Functional peaks-over-threshold analysis with an application to extreme European winter storms

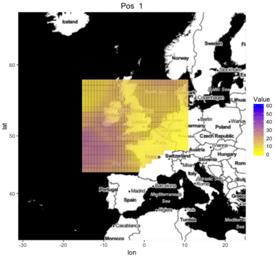
Raphaël de Fondeville

Chair of Statistics, EPFL
BIRS Workshop, Banff

December 2017

Motivation

3*s* maximum wind gust of the storm Lothar during winter 1999



- 169 km/h maximum observed windspeed in Paris (Parc Montsouris).
- Estimated loss around 8 billion dollars.

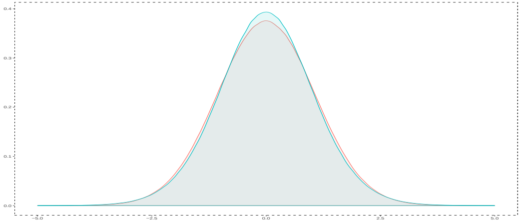
Motivation

- Classical techniques for risk estimation rely on historical catalogues and climate models:

  $\Rightarrow$ Cannot generate completely new extreme events.

- Aim to develop a windstorm generator producing storms with
  - unobserved intensities, i.e., extrapolation above known levels;
  - unobserved patterns, i.e., new storm tracks and shapes.

Motivation: geostatistical tools?



- Gaussian (red) and t_{20} (blue) density functions matched to have probabilities 0.05 for $|X| > 1.96$.
- Ratio of t_{20}/Gaussian probabilities for $|X| > x$:

x	2	3	4	5	6	7
Ratio of probabilities	1.01	1.7	6.1	58	1589	$1.7e5$

- Gaussian distribution has a quick tale decay which may strongly underestimate rare events:

$\Rightarrow$ Not suitable for extrapolation!

Motivation: geostatistical tools?

- For a threshold $u > 0$ and a bivariate vector with Fréchet margins and Gaussian copula,

$$\Pr(X_1 > u \mid X_2 > u) \sim C \times u^{-(1-\rho)/(\rho+1)} (\log u)^{-\rho/(1+\rho)},$$

and

$$\lim_{u \to \infty} \Pr(X_1 > u \mid X_2 > u) = 0.$$

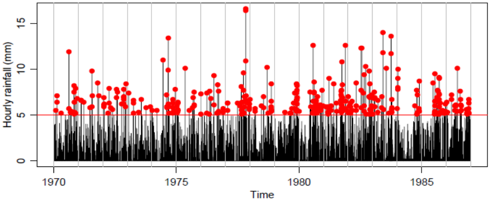
where $-1 < \rho < 1$ is the correlation coefficient.

- With a Gaussian spatial model, the extent of an extreme event decreases as u increases:

$\Rightarrow$ Strength of dependence should not depend on the intensity.

Extreme value theory (EVT)

- Extreme value theory describes the tail of the distribution.
- Historically it was developped for "block maxima", i.e., to model annual/monthly maxima with the Genralized Extreme Value (GEV) distribution.
- Max-stable processes, the functional equivalent of GEV, are mathematically very complex and thus limited application to few dozens of locations.
- To model single events, an alternative is the peaks-over-threshold analysis.



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Peaks-over-threshold models

For any random variable X, there exist sequences $a_n > 0$ and b_n such that

$$\left.\begin{array}{l} n\Pr\left[\frac{\{X-b_n\}_+}{a_n} > x\right] \\ n\Pr\left[\frac{\{b_n-X\}_+}{a_n} > x\right] \end{array}\right\} \to \nu_\xi(x), \quad n \to \infty,$$

and ν_ξ is either degenerate or

$$\nu_\xi(x) = \begin{cases} \left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}, & 1+\xi(x-\mu)/\sigma \geq 0, \quad \xi \neq 0; \\ \exp\left(-\frac{x-\mu}{\sigma}\right), & x \geq 0, \quad \xi = 0. \end{cases}$$

with, $\mu \in \mathbb{R}$, $\sigma > 0$.

ξ, the tail index, determines the strength of the tail and its support:

- $\xi > 0$: Fréchet type with $x \geq \mu$,
- $\xi = 0$: Gumbel type with $x \geq \mu$,
- $\xi < 0$: Weilbull type with $x \in (\mu; \mu - \sigma/1/\xi)$;

Univariate results for threshold exceedances

For a large enough threshold $u < \inf\{x : F(x) = 1\}$, we can use the approximation

$$\Pr(X - u > x \mid X > u) \approx \begin{cases} (1 + \xi x/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x/\sigma), & \xi = 0, \end{cases}$$

where $\sigma = \sigma(u) > 0$ and $a_+ = \max(a, 0)$:

$\Rightarrow$ The conditional distributions of exceedances over a high threshold can be approximated by a GP distribution.

(Generalized) functional regular variation

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

- Let $\{X(s)\}_{s \in S}$ be a stochastic process with sample paths in the space of continuous functions $C(S)$, where $S \subset \mathbb{R}^d$.
- Suppose there exist $\xi \in \mathbb{R}$, sequences $a_n > 0$ and b_n with $\lim_{n\to\infty} a_n(s) = \infty$ for all $s \in S$, such that

$$\left.\begin{array}{l} n \Pr\left[\left\{1 + \xi\left(\frac{X - b_n}{a_n}\right)\right\}^{1/\xi} \in \cdot\right] \\ n \Pr\left\{\exp\left(\frac{X - b_n}{a_n}\right) \in \cdot\right\} \end{array}\right\} \to \Lambda(\cdot), \quad n \to \infty. \tag{1}$$

- Λ is a measure on $C_+(S) \setminus \{0\}$ satsifying

$$\Lambda\{x \in tA\} = t^{-1}\Lambda\{x \in A\}, \quad t > 0, \quad A \in C(S) \setminus \{0\}.$$

- Condition (1), which we write $X \in \mathrm{GRV}(\Lambda)$, is a form of functional regular variation (Hult and Lindskog, 2005).

Risk functional and r-exceedances

- For a monotonic increasing functional r, an r-exceedance over a threshold $u \geqslant 0$ is an event $\{\mathrm{r}(x) \geqslant u\}$.
- r is called a risk functional. Common examples are
  - $\sup_{s \in S} X(s)$ for events where X exceeds a threshold at least one location;
  - $\sum_{t=1}^{T} \int_S X_t(s) ds$ for spatio-temporal accumulation;
  - $\sqrt{\int_S X(s)^2 ds}$ when the risk is determined by the energy inside a system;
  - $X(s_0)$, with $s_0 \in S$ for risks at a specific location, for instance a dam or a power plant.
- For simplicity of exposure, we now further suppose that r is linear.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Theorem (de Fondeville and Davison, 2018)

Let r be a risk functional and let $X \in \mathrm{GRV}(\Lambda)$. Then there exist $\xi \in \mathbb{R}$ and a measure σ_{ang} on

$$\mathcal{S}_{\mathrm{ang}} = \{x \in C(S) : \|x\|_1 = 1\},$$

such that for any $\mathcal{W} \subset \mathcal{S}_{\mathrm{ang}}$, and $\rho \geqslant 0$,

$$n\Pr\left[\frac{\mathrm{r}(\mathrm{X}) - \mathrm{r}(\mathrm{b_n})}{\mathrm{r}(\mathrm{a_n})} > \rho, \frac{\mathrm{X} - \mathrm{r}(\mathrm{b_n})}{\|\mathrm{X} - \mathrm{r}(\mathrm{b_n})\|_{\mathrm{ang}}} \in \mathcal{W}\right] \to \left(1 + \xi \frac{\mathrm{x} - \mu}{\sigma}\right)^{-1/\xi} \sigma_{\mathrm{ang}}(\mathcal{W}),$$

as $n \to \infty$, for $\xi \neq 0$, and

$$n\Pr\left[\frac{\mathrm{r}(\mathrm{X}) - \mathrm{r}(\mathrm{b_n})}{\mathrm{r}(\mathrm{a_n})} \geqslant \rho, \exp\frac{\mathrm{X} - \mathrm{r}(\mathrm{X})}{\mathrm{r}(\mathrm{a_n})} \in \mathcal{W}\right] \to \exp\left(-\frac{\mathrm{x} - \mu}{\sigma}\right) \sigma_{\mathrm{ang}}(\mathcal{W}),$$

for $\xi = 0$.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Generalized r-Pareto process (de Fondeville and Davison, 2018)

A generalized r-Pareto process P is defined by

$$P = \begin{cases} R\frac{W}{\mathrm{r}(W)}, & \xi \neq 0, \\ R + \log W - \mathrm{r}(\log W), & \xi = 0, \end{cases} \tag{2}$$

where

- R is a univariate generalized Pareto variable with tail parameter ξ, and distribution function

$$\Pr(\mathrm{R} > \rho) = \left\{1 + \xi \frac{\rho - \mathrm{u}}{\sigma}\right\}^{-1/\xi}, \quad \rho \geqslant \mathrm{u} \geqslant 0,$$

  with $\sigma > 0$;
- W is the stochastic process

$$W = A\{R_1 Q\}^{\xi} + B.$$

  where $A > 0, B \in C(S)$ with $\mathrm{r}(A) = 1$ and $\mathrm{r}(B) = 0$, R_1 is a unit Pareto distribution and Q is a stochastic process on $\{x \in C_+(S) : ||x||_1 = 1\}$ with probability measure σ_{ang}.

Properties of generalized r-Pareto processes

- r-Pareto processes are the only possible limits of rescaled threshold exceedances for a regularly varying stochastic process. This means for a large enough threshold $u > 0$,

$$\Pr\left(\mathrm{X} - \mathrm{u} \in \cdot \mid \mathrm{r}(\mathrm{X}) > \mathrm{u}\right) \approx \Pr\left(\mathrm{P} \in \cdot\right).$$

- The r-exceedance distribution of P is

$$\Pr\left\{\mathrm{r}(P) \geqslant \rho\right\} = \left\{1 + \xi \frac{\rho - u}{\sigma}\right\}^{-1/\xi}, \quad \rho \geqslant u.$$

- The generalized r-Pareto process has generalized Pareto marignals above a sufficiently high threshold $u_0 \geqslant 0$:

$$\Pr\left\{P(s_0) \geqslant \rho \mid P(s_0) \geqslant u_0\right\} = \left\{1 + \xi \frac{\rho - \mu(s_0)}{\sigma(u_0)}\right\}^{-1/\xi}, \quad \rho \geqslant u_0,$$

with $\sigma(u_0) > 0$ and $\mu(s_0) \in \mathbb{R}$.

Density function of exceedances

- In practice, choose a high threshold vector $u > 0$ such that the density function of the r-exceedances f_θ^{r} can be approximated by its limit

$$f_\theta^{\mathrm{r}}(x) \approx \frac{\lambda_\theta^{\mathrm{r}}(x)}{\Lambda_\theta\{A_{\mathrm{r}}(u)\}}, \quad x \in A_{\mathrm{r}}(u),$$

  with $A_{\mathrm{r}}(u) = \{x \in C(S) : \mathrm{r}(x) \geqslant u\}$ and where

$$\Lambda_\theta\{A_{\mathrm{r}}(u)\} = \int_{A_{\mathrm{r}}(u)} \lambda_\theta^{\mathrm{r}}(x) dx, \quad u \geqslant 0.$$

- For most models, the limiting measure Λ and its partial derivatives are known in Cartesian coordinates.

- Direct maximum likelihood estimation is in general not recommended and dimensionally limited because it requires $\Lambda_\theta\{A_{\mathrm{r}}(u)\}$.

- Model estimation in "moderately high" dimensions is possible within the framework of proper scoring rules (de Fondeville and Davison, 2017).

The gradient scoring rule

- An adaptation of the gradient scoring rule (de Fondeville and Davison, 2017) allows statistical inference using partial derivatives, with respect to $x_1, \dots, x_\ell$, of the log-density function,

$$\delta_{\mathrm{w}}(\lambda^{\mathrm{r}}_{\theta,u}, x) = \sum_{i=1}^{\ell} \left(2 w_i(x) \frac{\partial w_i(x)}{\partial x_i} \frac{\partial \log \lambda^{\mathrm{r}}_{\theta,u}(x)}{\partial x_i} + w_i(x)^2 \left[\frac{\partial^2 \log \lambda^{\mathrm{r}}_{\theta,u}(x)}{\partial x_i^2} + \frac{1}{2} \left\{ \frac{\partial \log \lambda^{\mathrm{r}}_{\theta,u}(x)}{\partial x_i} \right\}^2 \right] \right),$$

  where $w : A_{\mathrm{r}}(u) \to \mathrm{r}^{\ell}_{+}$ is a weighting function differentiable on $A_{\mathrm{r}}(u)$ and vanishing on the boundaries of $A_{\mathrm{r}}(u)$.
- Maximization of δ_{w} gives an asymptotically unbiased and normal estimator.

Brown–Resnick model

- Recall that for $\xi \neq 0$, P is

$$P = R \frac{A(R_1 Q)^\xi + B}{\mathrm{r}\{A(R_1 Q)^\xi + B\}}$$

- Suppose W follows a log-Gaussian distribution with stationary increments and semi-variogram γ.
- The ℓ-dimensional intensity function is

$$\lambda_\theta^{\mathrm{r}}(x) = \frac{|\Sigma_\theta|^{-1/2}}{x_1^2 x_2 \cdots x_\ell (2\pi)^{(\ell-1)/2}} \exp\left(-\frac{1}{2}\widetilde{x}^T \Sigma_\theta^{-1} \widetilde{x}\right), \quad x \in \mathbb{R}_+^\ell \setminus \{0\},$$

where $\widetilde{x}$ is the $(\ell-1)$-dimensional vector

$$\{\log(x_j/x_1) + \gamma_\theta(s_j - s_1) : j = 2, \ldots, \ell\}^T,$$

and Σ_θ is the $(\ell-1) \times (\ell-1)$ matrix

$$\{\gamma_\theta(s_i - s_1) + \gamma_\theta(s_j - s_1) - \gamma_\theta(s_i - s_j)\}_{i,j \in \{2,\ldots,\ell\}}.$$

Measure of extremal dependence

- The extremogram $\pi(h)$ is a tool to measure the strength of dependence ($\sim$ variogram for extremes);

$$\pi(h) = \mathrm{Pr}\left[\mathrm{X}(\mathrm{s}+\mathrm{h}) > \mathrm{u} \,\middle|\, \{\mathrm{X}(\mathrm{s}) > \mathrm{u}\} \cap \left\{\mathrm{r}\left(\frac{\mathrm{X}}{\mathrm{u}}\right) > 1\right\}\right].$$

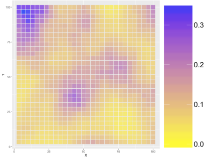
- For the Brown–Resnick model,

$$\pi(h) = 2\left[1 - \Phi\left\{\left(\frac{\gamma(h)}{2}\right)^{1/2}\right\}\right],$$

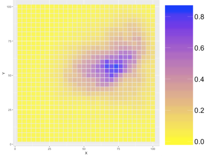
where γ is a valid semi-variogram and Φ is the cumulative distribution function of a Gaussian random variable.

Variogram and extremogram

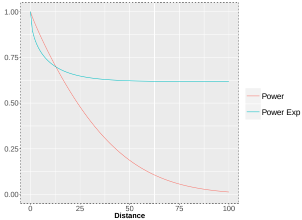
Simulation of generalized r-Pareto process with $\int_S X(s)ds \geqslant 100$.



$$\gamma(h) = \sigma\left[1 - \exp\left\{-\left(\frac{h}{\tau}\right)^{\alpha}\right\}\right]$$



$$\gamma(h) = \left(\frac{h}{\tau}\right)^{\alpha}$$



Theoretical $\pi(h)$ for two different variograms

Application: Extreme European winter storms

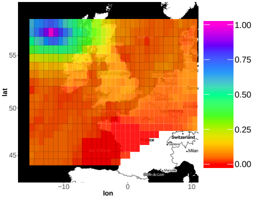
- 3s maximum windgust every 3 hours for the period 1979 to 2016 from ERA-Interim reanalysis model.
- Storms are defined as an exceedance of an 24 hours temporal aggregation of the spatial mean:

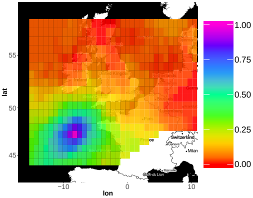
$$\mathrm{r}(X) = \sum_{i=1}^{8} \int_S X(s) ds.$$

- Time frame is centered on the 24 hour maximum of the spatial mean.
- 200 events are used to fit a Pareto process.

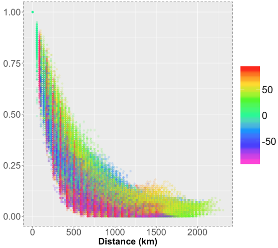
Application: Extreme European winter storms

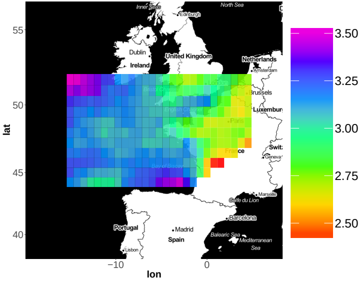
Estimated $\pi(h)$ for two different locations



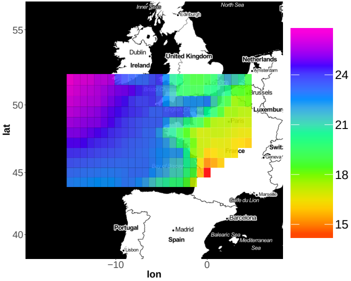


Extremogram cloud





Scales



Locations

Application: Extreme European winter storms

- Variogram model:

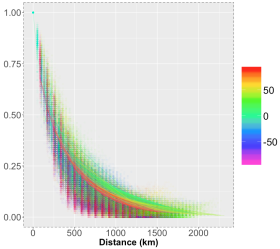
$$\gamma(s_i, s_j, t_i, t_j) = \left\| \frac{\Omega\{s_i - s_j + V(t_i - t_j)\}}{\tau} \right\|_2^\kappa, \quad s_i, s_j \in S, \quad t_i, t_j \in \{0, \ldots, 21\},$$

with $0 < \kappa \leqslant 2, \tau > 0$, wind vector $V \in \mathbb{R}^2$ and anisotropy matrix

$$\Omega = \begin{bmatrix} \cos\eta & -\sin\eta \\ a\sin\eta & a\cos\eta \end{bmatrix}, \quad \eta \in \left(-\frac{\pi}{4}; \frac{\pi}{4}\right], \quad a > 0.$$

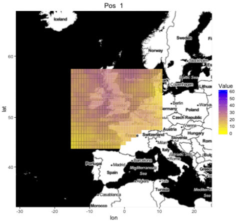
- Estimated parameters

κ	τ	a	η	V_1	V_2
1.17	348.6	1.25	-0.02	0.25	-0.01



Application: European extra-tropical cyclones

Simulated extreme windstorm over Europe



Summary and discussion

- Classical geostatistics should be avoided when modelling extreme events.
- Generalized r-Pareto process is the functional equivalent of the generalized Pareto distribution and allows one to model r-exceedances.
- The Brown–Resnick model uses classical variogram models, while the corresponding stochastic process is heavy-tailed.
- Inference using the gradient scoring rule enables inference in "moderately high" dimensions and is limited by matrix inversion.
- We developed a (too) simple spatio-temporal generator for extreme windstorms in Europe.
- Ongoing work:
  - Marginal modelling;
  - Complex dependence structure to better capture the characteristics of the dependence structure.

Bibliography

de Fondeville, R. and Davison, A. C. (2017). High-dimensional Peaks-over-threshold Inference. *arXiv:1605.08558*.

de Fondeville, R. and Davison, A. C. (2018). Functional Peaks-over-threshold and Generalized r-Pareto Process. *preprint*.

Hult, H. and Lindskog, F. (2005). Extremal Behavior of Regularly Varying Stochastic Processes. *Stochastic Processes and their Applications*, 115(2):249–274.

Hyvärinen, A. (2007). Some Extensions of Score Matching. *Computational Statistics & Data Analysis*, 51(5):2499–2512.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary Max-stable Fields Associated to Negative Definite Functions. *Annals of Probability*, 37(5):2042–2065.

Lindskog, F., Resnick, S. I., and Roy, J. (2014). Regularly Varying Measures on Metric Spaces: Hidden Regular Variation and Hidden Jumps. *Probability Surveys*, 11(1):270–314.

Risk functional

Risk functional

A monotonic increasing functional $\mathrm{r} : \mathcal{C}(S) \to \mathbb{R}$ satisfying

$$\begin{array}{lll} \mathrm{r} \text{ continuous at } \{B - A\xi^{-1}\} & \text{and } \mathrm{r}(B - A\xi^{-1}) < 0, & \xi > 0 \\ \mathrm{r}(x) \to -\infty, \quad x \to -\infty & & \xi \leqslant 0, \end{array}$$

and for which there are functions $A > 0$ and B such that

$$\lim_{n\to\infty} \sup_{s\in S} \left| \frac{a_n(s)}{r(a_n)} - A(s) \right| = 0, \quad \lim_{n\to\infty} \sup_{s\in S} \left| \frac{b_n(s) - \mathrm{r}(b_n)}{\mathrm{r}(a_n)} - B(s) \right| = 0,$$

is called a risk functional.

Properties of generalized r-Pareto processes

- The r-exceedance distribution of P is

$$\Pr\{\mathrm{r}(P) \geqslant \rho\} = \left\{1 + \xi \frac{\rho - u}{\sigma}\right\}^{-1/\xi}, \quad \rho \geqslant u.$$

- The generalized r-Pareto process has generalized Pareto marignals above a sufficiently high threshold $u_0 \geqslant 0$:

$$\Pr\{P(s_0) \geqslant \rho \mid P(s_0) \geqslant u_0\} = \left\{1 + \xi \frac{\rho - u_0 A(s_0) - B(s_0)}{\sigma(u_0)}\right\}^{-1/\xi}, \quad \rho \geqslant u_0,$$

with $\sigma(u_0) = \sigma A(s_0) + \xi \{u_0 - A(s_0)u - B(s_0)\}$.

The gradient score

Proposition (de Fondeville and Davison, 2017)

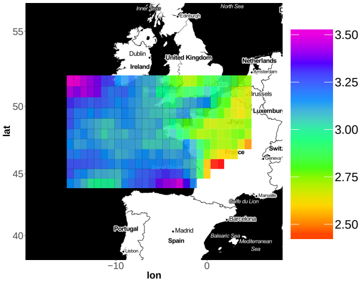
The scoring rule $\delta_{\mathrm{w}}(\lambda^{\mathrm{r}}_{\theta,u}, \cdot)$ is strictly proper, i.e., the estimator

$$\widehat{\theta}^{\mathrm{r}}_{\delta}\{x^1, \ldots, x^n\} = \arg\max_{\theta \in \Theta} \sum_{m=1}^{n} \epsilon\left\{\mathrm{r}\left(\frac{x^m}{u_n}\right) > 1\right\} \delta(\lambda^{\mathrm{r}}_{\theta,u_n}, x^m), \quad (3)$$

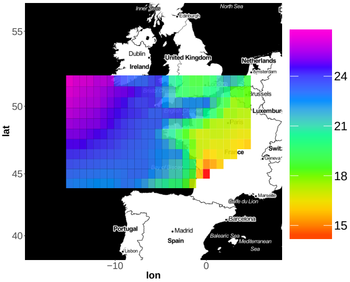
where $\epsilon\{\cdot\}$ is the indicator function and $x^1, \ldots, x^n$ are sampled from the random vector X with normalized marginals, is consistent and asymptotically normal as $n \to \infty$ and $u_n \to \infty$ with $N_{u_n} = o(n)$.

In a simulation study, we compared the gradient scoring rule with spectral likelihood and censored likelihood.

Extreme European winter storms: Marginal model



Scales



Locations