# On blocks estimators

## for cluster inference of heavy-tailed time series

Gloria Buriticá

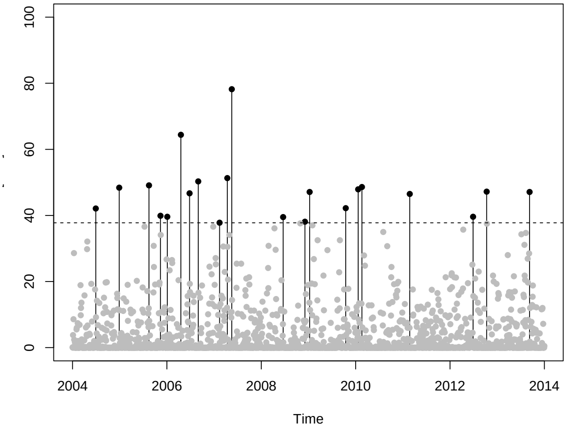
Joint work with O. Wintenberger (LPSM, Paris)

Research center for Statistics,
Université de Génève

May 16, 2024

UNIVERSITÉ DE GENÈVE

# Motivation



# Goal

- Model extremal temporal clustering of stationary time series.
- Infer cluster statistics using block methods.

**Key words**: Extreme value theory, cluster Poisson point process, blocks methods for cluster inference, extremal index.

# Table of contents

# Notation

We consider

- $(\mathbf{X}_t)$ stationary time series in $(\mathbb{R}^d, |\cdot|)$.

[1] Janssen (2019)

# Notation

We consider

- $(\mathbf{X}_t)$ stationary time series in $(\mathbb{R}^d, |\cdot|)$.
- $(\mathbf{X}_t)$ is regularly varying ($\mathrm{RV}_\alpha$): for all $h \geqslant 0$, $y > 1$,

$$\lim_{x \to +\infty} \mathbb{P}(|\mathbf{X}_0| > y\, x, \tfrac{\mathbf{X}_{[-h,h]}}{|\mathbf{X}_0|} \in \cdot \mid |\mathbf{X}_0| > x) \;=\; y^{-\alpha}\, \mathbb{P}(\Theta_{[-h,h]} \in \cdot).$$

[1] Janssen (2019)

# Notation

We consider

- $(\mathbf{X}_t)$ stationary time series in $(\mathbb{R}^d, |\cdot|)$.
- $(\mathbf{X}_t)$ is regularly varying ($\mathrm{RV}_\alpha$): for all $h \geqslant 0$, $y > 1$,

$$\lim_{x \to +\infty} \mathbb{P}\Big(|\mathbf{X}_0| > y\,x, \frac{\mathbf{X}_{[-h,h]}}{|\mathbf{X}_0|} \in \cdot \,\Big|\, |\mathbf{X}_0| > x\Big) \;=\; y^{-\alpha}\,\mathbb{P}(\Theta_{[-h,h]} \in \cdot).$$

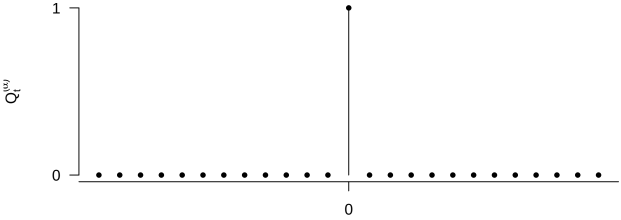
- $|\mathbf{\Theta}_t| \to 0$ as $t \to +\infty$ a.s.
- $\mathbf{Q} = \mathbf{\Theta}/\|\mathbf{\Theta}\|_\alpha$, where $\|\mathbf{x}\|_\alpha^\alpha = \sum_{t \in \mathbb{Z}} |\mathbf{x}_t|^\alpha$. [1]

[1] Janssen (2019)

# i.i.d. model

- $(\mathbf{X}_t)$ **i.i.d.**, $\mathbf{X}_1$ satisfies $\mathbf{RV}_\alpha$,

$$\mathbf{Q}_t = \mathbf{\Theta}_t = \mathbb{1}(t=0)\,\mathbf{\Theta}_0.$$

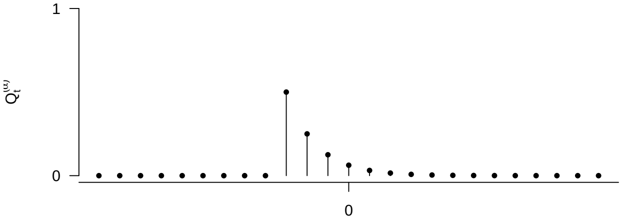


# Auto-regressive model

- $(X_t)$ a stationary **AR(1)**, $X_t = \varphi X_{t-1} + Z_t$ with $\varphi \in (0,1)$, and $(Z_t)$ i.i.d. satisfying $\mathbf{RV}_\alpha$,

$$Q_t^{(\alpha)} = \Theta_t / \|\Theta\|_\alpha = \varphi^t \Theta_0^Z \, \mathbb{1}(J + t \geqslant 0)\, (1 - \varphi^\alpha)^{1/\alpha},$$

$J$ independent of $\Theta_0^Z$, $\mathbb{P}(J = j) = (1 - \varphi^\alpha)\varphi^{j\alpha}, j \geqslant 0$.

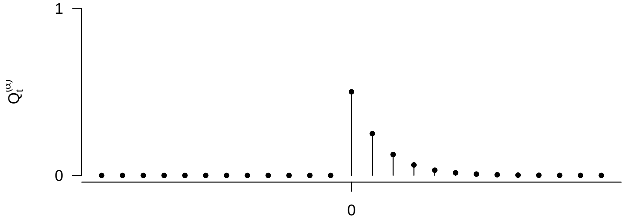


# Auto-regressive model

- $(X_t)$ a stationary **AR(1)**, $X_t = \varphi X_{t-1} + Z_t$ with $\varphi \in (0,1)$, and $(Z_t)$ i.i.d. satisfying $\mathbf{RV}_\alpha$,

$$Q_t^{(\alpha)} \quad = \quad \Theta_t / \|\Theta\|_\alpha \quad = \quad \varphi^t \Theta_0^Z \, \mathbb{1}(J + t \geqslant 0)\,(1 - \varphi^\alpha)^{1/\alpha},$$

$J$ independent of $\Theta_0^Z$, $\mathbb{P}(J = j) = (1 - \varphi^\alpha)\varphi^{j\alpha}, j \geqslant 0$.

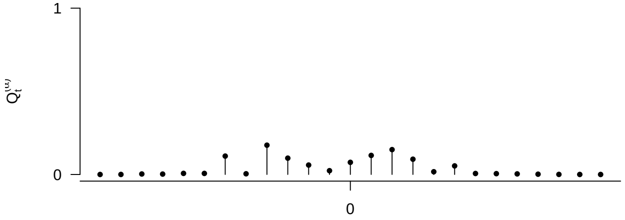


# Causal solution to SRE under Kesten-Goldie assumptions

- $(X_t)$ **causal solution to SRE**, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive i.i.d. and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \quad = \quad A_t \cdots A_1 \,, \quad t \geqslant 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.



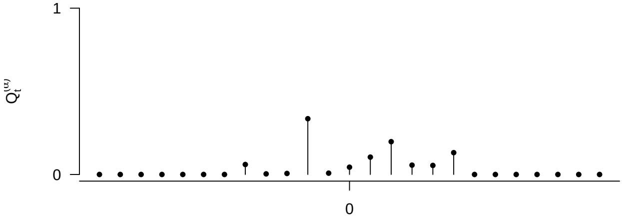

We take $A_t = e^{N_t - 1/2}$ such that $(N_t)$ is i.i.d. gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) [5] where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leqslant 0$.

# Causal solution to SRE under Kesten-Goldie assumptions

- $(X_t)$ **causal solution to SRE**, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive i.i.d. and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \quad = \quad A_t \cdots A_1\,, \quad t \geqslant 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.



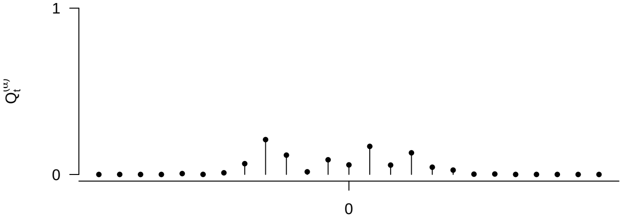

We take $A_t = e^{N_t - 1/2}$ such that $(N_t)$ is i.i.d. gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) [5] where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leqslant 0$.

# Causal solution to SRE under Kesten-Goldie assumptions

- $(X_t)$ **causal solution to SRE**, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive i.i.d. and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \quad = \quad A_t \cdots A_1 \,, \qquad t \geqslant 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.



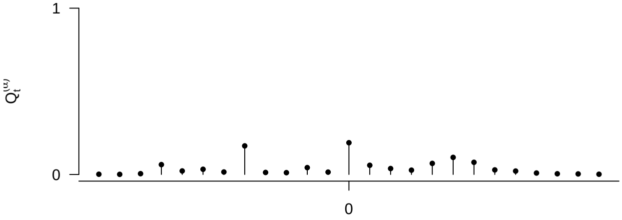

We take $A_t = e^{N_t - 1/2}$ such that $(N_t)$ is i.i.d. gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) [5] where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leqslant 0$.

# Causal solution to SRE under Kesten-Goldie assumptions

- $(X_t)$ **causal solution to SRE**, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive i.i.d. and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \quad = \quad A_t \cdots A_1 \,, \quad t \geqslant 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.




We take $A_t = e^{N_t - 1/2}$ such that $(N_t)$ is i.i.d. gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) [5] where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leqslant 0$.

# Cluster Poisson Point Process

**Theorem**[2] Buriticá, Meyer, Mikosch, Wintenberger (2021)

Assume $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$, **AC** and **MX**, then

$$N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1} X_i} \quad \to \quad N = \sum_{i=1}^{\infty} \sum_{j \in \mathbb{Z}} \varepsilon_{\Gamma_i^{-1/\alpha} \mathbf{Q}_{ij}} ,$$

in $\mathbb{R}_0$, where $n\mathbb{P}(|\mathbf{X}_1| > a_n) \to 1$,

• $\sum_{j\in\mathbb{Z}} \varepsilon_{\mathbf{Q}_{ij}}$, $i = 1, 2, \ldots$, is an iid sequence of point processes with state space $\mathbb{R}^d$ with generic element $\mathbf{Q}_i = (\mathbf{Q}_{ij})_{j\in\mathbb{Z}}$,

$$\mathbf{Q} = \left( \frac{\Theta_j}{\|\Theta\|_\alpha} \right)_{j\in\mathbb{Z}} .$$

• $(\Gamma_i)$ are points of a unit rate homogeneous Poisson process on $(0, \infty)$.
• $(\Gamma_i)$ and $(\mathbf{Q}_i)_{i=1,2,\ldots}$ are independent.

[2] Davis and Hsing (1995); see [1, 3]

# Cluster inference

## Blocks method

**Aim:** compute cluster statistic: $\mathbb{E}[f(Y\mathbf{Q})]$.

**So far:** $N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}X_i} \to N = \sum_{i=1}^{\infty} \sum_{j\in\mathbb{Z}} \varepsilon_{\Gamma_i^{-1/\alpha}\mathbf{Q}_{ij}}$ .

$$\mathbf{X}_{[1:n]} = \big(\underbrace{\mathbf{X}_{[1:b_n]}}_{:=\mathcal{B}_1}, \underbrace{\mathbf{X}_{b_n+[1:b_n]}}_{:=\mathcal{B}_2}, \dots, \underbrace{\mathbf{X}_{[n-b_n+1:b_n]}}_{:=\mathcal{B}_{m_n}}\big),$$

- select $k$ extremal blocks $\mathcal{B}_{(1)}, \dots, \mathcal{B}_{(k)}$,
- average $\frac{1}{k}\sum_{t=1}^{k} f(\mathcal{B}_{(t)}/a_n)$,
- e.g. count threshold exceedances in a block $f : (\mathbf{x}_t) \mapsto \sum \mathbb{1}(|\mathbf{x}_t| > 1)$.

# Cluster inference

## Blocks method

**Aim:** compute cluster statistic: $\mathbb{E}[f(Y\mathbf{Q})]$.

**So far:** $N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}X_i} \to N = \sum_{i=1}^{\infty} \sum_{j \in \mathbb{Z}} \varepsilon_{\Gamma_i^{-1/\alpha}\mathbf{Q}_{ij}}$ .

$$\mathbf{X}_{[1:n]} = \big(\underbrace{\mathbf{X}_{[1:b_n]}}_{:=\mathcal{B}_1}, \underbrace{\mathbf{X}_{b_n+[1:b_n]}}_{:=\mathcal{B}_2}, \ldots, \underbrace{\mathbf{X}_{[n-b_n+1:b_n]}}_{:=\mathcal{B}_{m_n}}\big),$$

- select $k$ extremal blocks $\mathcal{B}_{(1)}, \ldots, \mathcal{B}_{(k)}$,
- average $\frac{1}{k}\sum_{t=1}^{k} f(\mathcal{B}_{(t)}/a_n)$,
- e.g. count threshold exceedances in a block $f : (\mathbf{x}_t) \mapsto \sum \mathbb{1}(|\mathbf{x}_t| > 1)$.

**(Q)** How to choose those $k$ extremal blocks ?

# Large deviations of $\ell^p$-blocks

**Theorem** Buriticá, Mikosch, Wintenberger (2023)

Assume $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$, and $(x_n)$ satisfies $\mathbf{AC}(x_n)$, $\mathbf{CS}_\alpha(x_n)$, and $n\mathbb{P}(|\mathbf{X}_1| > x_{b_n}) \to 0$. Then,

$$\mathbb{P}\Big(\|\mathbf{X}_{[1,b_n]}\|_\alpha > y\, x_{b_n}, \frac{\mathbf{X}_{[1,b_n]}}{\|\mathbf{X}_{[1,b_n]}\|_\alpha} \in \cdot \,\Big|\, \|\mathbf{X}_{[1,b_n]}\|_\alpha > x_{b_n}\Big) \to y^{-\alpha}\mathbb{P}(\mathbf{Q} \in \cdot)$$

# Large deviations of $\ell^p$-blocks

**Theorem** Buriticá, Mikosch, Wintenberger (2023)

Assume $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$, and $(x_n)$ satisfies $\mathbf{AC}(x_n)$, $\mathbf{CS}_p(x_n)$, and $n\mathbb{P}(|\mathbf{X}_1| > x_{b_n}) \to 0$. Then, for $p > 0$,

$$\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > y\, x_{b_n}, \frac{\mathbf{X}_{[1,b_n]}}{\|\mathbf{X}_{[1,b_n]}\|_p} \in \cdot \mid \|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n}) \to y^{-\alpha}\mathbb{P}(\mathbf{Q}^{(p)} \in \cdot)$$

and

$$\lim_{n\to\infty} \mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n})/(b_n\mathbb{P}(|\mathbf{X}_0| > x_{b_n})) = c(p) = \mathbb{E}[\|\mathbf{Q}\|_p^\alpha],$$

$\theta_{|\mathbf{X}|} = c(\infty) \leqslant c(p) \leqslant c(\alpha) = 1$, for $p \in (\alpha, \infty)$.

# Blocks method

To infer $\mathbb{E}[f(Y\mathbf{Q}^{(p)})]$,

$$\widehat{f^{\mathbf{Q}}}(p) \; := \; \frac{1}{k}\sum_{t=1}^{m} f\big(\mathcal{B}_t/\|\mathcal{B}_t\|_{p,(k+1)}\big)\mathbb{1}(\|\mathcal{B}_t\|_p > \|\mathcal{B}_t\|_{p,(k+1)}\big),$$

The same quantity $f^{\mathbf{Q}}$ can be estimated using different pairs $p', f'$ as

$$\begin{aligned} f^{\mathbf{Q}} \; = \; \mathbb{E}[f(Y\mathbf{Q})] \; &= \; \frac{\mathbb{E}[\|\mathbf{Q}^{(p')}\|_\alpha^\alpha f(Y\mathbf{Q}^{(p')}/\|\mathbf{Q}^{(p')}\|_p)]}{\mathbb{E}[\|\mathbf{Q}^{(p')}\|_\alpha^\alpha]} \\ &= \; c(p')\mathbb{E}[f'(Y\mathbf{Q}^{(p')})]. \\ c(p') \; &= \; \mathbb{E}[\|\mathbf{Q}\|_{p'}^\alpha]. \end{aligned}$$

# Asymptotic normality

We propose to estimate the statistic $f_\alpha^{\mathbf{Q}} = \mathbb{E}[f_\alpha(Y\mathbf{Q})]$ by

$$\widehat{f_\alpha^{\mathbf{Q}}} := \frac{1}{k}\sum_{t=1}^{m} f_{\widehat{\alpha}}\big(\mathcal{B}_t/\|\mathcal{B}_t\|_{\widehat{\alpha},(k+1)}\big)\mathbb{1}(\|\mathcal{B}_t\|_{\widehat{\alpha}} > \|\mathcal{B}_t\|_{\widehat{\alpha},(k+1)}),$$

## Theorem Buriticá, Wintenberger (2024)

Under moment, bias and mixing conditions. There exists $k = k_n \to \infty$, $m/k \to \infty$, such that for suitable $f_\alpha : \ell^\alpha \to \mathbb{R}$.

$$\sqrt{k_n}\left(\widehat{f_{\widehat{\alpha}}^{\mathbf{Q}}} - f_\alpha^{\mathbf{Q}}\right) \xrightarrow{d} \mathcal{N}\Big(0, Var\big(f_\alpha(Y\mathbf{Q})\big) + \kappa^2\sigma_\alpha^2\Big), \quad n \to \infty,$$

and $k_n/k_n' \to \kappa$, with $\kappa \geqslant 0$, $Y$ independent of $\mathbf{Q}$, and $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$.

# Tail index Hill estimator

$$\frac{1}{\widehat{\alpha}^n} := \frac{1}{\widehat{\alpha}^n(k')} := \frac{1}{k'} \sum_{t=1}^{n} \log(|\mathbf{X}_t|/|\mathbf{X}|_{(k'+1)}),$$

where $|\mathbf{X}|_{(1)} \geqslant |\mathbf{X}|_{(2)} \geqslant \cdots \geqslant |\mathbf{X}|_{(n)}$, and $k' = k'(n)$ is a tuning sequence for the Hill estimator satisfying $k' \to \infty$, $n/k' \to \infty$, as $n \to \infty$.

# Remarks

- $\ell^\infty$-block methods for cluster inference studied in Drees and Rootzén (2010) [4] with high threshold $x : |\mathbf{X}_t| > x_{b_n}$. We extend the analysis to $\ell^{\widehat{\alpha}}$-cluster inference selecting the blocks whose $\ell^{\widehat{\alpha}}$-norm exceed the high threshold $x$: $\|\mathcal{B}_t\|_{\widehat{\alpha}} > x_{b_n}$.

# Remarks

- $\ell^\infty$-block methods for cluster inference studied in Drees and Rootzén (2010) [4] with high threshold $x : |\mathbf{X}_t| > x_{b_n}$. We extend the analysis to $\ell^{\widehat{\alpha}}$-cluster inference selecting the blocks whose $\ell^{\widehat{\alpha}}$-norm exceed the high threshold $x$: $\|\mathcal{B}_t\|_{\widehat{\alpha}} > x_{b_n}$.
- We promote the use of order statistics of $\ell^{\widehat{\alpha}}$-norm blocks such that

$$\|\mathcal{B}\|_{\widehat{\alpha},(k+1)}/x_{b_n} \overset{\mathbb{P}}{\to} 1.$$

  where $k_n = \lceil m_n \mathbb{P}(\|\mathcal{B}_{1,b_n}\|_\alpha > x_{b_n}) \rceil$. In this way $k_n$ points to the bias-variance trade-off in extreme value statistics.

# Remarks

- $\ell^\infty$-block methods for cluster inference studied in Drees and Rootzén (2010) [4] with high threshold $x : |\mathbf{X}_t| > x_{b_n}$. We extend the analysis to $\ell^{\widehat{\alpha}}$-cluster inference selecting the blocks whose $\ell^{\widehat{\alpha}}$-norm exceed the high threshold $x$: $\|\mathcal{B}_t\|_{\widehat{\alpha}} > x_{b_n}$.
- We promote the use of order statistics of $\ell^{\widehat{\alpha}}$-norm blocks such that

$$\|\mathcal{B}\|_{\widehat{\alpha},(k+1)}/x_{b_n} \xrightarrow{\mathbb{P}} 1.$$

  where $k_n = \lceil m_n \mathbb{P}(\|\mathcal{B}_{1,b_n}\|_\alpha > x_{b_n}) \rceil$. In this way $k_n$ points to the bias-variance trade-off in extreme value statistics.
- It is common to take $k_n/k_n' \to 0$. In this case the asymptotic variance simplifies to $Var\big(f_\alpha(Y\mathbf{Q})\big)$.

# Number of extreme blocks

Denote $k_n(p) = \lceil m_n \mathbb{P}(\|\mathcal{B}_{1,b_n}\|_p > x_{b_n}) \rceil$ the extremal $\ell^p$-blocks, for a sequence of levels $(x_n)$ satisfying **AC**, $\mathbf{CS}_p$.
For i.i.d. sequence $k_n = \lceil n\mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \rceil \sim k_n(\infty) \sim k_n(p) \sim k_n(\alpha)$ exceedances.

## Heuristic on the number of extreme blocks:

$$k_n(p) \sim m_n \mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n}) \sim c(p) n \mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \sim c(p) k_n \,,$$
$$k_n(\alpha) \sim m_n \mathbb{P}(\|\mathcal{B}_1\|_\alpha > x_{b_n}) \sim n \mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \sim k_n \,,$$

Assuming also $\mathbf{CS}_\alpha$, $\alpha$-cluster inference is justified. In this case the tuning parameter $k_n$ does not dependent on the underlying time dependencies.

# Extremal index

### Maximum domain of attraction

There exists $(a_n)$ such that

$$(\mathbb{P}(|\mathbf{X}_1| \leqslant x a_n))^n \quad \to \quad G(x) := \mathbb{P}((\Gamma_1)^{-1/\alpha} \leqslant x), \quad n \to \infty,$$

where $G(x) = \exp\{-x^{-\alpha}\}$, for $\alpha > 0$, $x > 1$, $n\mathbb{P}(X_1 > a_n) \to 1$.

# Extremal index

### Maximum domain of attraction

There exists $(a_n)$ such that

$$(\mathbb{P}(|\mathbf{X}_1| \leqslant x a_n))^n \quad \to \quad G(x) := \mathbb{P}((\Gamma_1)^{-1/\alpha} \leqslant x), \quad n \to \infty,$$

where $G(x) = \exp\{-x^{-\alpha}\}$, for $\alpha > 0$, $x > 1$, $n\mathbb{P}(X_1 > a_n) \to 1$.

(Leadbetter 1983) there exists $\theta \in (0, 1]$ such that

$$\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_\infty \leqslant x\, a_n) \quad \to \quad (G(x))^{\theta}, \quad n \to \infty.$$

$$\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_\infty \leqslant x\, a_n) \quad \to \quad (G(x))^{\theta}, \quad n \to \infty.$$

$$\Longrightarrow \ (\mathbb{P}(X_1 \leqslant x\, a_{b_n}))^{\theta\, b_n} \sim \mathbb{P}(\|\mathcal{B}_{1,b_n}\|_\infty \leqslant x\, a_{b_n}) \sim G(x), \quad n \to \infty.$$

for $x_{b_n} = x\, a_{b_n}$ with $\|\mathcal{B}_{1,b_n}/x_{b_n}\|_\infty \xrightarrow{\mathbb{P}} 0$, as $\log(1-x)/x \to 0$ as $x \to 0$,

$$\frac{\mathbb{P}(\|\mathcal{B}_{1,b_n}\|_\infty > x_{b_n})}{b_n \mathbb{P}(|\mathbf{X}_1| > x_{b_n})} \to \theta_{|\mathbf{X}|}, \quad n \to \infty.$$

$\Longrightarrow$ Blocks estimator based in (Hsing 1991):

$$\widehat{\theta}^{B}_{|\mathbf{X}|} \ = \ \frac{1}{k_n b_n} \sum_{t=1}^{m_n} \mathbb{1}(\|\mathcal{B}_t\|_\infty > |\mathbf{X}|_{(k+1)})\,.$$

# Example: extremal index

## Cluster-based extremal index inference

For example, if $f_\alpha : (\mathbf{x}_t) \mapsto \|(\mathbf{x}_t)\|_\infty^\alpha / \|(\mathbf{x}_t)\|_\alpha^\alpha$, then,

$$\theta_{|\mathbf{X}|} \quad = \quad \mathbb{E}[\|\mathbf{Q}\|_\infty^\alpha].$$

$\Longrightarrow$ Estimator of the extremal index based on extremal $\ell^\alpha$-blocks.

$$\widehat{\theta}_{|\mathbf{X}|} \quad = \quad \frac{1}{k} \sum_{t=1}^{m} \frac{\|\mathcal{B}_t\|_\infty^{\widehat{\alpha}}}{\|\mathcal{B}_t\|_{\widehat{\alpha}}^{\widehat{\alpha}}} \mathbb{1}(\|\mathcal{B}_t\|_{\widehat{\alpha}} > \|\mathcal{B}\|_{\widehat{\alpha},(k+1)}),$$

# Causal linear model

**Theorem** Buriticá, Wintenberger (2024)[3]

Let $(\mathbf{X}_t) = \sum_{t \geqslant 0} \varphi_j \mathbf{Z}_{t-j}$, such that $(\mathbf{Z}_t)$ is i.i.d. and satisfies $\mathbf{RV}_\alpha$. For $\rho > 0$, assume $\varphi_t = O(t^{-\rho})$. Assume

1) $f_p$ is bounded and $\rho > 3 + 2/\alpha$.
2) there exist $\kappa' > 0$, and $(k_n)$ satisfying $k_n = O(n\, b_n^{-\kappa'-1})$,

Then, if a bias assumption holds and $k/k' \to 0$,

$$\sqrt{k}(\widehat{f^{\mathbf{Q}}_{\widehat{\alpha}}} - f^{\mathbf{Q}}_\alpha) \quad \xrightarrow{d} \quad \mathcal{N}\big(0, Var(\, f_\alpha(\, Y\mathbf{Q})\,)\big), \quad n \to \infty.$$

In particular, the $\alpha$-cluster based estimators for the extremal index has null asymptotic variance!

[3]see [2]

# Implementation extremal index

If $f_\alpha : (\mathbf{x}_t) \mapsto \|(\mathbf{x}_t)\|_\infty^\alpha / \|(\mathbf{x}_t)\|_\alpha^\alpha$, then for $p = \alpha$,

$$\theta_{|\mathbf{X}|} \quad = \quad \mathbb{E}[\|\mathbf{Q}\|_\infty^\alpha].$$

$\implies$ Estimator of the extremal index based on extremal $\ell^\alpha$-blocks.

$$\widehat{\theta}_{|\mathbf{X}|} \quad = \quad \frac{1}{k} \sum_{t=1}^{m} \frac{\|\mathcal{B}_t\|_\infty^{\widehat{\alpha}}}{\|\mathcal{B}_t\|_{\widehat{\alpha}}^{\widehat{\alpha}}} \mathbb{1}(\|\mathcal{B}_t\|_{\widehat{\alpha}} > \|\mathcal{B}\|_{\widehat{\alpha},(k+1)}),$$

For the autoregressive process AR(1): $\mathbb{V}ar(f_\alpha(\mathbf{Q})) = 0$.

# Implementation extremal index

Blocks estimator based in (Hsing 1991):

$$\widehat{\theta}^B_{|\mathbf{X}|} = \frac{1}{k_n b_n} \sum_{t=1}^{m_n} \mathbb{1}(\|\mathcal{B}_t\|_\infty > |\mathbf{X}|_{(k+1)}) .$$

Direct computations from Example 10.4.2 in (Kulik and Soulier 2020) yield

$$\sqrt{k}(\widehat{\theta}^B_{|\mathbf{X}|} - \theta_{|\mathbf{X}|}) \xrightarrow{d} \mathcal{N}(0, \sigma^2_\theta), \quad n \to \infty,$$

where $\sigma^2_\theta \in [0, \infty)$, and

$$\sigma^2_\theta = \theta^2_{|\mathbf{X}|} \sum_{j \in \mathbb{Z}} \sum_{t \in \mathbb{Z}} \mathbb{E}[|\mathbf{Q}^{(\alpha)}_{j+t}|^\alpha \wedge |\mathbf{Q}^{(\alpha)}_t|^\alpha] - \theta_{|\mathbf{X}|}.$$

For the autoregressive process AR(1): $\sigma^2_\theta = 1 - \theta_{|\mathbf{X}|} > 0$.

# Case study

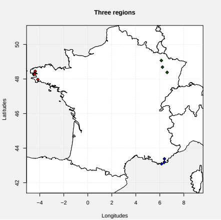
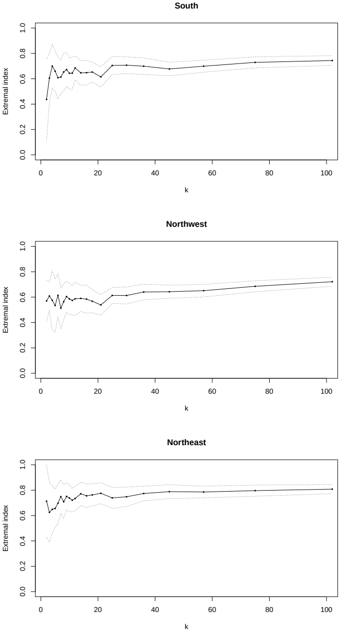



Figure: Location of weather stations in France.



# Take away

- How to choose extreme blocks plays an important role in inference.
- Estimation of the tail-index can help to stabilize the method.
- $\ell^{\widehat{\alpha}}$-blocks inference yields robust result.

# Questions?

Thank you for your attention!

# Further perspectives

- How to define extremal directions of extremes in space and time?
- Different extremes episodes can have different causes, identify extreme with comparable features could help environmental scientists detect and characterize the natural phenomena leading to an extreme event.

# References I


G. Buriticá, N. Meyer, T. Mikosch, and O. Wintenberger.
Some variations on the extremal index.
*Zap. Nauchn. Semin. POMI.*, 30:52–77, 2021.
To be translated in J.Math.Sci. (Springer).

G. Buriticá and O. Wintenberger.
On the asymptotics of extremal lp-blocks cluster inference.
*arXiv:2212.13521*, 2024.

R. A. Davis and T. Hsing.
Point process and partial sum convergence of weakly dependent random variables with infinite variance.
*The Annals of Probability*, 23:879–917, 1995.

H. Drees and H. Rootzén.
Limit theorems for empirical processes of cluster functionals.
*Ann. Stat.*, 38:2145–2186, 2010.

A. Janssen and J. Segers.
Markov tail chains.
*Journal of Applied Probability*, 51:1133–1153, 2014.

# Simulation setup

$$\widehat{\theta}_{|\mathbf{X}|,\alpha} = k_n^{-1}\textstyle\sum_{t=1}^{m_n} \frac{\|\mathcal{B}_t\|_\infty^\alpha}{\|\mathcal{B}_t\|_\alpha^\alpha} \mathbb{1}(\|\mathcal{B}_t\|_\alpha > \|\mathcal{B}\|_{\alpha,(k)}), \tag{1}$$

$$\widehat{\theta}^{B}_{|\mathbf{X}|} = \frac{1}{k_n b_n} \sum_{t=1}^{m_n} \mathbb{1}(\|\mathcal{B}_t\|_\infty > |\mathbf{X}|_{(k+1)}) . \tag{2}$$

- We simulate 1 000 AR(1) trajectories $(X_t)_{t=1,\dots,n}$, $X_t = \varphi X_{t-1} + Z_t$, for $n = 8\,000, 3\,000, 1\,000$.
- We fix $k = k_n = n/b_n^2$ and we use that $k_n(p) = o(n/b^{1+\kappa'})$,
- In this setting,

$$0 = Var(f_\alpha(YQ^{(\alpha)})) < \sigma_\theta^2 = 1 - \theta_{|\mathbf{X}|}.$$

# Extremal index comparison

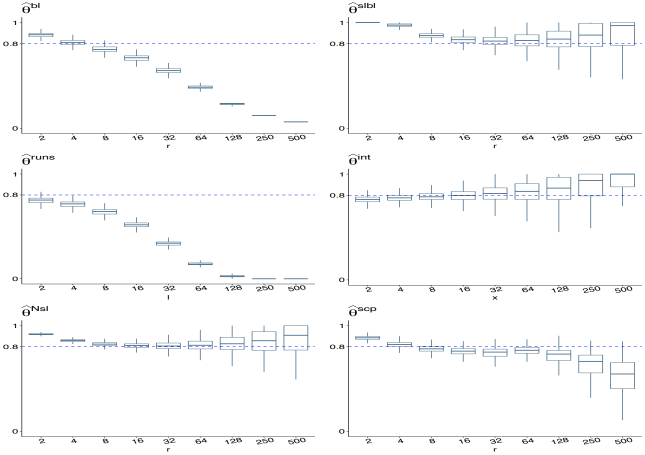



Figure: Boxplots based on 1 000 simulations of $(X_t)_{t=1,\dots,n}$ with $n = 5\,000$ for the estimation of $\theta_{|X|} = 0.8$ in the $AR(1)$ model with $\varphi = 0.2$ and iid student(1) noise.