Estimation of risk measures for extreme pluviometrical measurements

by

Jonathan EL METHNI

in collaboration with

Laurent GARDES & Stéphane GIRARD

Workshop "Extreme value modeling and water resources"
CNRS center in Aussois – June 2016

Outline

Return level

- The rainfall is modeled by a random variable Y with survival function

$$\overline{F}(y) = 1 - F(y) = \mathbb{P}(Y \geq y)$$

- We have an ordered sample of annual rainfall

$$Y_{1,n} \leq \cdots \leq Y_{n,n}.$$

- We want to estimate the level of rain H which is exceeded on average once in T years *i.e.* we want to estimate H such that

$$1/T = \mathbb{P}(Y \geq H) = \overline{F}(H)$$

i.e. we want to estimate

$$H = \overline{F}^{-1}(1/T)$$

- H is the return level corresponding to a return period T.
- It is the standard quantity of interest in environmental studies.

Return level and quantile

The quantile of order $\alpha \in]0,1[$ of the survival function is defined by

$$q(\alpha) = \overline{F}^{-1}(\alpha).$$

$\Longrightarrow$ The return level $H = \overline{F}^{-1}(1/T)$ is by consequence a quantile of order $\alpha = 1/T$.

- What happens if the return period T is greater than the observed period ?
  In other words what happens if

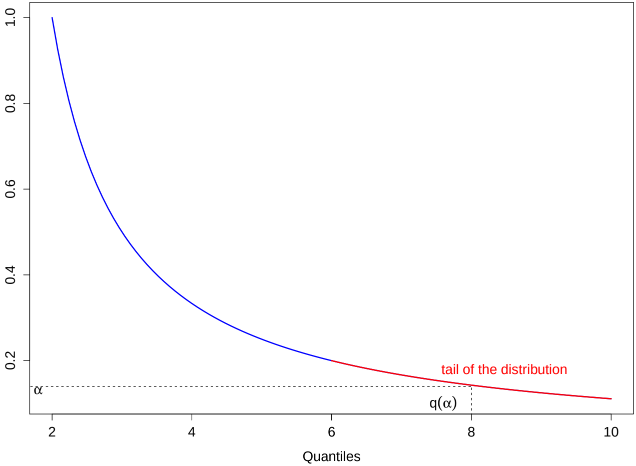
$$T > n \Longleftrightarrow \alpha = 1/T < 1/n \underset{n \to \infty}{\longrightarrow} 0 \quad ?$$

We want to estimate extreme quantiles $q(\alpha_n)$ of order α_n defined by

$$q(\alpha_n) = \overline{F}^{-1}(\alpha_n) \quad \text{with} \quad \alpha_n \to 0 \quad \text{when} \quad n \to \infty$$

- A return level with a return period greater than the observed period is an extreme quantile.

Extremes Quantiles



What happens if $q(\alpha_n) > Y_{n,n}$?

Extremes Quantiles

We can prove that

$$\mathbb{P}(q(\alpha_n) > Y_{n,n}) = \exp(-n\alpha_n(1 + o(1))$$

- First case :

$$\text{If } n\alpha_n \to \infty \text{ then } \mathbb{P}(q(\alpha_n) > Y_{n,n}) \to 0.$$

  A natural estimator is the order statistic $Y_{n-\lfloor n\alpha_n \rfloor+1,n}$ (where $\lfloor . \rfloor$ is the floor function).

- Second case :

$$\text{If } n\alpha_n \to 0 \text{ then } \mathbb{P}(q(\alpha_n) > Y_{n,n}) \to 1.$$

  In this case, we can not estimate $q(\alpha_n)$ by simply inverting the empirical cumulative distribution function :

$$\widehat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{Y_i \leq y\} \quad \text{because} \quad \widehat{F}_n(y) = 1 \quad \text{for} \quad y \geq Y_{n,n}.$$

The behavior of $Y_{n,n}$ is caracterised by its cumulative distribution function $F_{Y_{n,n}}(y) = F^n(y)$ which has a degenerate distribution.

The Fisher-Tippett-Gnedenko theorem [1943]

Theorem

Let $(Y_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables with cumulative distribution function F. If there exist two normalizing sequences $(a_n)_{n\geq 1} > 0$ and $(b_n)_{n\geq 1} \in \mathbb{R}$ and a non degenerate distribution $\mathcal{H}_\gamma$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y_{n,n} - b_n}{a_n} \leq y\right) = \lim_{n\to\infty} F^n(a_n y + b_n) = \mathcal{H}_\gamma(y),$$

then we have

$$\mathcal{H}_\gamma(y) = \exp\left(-(1+\gamma y)_+^{-1/\gamma}\right),$$

where $\gamma \in \mathbb{R}$ and $z_+ = \max(0, z)$.

- $\mathcal{H}_\gamma$ is called the cumulative distribution function of extreme value distribution.
- a_n and b_n are normalizing sequences.
- If F verifies the Fisher-Tippett-Gnedenko theorem we say that F belongs to the domain of attraction of $\mathcal{H}_\gamma$.
- This distribution depends on the unique shape parameter γ called extreme value index or tail index.

Three domains of attraction

According to the sign of γ, we distinguish between three domains of attraction :

- if $\gamma < 0$, we say that F belongs to the domain of attraction of Weibull. It containts distributions with survival function without tail distribution, *i.e.* short-tailed distributions.
- if $\gamma = 0$, we say that F belongs to the domain of attraction of Gumbel. It containts distributions with survival function exponentially decreasing, *i.e.* light-tailed distributions.
- if $\gamma > 0$, we say that F belongs to the domain of attraction of Fréchet. It containts distributions with survival function polynomially decreasing, *i.e.* heavy-tailed distributions.

Fréchet ($\gamma > 0$)	Gumbel ($\gamma = 0$)	Weibull ($\gamma < 0$)
Pareto	Normale	Uniforme
Student	Exponentielle	Beta
Burr	Log-normale	ReverseBurr
Chi-deux	Gamma	
Fréchet	Weibull	
Log-gamma	Logistique	
Log-logistique	Gumbell	
Cauchy		

Heavy-tailed distributions

All distributions belonging to the domain of attraction of Fréchet can be rewritten

$$\overline{F}(y) = y^{-1/\gamma}\ell(y) \quad \text{with} \quad \gamma > 0,$$

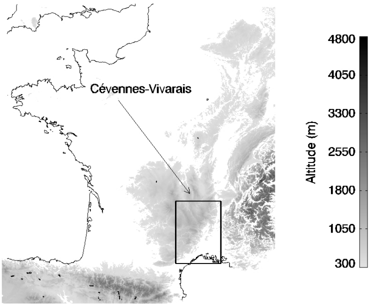
and ℓ is a slowly varying function at infinity *i.e.* $\forall \lambda \geq 1$,

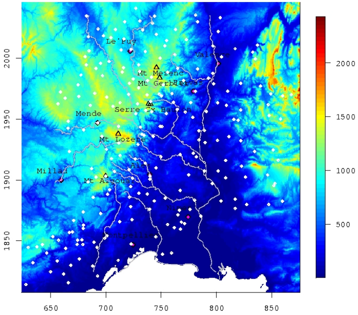
$$\lim_{y \to \infty} \frac{\ell(\lambda y)}{\ell(y)} \to 1.$$

- The parameter γ controls the behaviour of the tail of the survival function and by consequence the behaviour of the extreme values.

In numerous applications the variable of interest is recorded simultaneously with some covariate.

The Cévennes-Vivarais region

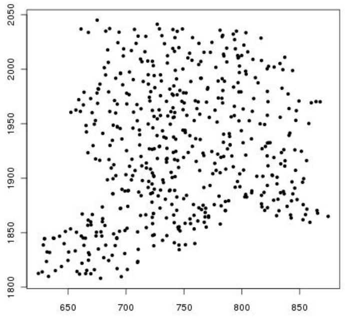




Horizontally we have the longitude (in km), verticaly the latitude (in km) and in the colour scale the altitude (in m). Some stations of interest (white rombus).

Description of the real data set

The 523 stations of interest



- Data set provided by Météo-France.
- Y : daily rainfall measured in mm.
- $X = \{$longitude, lattitude, altitude$\}$.

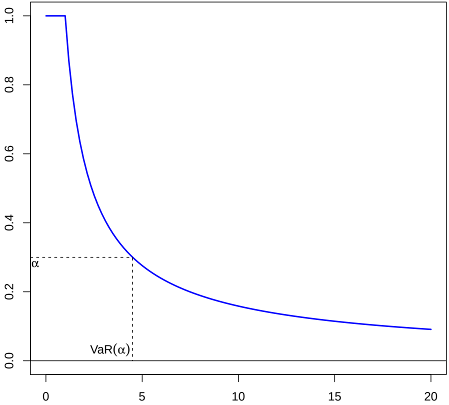
- 1958 $\Longrightarrow$ 2000 =43 years of data.
- 523 stations denoted $\{x_t; t = 1, \dots, 523\}$.
- Observations = 5 513 734.

Aim $\Longrightarrow$ to obtain maps of estimated extreme risk measures in all point of the region.

The Value-at-Risk

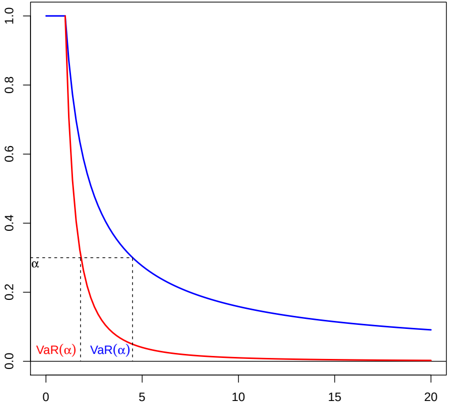
- The Value-at-Risk of level $\alpha \in (0,1)$ denoted by $\mathrm{VaR}(\alpha)$ introduced in 1993 is defined by

$$\mathrm{VaR}(\alpha) := \overline{F}^{-1}(\alpha).$$



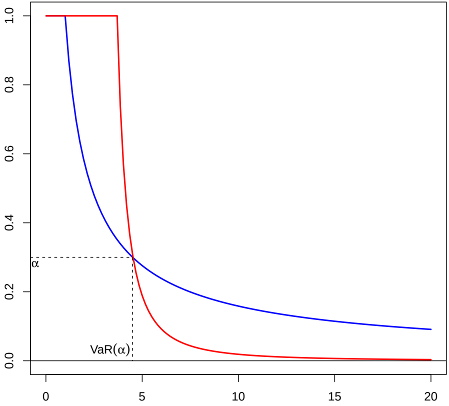
- The $\mathrm{VaR}(\alpha)$ is the quantile of level α of the survival function of the *r.v.* Y.

- Let us consider Y_1 and Y_2 two loss *r.v.* with survival function associated $\overline{F}_1$ and $\overline{F}_2$.



- The $\mathrm{VaR}(\alpha)$ is different according to the survival function $\overline{F}_1$ or $\overline{F}_2$.

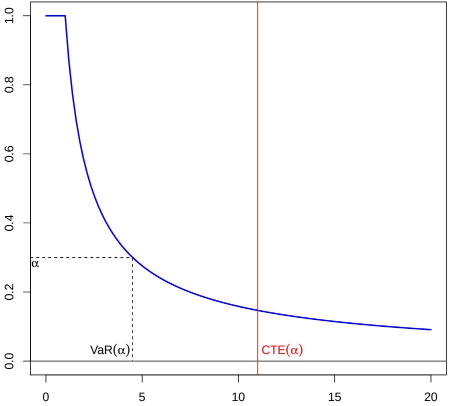
- Let us consider Y_1 and Y_2 two loss *r.v.* with survival function associated $\overline{F}_1$ and $\overline{F}_2$.



- In particular, *r.v.* with light tail probabilities and with heavy tail probabilities (Embrechts et al. 1997) may have the same VaR(α). This point is one main criticism against VaR as a risk measure.

- The Conditional Tail Expectation of level $\alpha \in (0,1)$ denoted CTE(α) (see Embrechts *et al.* [1997]) is defined by

$$\mathrm{CTE}(\alpha) := \mathbb{E}(Y|Y > \mathrm{VaR}(\alpha)).$$

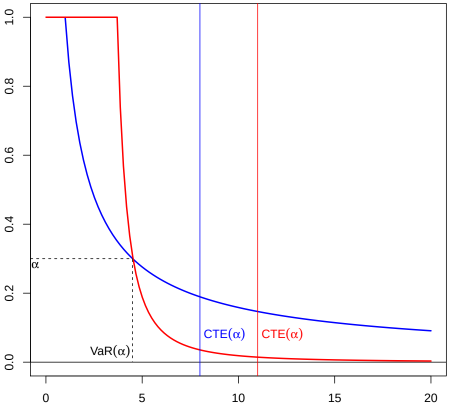


- The CTE(α), takes into account the whole information contained in the upper part of the tail distribution and gives informations on the distribution of Y given that $Y > \mathrm{VaR}(\alpha)$ and contrary to the $\mathrm{VaR}(\alpha)$, on the heaviness of the tail distribution.

The Conditional Tail Expectation

- The Conditional Tail Expectation of level $\alpha \in (0,1)$ denoted CTE(α) (see Embrechts *et al.* [1997]) is defined by

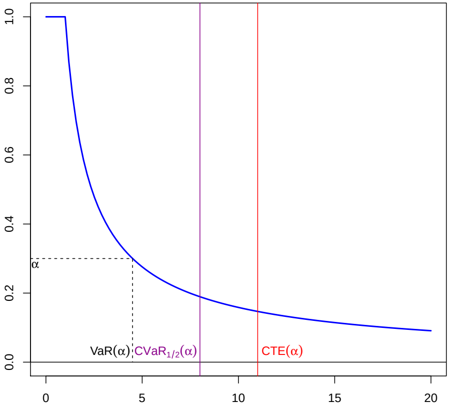
$$\mathrm{CTE}(\alpha) := \mathbb{E}(Y|Y > \mathrm{VaR}(\alpha)).$$



- The CTE(α), takes into account the whole information contained in the upper part of the tail distribution and gives informations on the distribution of Y given that $Y > \mathrm{VaR}(\alpha)$ and contrary to the $\mathrm{VaR}(\alpha)$, on the heaviness of the tail distribution.

- The Conditional-Value-at-Risk of level $\alpha \in (0,1)$ denoted $\mathrm{CVaR}_\lambda(\alpha)$ and introduced by Rockafellar and Uryasev [2000] is defined by

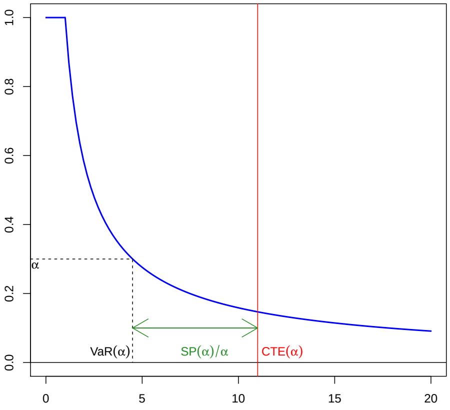
$$\mathrm{CVaR}_\lambda(\alpha) := \lambda \mathrm{VaR}(\alpha) + (1-\lambda)\mathrm{CTE}(\alpha) \quad \text{with} \quad 0 \leq \lambda \leq 1.$$



- We can remark that $\mathrm{CVaR}_1(\alpha)$=$\mathrm{VaR}(\alpha)$ and $\mathrm{CVaR}_0(\alpha)$=$\mathrm{CTE}(\alpha)$.

- The risk measure Stop-loss Premium reinsurance with a retention level equals to VaR(α) (see Cai and Tan [2007]) is defined by

$$\mathrm{SP}(\alpha) := \mathbb{E}((Y - \mathrm{VaR}(\alpha))_+) = \alpha\,(\mathrm{CTE}(\alpha) - \mathrm{VaR}(\alpha))\,.$$



- This risk measure thus permits to emphasize the dangerous cases.

Our contributions consist in adding two difficulties in the framework of the estimation of risk measures.

1. We add the presence of a random covariate $X \in \mathbb{R}^p$.

2. We are interested in the estimation of risk measures in the case of extremes losses.

$\Longrightarrow$ To this end, we replace the fixed order $\alpha \in (0,1)$ by a sequence $\alpha_n \to 0$ as the sample size $n \to \infty$.

Denoting by $\overline{F}(.|x)$ the conditional survival function of Y given that $X = x$, we define the Regression Value-at Risk by

$$\mathrm{RVaR}(\alpha_n|x) := \overline{F}^{-1}(\alpha_n|x),$$

and the Regression Conditional Tail Expectation by

$$\mathrm{RCTE}(\alpha_n|x) := \mathbb{E}(Y|Y > \mathrm{RVaR}(\alpha_n|x), X = x),$$

$\Longrightarrow$ All the risk measures depend on the RVaR and on the RCTE.

$$\begin{aligned}
\mathrm{RCVaR}_\lambda(\alpha_n|x) &= \lambda \mathrm{RVaR}(\alpha_n|x) + (1-\lambda)\mathrm{RCTE}(\alpha_n|x), \\
\mathrm{RSP}(\alpha_n|x) &= \alpha_n(\mathrm{RCTE}(\alpha_n|x) - \mathrm{RVaR}(\alpha_n|x)).
\end{aligned}$$

$\Longrightarrow$ We want to estimate all these risk measures.

We define the conditional moment of order $a \geq 0$ of Y given that $X = x$ by

$$\varphi_a(y|x) = \mathbb{E}\left(Y^a \mathbb{I}\{Y > y\}|X = x\right),$$

where $\mathbb{I}\{.\}$ is the indicator function.

Since $\varphi_0(y|x) = \overline{F}(y|x)$, it follows

$$\begin{aligned}
\mathrm{RVaR}(\alpha_n|x) &= \varphi_0^{-1}(\alpha_n|x), \\
\mathrm{RCTE}(\alpha_n|x) &= \frac{1}{\alpha_n}\varphi_1(\varphi_0^{-1}(\alpha_n|x)|x).
\end{aligned}$$

Estimator of $\varphi_a(.|x)$:

To estimate the moment of order $a \geq 0$ of Y given that $X = x$, we propose to use a classical kernel estimator defined for $(x, y) \in \mathbb{R}^p \times \mathbb{R}$ by

$$\widehat{\varphi}_{a,n}(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) Y_i^a \mathbb{I}\{Y_i > y\} \Bigg/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

- The fonction K is called kernel. It is a bounded density on $\mathbb{R}^p$, with support S included in the unit ball of $\mathbb{R}^p$.
- (h_n) is a non-random sequence such that $h_n \to 0$ when $n \to \infty$ called the window-width.

Estimator of $\varphi_a^{-1}(.|x)$:

Since $\hat{\varphi}_{a,n}(.|x)$ is a non-increasing function, we can define an estimator of $\varphi_a^{-1}(.|x)$ by

$$\hat{\varphi}_{a,n}^{-1}(\alpha|x)$$

The Regression Value-At-Risk of level α_n is thus estimated by

$$\widehat{\mathrm{RVaR}}_n(\alpha_n|x) = \widehat{\varphi}_{0,n}^{-1}(\alpha_n|x),$$

and the Regression Conditional Tail Expectation is estimated by :

$$\widehat{\mathrm{RCTE}}_n(\alpha_n|x) = \frac{1}{\alpha_n}\widehat{\varphi}_{1,n}\left(\widehat{\varphi}_{0,n}^{-1}(\alpha_n|x)|x\right)$$

$\Longrightarrow$ We thus can now estimate all the above mentioned risk measures.

To do it, we need the asymptotic joint distribution of

$$\left\{\widehat{\mathrm{RCTE}}_n(\alpha_n|x)\ ,\ \widehat{\mathrm{RVaR}}_n(\alpha_n|x)\right\}$$

(F) We assume that the conditional survival distribution function of Y given $X = x$ is heavy-tailed and admits a probability density function.

This is equivalent to assume that

$$\forall y > 0, \quad \text{we have} \quad \overline{F}(y|x) = y^{-1/\gamma(x)}\ell(y|x)$$

where in this context,

- $\gamma(.)$ is an unknown and positive function of the covariate x and will be called conditional tail index since it tunes the tail heaviness of the conditional distribution of Y given $X = x$.
- $\ell(.|x)$ is a slowly varying function at infinity. We have (for x fixed), for all $\lambda > 0$,

$$\lim_{y\to\infty} \frac{\ell(\lambda y|x)}{\ell(y|x)} = 1.$$

This hypothesis amounts to assuming that the conditional distribution of Y given $X = x$ is heavy-tailed.

Theorem 1

Suppose **(F)** hold and for all $x \in \mathbb{R}^p$ such that $g(x) > 0$ and $0 < \gamma(x) < 1/2$ we have a sequence $(\alpha_n)_{n\geq 1}$ such that $\alpha_n \to 0$ and $nh_n^p \alpha_n \to \infty$ as $n \to \infty$, then, the random vector

$$\sqrt{nh_n^p \alpha_n} \left\{ \left(\frac{\widehat{\mathrm{RCTE}}_n(\alpha_n|x)}{\mathrm{RCTE}(\alpha_n|x)} - 1 \right)_{j\in\{1,\ldots,J\}}, \left(\frac{\widehat{\mathrm{RVaR}}_n(\alpha_n|x)}{\mathrm{RVaR}(\alpha_n|x)} - 1 \right) \right\}$$

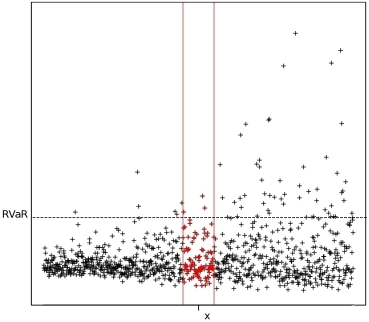
is asymptotically Gaussian, centred, with a covariance matrix

$$\Sigma(x) = \gamma^2(x) \frac{\|K\|_2^2}{g(x)} \begin{pmatrix} \frac{2(1-\gamma(x))}{1-2\gamma(x)} & 1 \\ 1 & 1 \end{pmatrix}.$$

- $\Sigma(x)$ is proportional to $\gamma^2(x) \Longrightarrow$ the higher the $\gamma(x)$ (*i.e.* the heavier is the tail) the more the variance of our estimators increases.
- The density $g(x)$ of the covariate is part of the denominator of the covariance matrix $\Longrightarrow$ the less the point (*i.e.* the smaller is the density) the more the variance of our estimators increases.

Conditions on the sequences α_n and h_n

- $nh_n^p\alpha_n \to \infty$: necessary and sufficient condition for the almost sure presence of at least one point in the region $B(x, h_n) \times [\mathrm{RVaR}(\alpha_n|x), +\infty)$ of $\mathbb{R}^p \times \mathbb{R}$.



- If $\alpha_n = \alpha$ is fixed we find the classical condition of asymptotic normality : $nh_n^p \to \infty$.
- If $h_n = h$ is fixed we find the classical condition of asymptotic normality : $n\alpha_n \to \infty$.

Extrapolation

- In Theorem 1, the condition $nh_n^p \alpha_n \to \infty$ provides a lower bound on the level of the risk measure to estimate.
- This restriction is a consequence of the use of a kernel estimator which cannot extrapolate beyond the maximum observation in the ball $B(x, h_n)$.
- In consequence, α_n must be an order of an extreme quantile within the sample.

Definition

Let us consider $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ two positives sequences such that $\alpha_n \to 0$, $\beta_n \to 0$ and $0 < \beta_n < \alpha_n$. A kernel adaptation of Weissman's estimator [1978] is given by

$$\widehat{\mathrm{RCTE}}_n^W(\beta_n|x) = \widehat{\mathrm{RCTE}}_n(\alpha_n|x) \left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n(x)}$$

Extrapolation

- In Theorem 1, the condition $nh_n^p \alpha_n \to \infty$ provides a lower bound on the level of the risk measure to estimate.
- This restriction is a consequence of the use of a kernel estimator which cannot extrapolate beyond the maximum observation in the ball $B(x, h_n)$.
- In consequence, α_n must be an order of an extreme quantile within the sample.

Definition

Let us consider $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ two positives sequences such that $\alpha_n \to 0$, $\beta_n \to 0$ and $0 < \beta_n < \alpha_n$. A kernel adaptation of Weissman's estimator [1978] is given by

$$\widehat{\mathrm{RCTE}}_n^W(\beta_n|x) = \widehat{\mathrm{RCTE}}_n(\alpha_n|x) \left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n(x)}$$

where

- $\widehat{\mathrm{RCTE}}_n(\alpha_n|x)$ is the previous kernel estimator.

Extrapolation

- In Theorem 1, the condition $nh_n^p\alpha_n \to \infty$ provides a lower bound on the level of the risk measure to estimate.
- This restriction is a consequence of the use of a kernel estimator which cannot extrapolate beyond the maximum observation in the ball $B(x, h_n)$.
- In consequence, α_n must be an order of an extreme quantile within the sample.

Definition

Let us consider $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ two positives sequences such that $\alpha_n \to 0$, $\beta_n \to 0$ and $0 < \beta_n < \alpha_n$. A kernel adaptation of Weissman's estimator [1978] is given by

$$\widehat{\mathrm{RCTE}}_n^W(\beta_n|x) = \widehat{\mathrm{RCTE}}_n(\alpha_n|x)\left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n(x)}$$

where

- $\widehat{\mathrm{RCTE}}_n(\alpha_n|x)$ is the previous kernel estimator.
- $(\alpha_n/\beta_n)^{\hat{\gamma}_n(x)}$ is the term which allows the extrapolation.

Extrapolation

- In Theorem 1, the condition $nh_n^p\alpha_n \to \infty$ provides a lower bound on the level of the risk measure to estimate.
- This restriction is a consequence of the use of a kernel estimator which cannot extrapolate beyond the maximum observation in the ball $B(x, h_n)$.
- In consequence, α_n must be an order of an extreme quantile within the sample.

Definition

Let us consider $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ two positives sequences such that $\alpha_n \to 0$, $\beta_n \to 0$ and $0 < \beta_n < \alpha_n$. A kernel adaptation of Weissman's estimator [1978] is given by

$$\widehat{\mathrm{RCTE}}_n^W(\beta_n|x) = \widehat{\mathrm{RCTE}}_n(\alpha_n|x)\left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n(x)}$$

where

- $\widehat{\mathrm{RCTE}}_n(\alpha_n|x)$ is the previous kernel estimator.
- $(\alpha_n/\beta_n)^{\hat{\gamma}_n(x)}$ is the term which allows the extrapolation.
- $\hat{\gamma}_n(x)$ is an estimator of the conditional tail index.

Theorem 2

Suppose the assumptions of Theorem 1 hold. Let us consider $\hat{\gamma}_n(x)$ an estimator of the tail index such that

$$\sqrt{nh_n^p\alpha_n}(\hat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \mathcal{N}\left(0, v^2(x)\right),$$

with $v(x) > 0$. If $(\beta_n)_{n\geq 1}$ is a positive sequence such that $\beta_n \to 0$ and $\beta_n/\alpha_n \to 0$ as $n \to \infty$, then for all $x \in \mathbb{R}^p$, we have

$$\frac{\sqrt{nh_n^p\alpha_n}}{\log(\alpha_n/\beta_n)}\left(\frac{\widehat{\mathrm{RCTE}}_n^W(\beta_n|x)}{\mathrm{RCTE}(\beta_n|x)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, v^2(x)\right).$$

- The condition $\beta_n/\alpha_n \to 0$ allows us to extrapolate and choose a level β_n arbitrarily small.
- Daouia *et al.* [2011] have established the asymptotic normality of

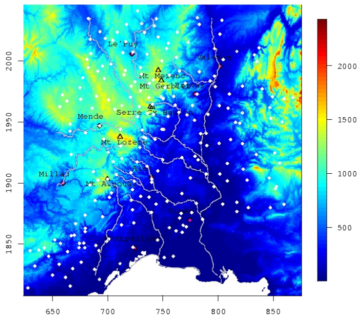
$$\widehat{\mathrm{RVaR}}_n^W(\beta_n|x) = \widehat{\mathrm{RVaR}}_n(\alpha_n|x)\left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n(x)}$$

Outline

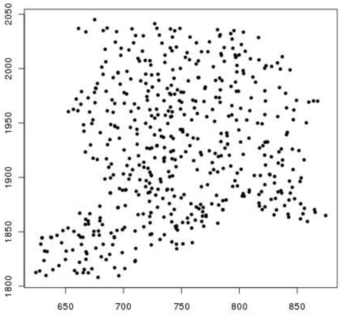
Application to an hydrological data set

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



The 523 stations of interest

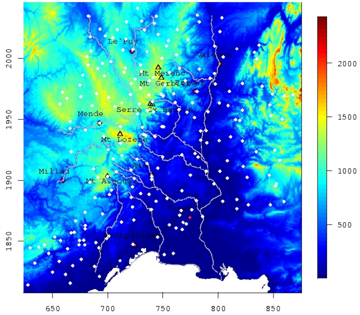


Aim $\Longrightarrow$ estimate $\mathrm{RVaR}(\beta_n|x)$ and $\mathrm{RCTE}(\beta_n|x)$ for $\beta_n = 1/(100 * 365.25)$.

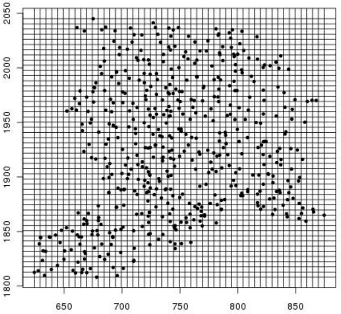
Application to an hydrological data set

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



Grid 200 $\times$ 200 points

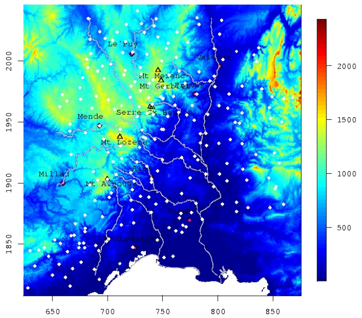


Aim $\Longrightarrow$ estimate $\mathrm{RVaR}(\beta_n|x)$ and $\mathrm{RCTE}(\beta_n|x)$ for $\beta_n = 1/(100 * 365.25)$.

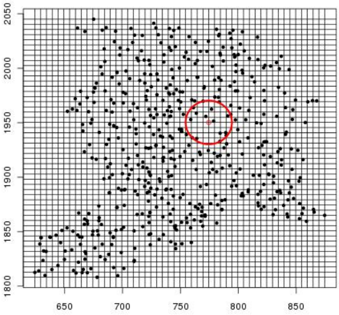
Application to an hydrological data set

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



Work in $B(x, h_n)$



Aim $\Longrightarrow$ estimate $\mathrm{RVaR}(\beta_n|x)$ and $\mathrm{RCTE}(\beta_n|x)$ for $\beta_n = 1/(100 * 365.25)$.

Procedure to choose (h_n, α_n)

Our estimators depend on the following two tuning parameters :

- h_n the bandwidth : recurrent issue in non-parametric statistics.
- α_n number of upper order statistics used : classical issue since it raises a compromise bias/variance in extreme value theory.

A high value of $\alpha_n \Longrightarrow$ large bias since we move out of the tail distribution.
A small value of $\alpha_n \Longrightarrow$ large variance since we use few observations.

$\Longrightarrow$ Procedure to select simultaneously (h_n, α_n).

Our procedure is based on the estimation of the function $\gamma(x)$ since it controls :

- the tail heaviness of the distribution (see assumption **(F)**),
- and the extrapolation.

The main idea of our procedure is to select the empirical pair $(h_{emp}, \alpha_{emp}) \in \mathcal{H} \times \mathcal{A}$ for which two different estimations of the tail index $\gamma(x_t)$ for each station t are closed.

- Hydrologists $\Longrightarrow$ important to combine locals and regionals informations.

The Hill estimator

- Without covariate

Let us consider $(\alpha_n)_{n\geq 1}$ a positive sequence such that $\alpha_n \to 0$, the Hill estimator [1975] is defined by :

$$\hat{\gamma}_{n,\alpha_n} = \frac{1}{\lfloor n\alpha_n \rfloor - 1} \sum_{i=1}^{\lfloor n\alpha_n \rfloor - 1} \log\left(Y_{n-i+1,n}\right) - \log\left(Y_{n-\lfloor n\alpha_n \rfloor+1,n}\right).$$

- With a covariate

Let us consider $(\alpha_n)_{n\geq 1}$ a positive sequence such that $\alpha_n \to 0$. A kernel version of the Hill estimator (Gardes and Girard [2008]) is given by

$$\hat{\gamma}_{n,\alpha_n}(x) = \sum_{j=1}^{J} (\log(\widehat{\mathrm{RVaR}}_n(\tau_j\alpha_n|x)) - \log(\widehat{\mathrm{RVaR}}_n(\tau_1\alpha_n|x))) \Bigg/ \sum_{j=1}^{J} \log(\tau_1/\tau_j),$$

where $J \geq 1$ and $(\tau_j)_{j\geq 1}$ is a positive and decreasing sequence of weights.

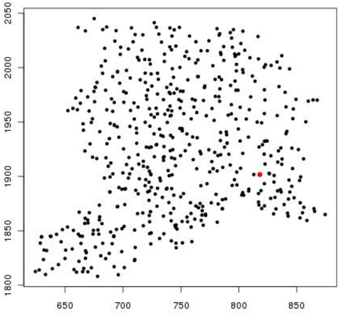
A leave one out cross validation procedure to select h_n and α_n : Step 1

- Double loop on $\mathcal{H} = \{h_i; i = 1, \ldots, M\}$ and on $\mathcal{A} = \{\alpha_j; j = 1, \ldots, R\}$.
- Loop on all stations $\{x_t; t = 1, \ldots, 523\}$.

A leave one out cross validation procedure to select h_n and α_n : Step 1

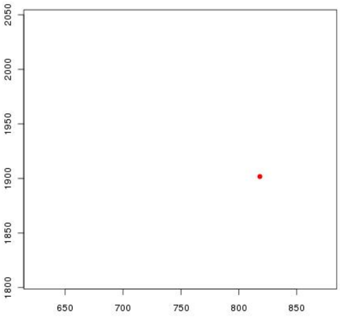
- Double loop on $\mathcal{H} = \{h_i; i = 1, \ldots, M\}$ and on $\mathcal{A} = \{\alpha_j; j = 1, \ldots, R\}$.
- Loop on all stations $\{x_t; t = 1, \ldots, 523\}$.

Place ourselves at the station x_t



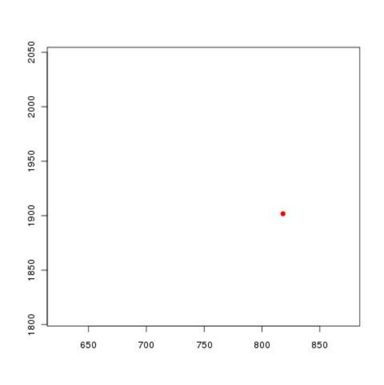
- Double loop on $\mathcal{H} = \{h_i; i = 1, \ldots, M\}$ and on $\mathcal{A} = \{\alpha_j; j = 1, \ldots, R\}$.
- Loop on all stations $\{x_t; t = 1, \ldots, 523\}$.

Remove all the others stations



- Double loop on $\mathcal{H} = \{h_i; i = 1, \ldots, M\}$ and on $\mathcal{A} = \{\alpha_j; j = 1, \ldots, R\}$.
- Loop on all stations $\{x_t; t = 1, \ldots, 523\}$.

Remove all the others stations

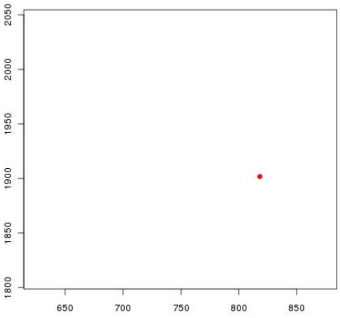


- We estimate $\gamma > 0$ using the classical Hill estimator.
- It only depends on α_j.
- The α_j are chosen such that we stay in the tail of the distribution $\max_{j \in \{1,\ldots,R\}} (\alpha_j) < 0.1$

A leave one out cross validation procedure to select h_n and α_n : Step 1

- Double loop on $\mathcal{H} = \{h_i; i = 1, \ldots, M\}$ and on $\mathcal{A} = \{\alpha_j; j = 1, \ldots, R\}$.
- Loop on all stations $\{x_t; t = 1, \ldots, 523\}$.

Remove all the others stations

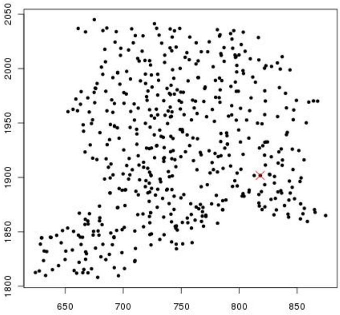


- We estimate $\gamma > 0$ using the classical Hill estimator.
- It only depends on α_j.
- The α_j are chosen such that we stay in the tail of the distribution $\max_{j \in \{1,\ldots,R\}} (\alpha_j) < 0.1$

$\Longrightarrow$ We obtain $\hat{\gamma}_{n,t,\alpha_j}$

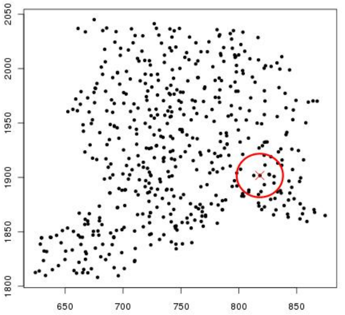
A leave one out cross validation procedure to select h_n and α_n : Step 2

We remove this station x_t

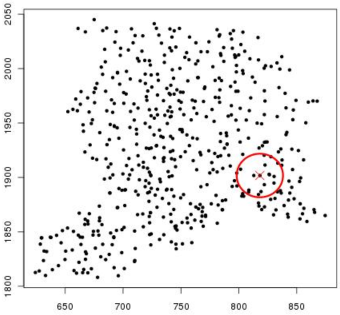


A leave one out cross validation procedure to select h_n and α_n : Step 2

We work in $B(x_t, h_i) \setminus \{x_t\}$



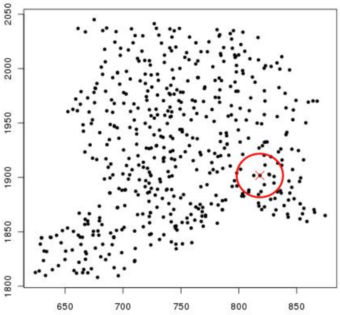
We work in $B(x_t, h_i) \setminus \{x_t\}$



- We estimate $\gamma(x) > 0$ using the conditional Hill estimator.
- It depends on α_j and on h_i.
- The h_i are chosen such that there is at least one station in $B(x_t, h_i) \setminus \{x_t\}$.

A leave one out cross validation procedure to select h_n and α_n : Step 2

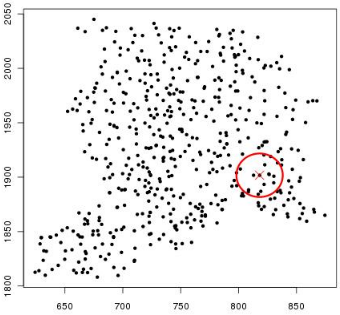
We work in $B(x_t, h_i) \setminus \{x_t\}$



- We estimate $\gamma(x) > 0$ using the conditional Hill estimator.
- It depends on α_j and on h_i.
- The h_i are chosen such that there is at least one station in $B(x_t, h_i) \setminus \{x_t\}$.

$\Longrightarrow$ We obtain $\hat{\gamma}_{n,h_i,\alpha_j}(x_t)$

We work in $B(x_t, h_i) \setminus \{x_t\}$



- We estimate $\gamma(x) > 0$ using the conditional Hill estimator.
- It depends on α_j and on h_i.
- The h_i are chosen such that there is at least one station in $B(x_t, h_i) \setminus \{x_t\}$.

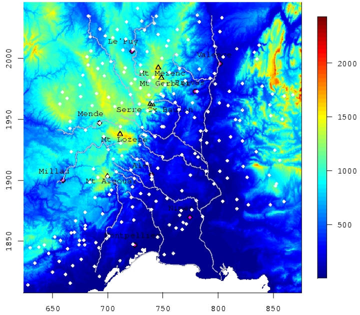
$\Longrightarrow$ We obtain $\hat{\gamma}_{n,h_i,\alpha_j}(x_t)$

$$(h_{emp}, \alpha_{emp}) = \underset{(h_i,\alpha_j)\in\mathcal{H}\times\mathcal{A}}{\arg\min} \ \text{median} \ \left\{\left(\hat{\gamma}_{n,t,\alpha_j} - \hat{\gamma}_{n,h_i,\alpha_j}(x_t)\right)^2, t \in \{1, \ldots, 523\}\right\}.$$

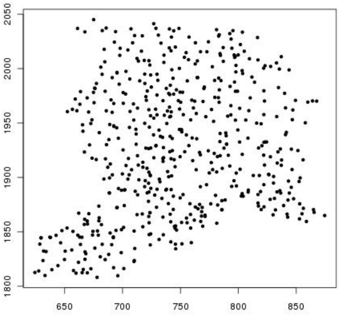
Applications to an hydrological data set

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



The 523 stations of interest

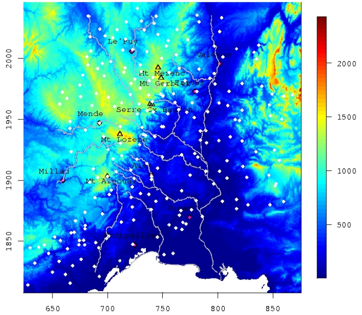


$\Longrightarrow$ Results of the procedure : $(h_{emp}, \alpha_{emp}) = (24, 1/(3 * 365.25))$.

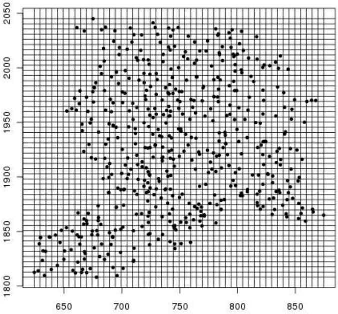
Applications to an hydrological data set

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



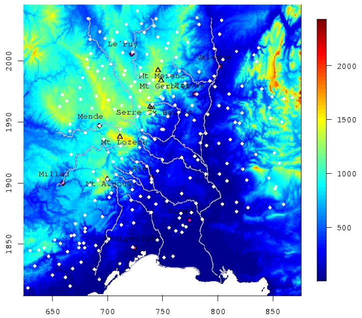
Grid 200 $\times$ 200 points



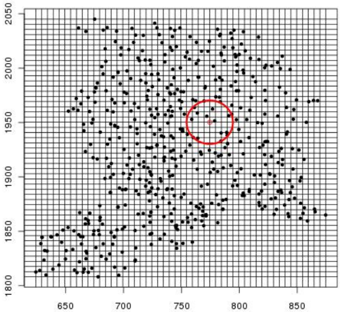
$\Longrightarrow$ Results of the procedure : $(h_{emp}, \alpha_{emp}) = (24, 1/(3 * 365.25))$.

Y : daily rainfall measured in mm. $X = \{\text{longitude, latitude, altitude}\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



Work in $B(x, h_n)$

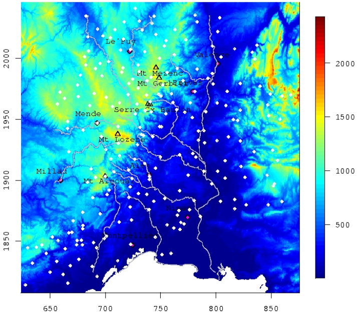


$\Longrightarrow$ Results of the procedure : $(h_{emp}, \alpha_{emp}) = (24, 1/(3 * 365.25))$.

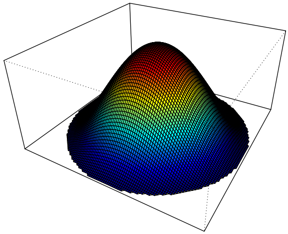
Applications to an hydrological data set

Y : daily rainfall measured in mm. $X = \{$longitude, latitude, altitude$\}$. 1958 $\Longrightarrow$ 2000.

The Cévennes-Vivarais region



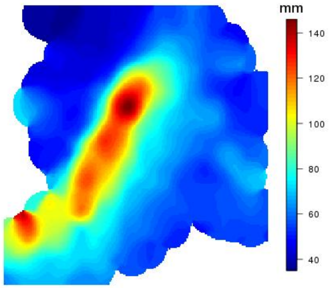
Bi-quadratic kernel



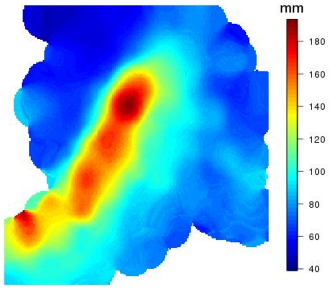
$\Longrightarrow$ Results of the procedure : $(h_{emp}, \alpha_{emp}) = (24, 1/(3 * 365.25))$.

Non extrapolated risk measures in the Cévennes-Vivarais region

$\widehat{\mathrm{RVaR}}_n(1/(3 * 365.25)|x)$

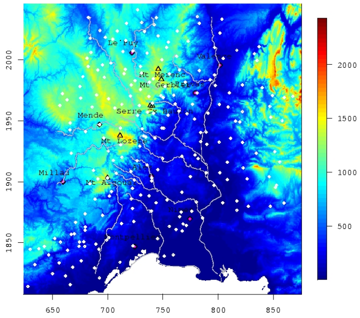


$\widehat{\mathrm{RCTE}}_n(1/(3 * 365.25)|x)$

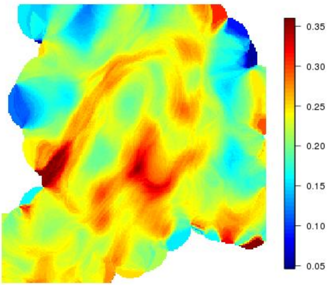


Risk measures in the Cévennes-Vivarais region

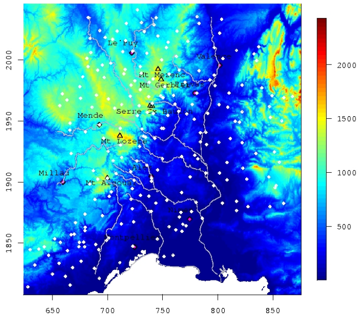
The Cévennes-Vivarais region



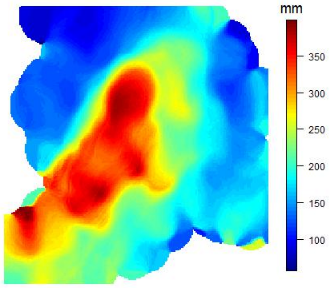
$\hat{\gamma}_{n,(1/(3*365.25))}(x)$



The Cévennes-Vivarais region

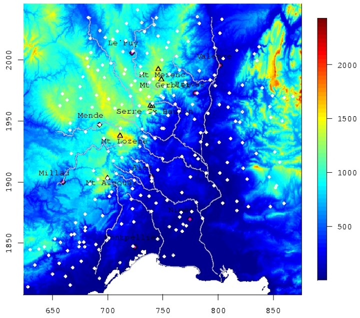


$\widehat{\mathrm{RVaR}}_n^W(1/(100 * 365.25)|x)$

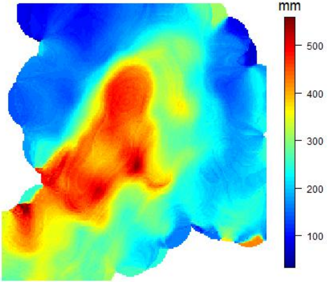


Risk measures in the Cévennes-Vivarais region

The Cévennes-Vivarais region



$\widehat{\mathrm{RCTE}}_n^W(1/(100 * 365.25)|x)$



Outline

Conclusions

Keys points

- Estimation of VaR, CTE, CVaR, and SP in the case of extreme losses for heavy-tailed distributions in the presence of a covariate.
- Extrapolate those risk measures to arbitrary small levels.
- Application to an hydrological data set.

Commentaries

+ Maps obtained are coherent according to hydrologists.
+ New tool for the prevention of risk in hydrology.
+ Theoretical properties similar to the univariate case (extreme or not) and with or without a covariate.

− Curse of dimensionality.

Perspectives

- Extend our work to all domains of attraction $\Longrightarrow \gamma(x) \in \mathbb{R}$

Further readings

This presentation is based on two research articles :

El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, vol. 41 (4), pp. 988–1012.

El Methni, J., Gardes, L. and Girard, S. (2015). Estimation de mesures de risque pour des pluies extrêmes dans la région Cévennes Vivarais, *La Houille Blanche*, vol. 4, pp. 46–51

Further readings

This presentation is based on two research articles :

El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, vol. 41 (4), pp. 988–1012.

El Methni, J., Gardes, L. and Girard, S. (2015). Estimation de mesures de risque pour des pluies extrêmes dans la région Cévennes Vivarais, *La Houille Blanche*, vol. 4, pp. 46–51

Thank you for your attention

Principals references

Cai, J. and Tan, K. (2007). Optimal retention for a stop-loss reinsurrance under the VaR and CTE risk measures, *Astin Bulletin*, **37**(1), 93–112.

Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, **20**, 311–333.

Embrechts, P., Kluppelberg, C. and Mikosh, T. (1997). Modelling Extremal Events, *Springer editions*.

Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index, *Journal of Multivariate Analysis*, **99**, 2368–2388.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, 1163–1174.

Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–42.

Valdez, E.A. (2005). Tail conditional variance for elliptically contoured distributions, *Belgian Actuarial Bulletin*, **5**, 26–36.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, **73**, 812–815.