Pr(Y_1 > y, Y_2 > y) \\
&\neq \frac{1}{t}\Pr(Y_1 > y, Y_2 > y).
\end{aligned}$$

- In case of perfect independence, we observe a faster decay rate than regular variation.

Measuring asymptotic independence

- Ledford and Tawn (1996) propose to study

$$\begin{aligned}\Pr(Y_1 > ty, Y_2 > ty) &\sim \frac{1}{t^{1/\eta}} \Pr(Y_1 > y, Y_2 > y) \\ &\sim \frac{1}{t^{1/\eta}} \mathcal{L}(y),\end{aligned}$$

  where $0 < \eta \neq 1$ and $\mathcal{L}$ is a "slowly varying function".

- The coefficient η measure the strength of "subasymptotic dependence", i.e., the speed at which the vector converge to independence.

- For $\eta = 1$, we retrieve regular variation and asmyptotic independence, so the larger is η the slower is the convergence.

Alternative measure of asymptotic independence

- We define the coefficient

$$\bar{\chi} = 2\eta - 1, \quad \bar{\chi} \in (-1, 1].$$

- $\bar{\chi} = 1$ asymptotic dependence with $\chi = \lim_{y\to\infty} \mathcal{L}(y) > 0$;
- $\bar{\chi} < 1$ asymptotic independence and $\chi = 0$;
- We distinguish three regimes of asymptotic independence:
  - $0 < \bar{\chi} < 1$ positive extremal association;
  - $\bar{\chi} = 0$ near extremal independence;
  - $-1 < \bar{\chi} < 0$ negative extremal association;

Example: Gaussian Copula

- Suppose that the dependence of Y is driven by a Gaussian copula with correlation ρ,

$$C_U(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \phi(y, \rho) dy,$$

  where
  - ϕ is the density function of a bivariate normal distribution with mean 0 and correlation $\rho \in (-1; 1)$,
  - Φ is the cumulative distribution function of a univariate Gaussian distribution.
- In this case, we have

$$\chi = 0, \quad \bar{\chi} = \rho < 1.$$

- The Gaussian copula motivated the parametrization of $\bar{\chi}$.

Measure of extremal dependence: Summary

Asymptotic dependence	Asymptotic independence
$\chi > 0$	$\chi = 0$
$\bar{\chi} = 1$	$-1 < \bar{\chi} < 1$

- Recall that χ and $\bar{\chi}$ are *asymptotic* quantities.
- In practice, we need to estimate these measures from a finite number of observations / finite threshold, i.e., for large enough threshold

$$\Pr(Y_1 > u | Y_2 > u) = \chi(u) \approx \chi$$

and for $t > 1$,

$$\begin{aligned}\Pr(Y_1 > tu, Y_2 > tu | Y_1 > u, Y_2 > u) &= \Pr(\min(Y_1, Y_2) > tu | \min(Y_1, Y_2) > u) \\ &\approx \left\{\frac{u}{t}\right\}^{1/\eta(u)}\end{aligned}$$

Estimating χ and $\bar{\chi}$

- Suppose we have observations $X_1, \ldots, X_n$ that have been normalized to $Y_1, \ldots, Y_n$ with unit Pareto margins.

- Empirical estimator for χ for $u \gg 0$:

$$\hat{\chi}(u) = \frac{\sum_{i=1}^{n} 1\{Y_{1,i} > u, Y_{2,i} > u\}}{\sum_{i=1}^{n} 1\{Y_{2,i} > u\}}$$

- Estimator for $\widehat{\bar{\chi}} = 2\hat{\eta} - 1$:
  - compute pairwise minimas $Z_i = \min(Y_{1,i}, Y_{2,i})$;
  - Estimate the tail index $\hat{\eta}$ of the Pareto distribution using observations Z_i exceeding threshold u.

Link between χ and $\bar{\chi}$ and copulas

- The is an explicit link between the bivariate copula C_U and both extremal coefficients (Coles et al., 1999).

- For $u \in (0;1)$,

$$\chi(u) = \frac{1 - 2u + C_U(u,u)}{1-u} = 2 - \frac{\log C_U(u,u)}{\log u}$$

and

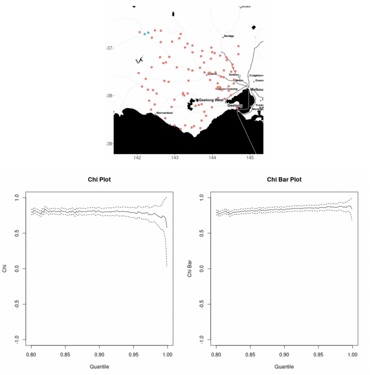
$$\bar{\chi}(u) = \frac{2\log(1-u)}{\bar{C}_U(u,u)} - 1,$$

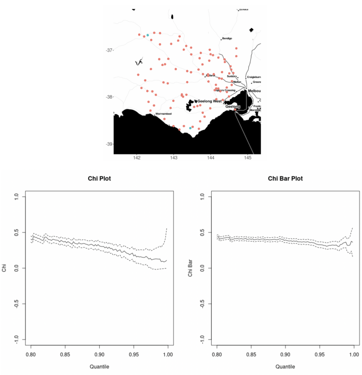
where $\bar{C}_U$ is the inverted copular, i.e., $C(u) = \bar{C}_U(1-u, 1-u)$.

- Both quantities can thus be estimated by replacing C_U and $\bar{C}_U$ by their empirical estimates.

Estimating χ and $\bar{\chi}$ with R

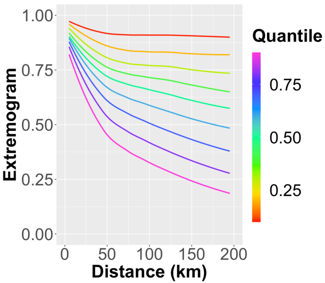
- The function `chiplot` in the package `evd` will estimate both coefficient as function of the threshold u.
- Example of rainfall in the state of Victoria, AU

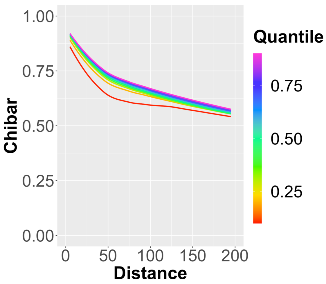




Spatial dependence

- In a spatial setting, we can estimate χ and $\bar{\chi}$ as function of the distance.





Bibliography

Barnett, V. (1976). The ordering of multivariate data. *Source: Journal of the Royal Statistical Society. Series A (General)*, 139:318–355.

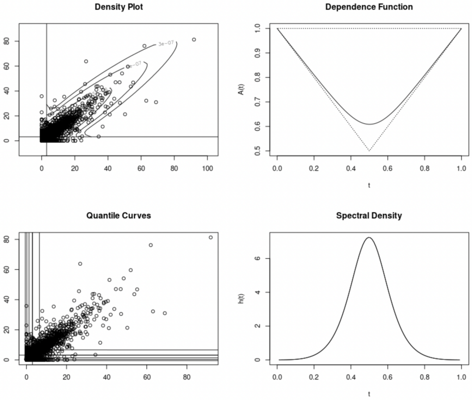
Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339–365.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate to structural design extremes: an application to structural design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43:1–48.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187.

Parametric Copula Models

- The function `fbvpot` (resp. `fbvevd` for block maxima) in the package `evd` enable automatic fit of a large number of parametric models.
- Example of an asymmetric logistic model fit for rainfall measurements from station 6 and 16 in the state of Victoria, AU:



Parametric Copula Models

Model	AIC	χ
Logistic	34830.66	0.784
Bi-logistic	34832.67	0.784
Asymmetric logistic	34690.47	0.783
Negative logistic	34879.74	0.785
Negative bi-logistic	34881.73	0.785
Asymmetric negative logistic	36498.8	0.5
Coles-Tawn	35091.2	0.760
Husler-Reiss	35187.61	0.762
Asymmetric mixed	36401.8	0.5

Model fit summary for all models available in `fbvpot`.

Disclaimer

This presentation has been developed as teaching material for the course "Extreme Value Statistics" and does not reflect the position of the Swiss Confederation.