1. Introduction

Estimating the risk of extreme natural hazards has been a major issue in recent decades, but has been limited to the exploitation of catalogs of historical events, which usually do not exceed 40 to 50 years, and to numerical models, which require heavy computation while being unreliable for extrapolation above observed intensities.

Extreme Value Theory provides a theoretical framework to describe and model tails of statistical distributions. This posterior describes a statistical procedure to estimate the extremal behavior of rare events in order to build models that can quantify the recurrence of past events as well as safely extrapolate above observed intensity levels. With this methodology, we develop a stochastic weather generator for extreme windstorms over Europe.

2. Univariate extreme value theory

Under mild conditions on a random variable $X$, the central limit theorem gives

$$
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( -X_i + \mu \right) = \mathcal{N}(0, \sigma^2),
$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$, and thus normal distributions approximate $\sum_{i=1}^{n\text{large}} \left( -X_i + \mu \right)$ up to an affine rescaling. The generalized Pareto distribution

$$
H_{\xi}(x) = \left( 1 + \xi (x/a)^{\xi} / \xi \right)^{-1}, \quad \xi \neq 0,
$$

$$
\lim_{x \to 0} H_{\xi}(x) = \left\{ \begin{array}{ll}
1 + \xi x/a & \xi > 0, \\
\log(x/a) & \xi = 0,
\end{array} \right.
$$

where $a > 0$ and $x = \mu + \sigma u$ has a similar role in approximating tail distributions. Indeed, for suitable sequences $u_n \to 0$ and $h_n$,

$$
\lim_{n \to \infty} \mathbb{P}( \frac{X - \mu}{\sigma} > n ) = H_{\xi}(a).
$$

The tail index $\xi$ determines the regime of tail decay:

- $\xi > 0$: Frechet type with $\xi > 0$ and polynomial tail decay.
- $\xi = 0$: Gumbel type with $x \sim n \sigma / \xi$ and exponential tail decay.
- $\xi < 0$: Weibull type with bounded tail, $x \sim n^{1/\xi}$.

Since the generalized Pareto distribution is the only limit distribution for threshold exceedances, for any random variable $X$ and large enough threshold $\theta = \mu (F_X(1) - 1)$, the approximation

$$
\mathbb{P}( X > \theta ) \sim \mathbb{P}( \frac{X - \mu}{\sigma} > n ) = H_{\xi}(a).
$$

This provides a model for tails, which is, due to (1), safe for extrapolation above observed intensities. However, severe climatic events, such as floods, windstorms, heatwaves, cannot be modelled using only univariate extreme value theory, as it fails to capture their spatio-temporal nature.

3. Functional peaks-over-threshold analysis

3.1 Functional exceedance

Characterization of an exceedance for a univariate quantity is straightforward: for some threshold $\omega > 0$, any observation $X$ such that $X > \omega$ is defined as an exceedance. When $X$ is a function the definition of an exceedance more delicate.

If $X \in \mathcal{C}(S)$ is a continuous function over a compact subset $S \subset \mathbb{R}^d$, we define an exceedance as an event $(i(X) > \omega)$, where $i(X)$ is a monotonically increasing functional, called a ’risk functional’. Common examples are

- $i_{\text{up,sp}}(X)(x)$ for events where $X$ exceeds a threshold at least at one location;
- $\int_{S} X(x, t) \text{d}x$ for spatio-temporal accumulation;
- $\int_{S} |X(x)|^p \text{d}x$ as a proxy of the energy inside a system;
- $X(x, t)$ for risks at a specific location $x \in S$.

For simplicity, now also suppose that $i$ is linear.

3.2 Generalized r-Pareto process

A generalized r-Pareto process $P$ (de Fondeville and Davison, 2018) is defined by

$$
P = \frac{R \mathbb{W}}{R + \log W - \log(M)}, \quad \xi \neq 0,
$$

$$
\mathbb{P}(X > \theta) = 1 - \left( 1 + \xi (x/a)^{\xi} / \xi \right)^{-1}, \quad \theta > 0,
$$

with $\sigma > 0$. It determines the intensity of the r-exceedance,

- $\mathbb{W}$, the angular component, is a stochastic process on the unit sphere $\{ Z \in \mathbb{S}^2 : |\mu_i| = 1 \}$ with probability measure $P_{\mu}$, which determines the dependence and marginal tail distribution of $P$.

- The r-exceedance distribution of $P$ is

$$
\mathbb{P}((X > \theta) \cap (Y > \theta)) = \mathbb{P}(\mathbb{W} > \theta), \quad \theta > 0.
$$

The generalized r-Pareto process has generalized Pareto marginals above a sufficiently high threshold $\omega_0 > 0$,

$$
\mathbb{P}(\mathbb{W} > \theta) = \left( 1 + \xi (x/a)^{\xi} / \xi \right)^{-1}, \quad \theta > 0,
$$

with $a(\omega_0) > 0$ and $\mu(\omega_0) \in \mathbb{R}$.

3.3 Convergence

The generalized r-Pareto process is the only possible limit of a rescaled regularly varying stochastic process, i.e.,

$$
\mathbb{P}(X > \omega) \sim \mathbb{P}(\mathbb{W} > \omega), \quad \omega \to \infty,
$$

and thus for a stochastic process $X$ and large enough threshold $\omega_0 > 0$,

$$
\mathbb{P}(X > \omega_0) \sim \mathbb{P}(\mathbb{W} > \omega_0), \quad \omega_0 \to \infty.
$$

Consequently the generalized r-Pareto process models the tail of any stochastic process $X$ and the convergence (2) ensures safe extrapolation above observed intensities. In equation (3), $P(X > \omega)$ represents the overall probability of observing an extreme event, while the part relative to the generalized r-Pareto process describes the intensity and pattern of the r-exceedances.

4. Application to extreme windstorm in Europe

We apply the previous results to extreme windstorms over Europe. To do so, we study maximum wind gusts measured every 3 hours over the period 1973 to 2016 using data from ERA-Interim reanalysis (Dee et al., 2011).

4.1 Storm definition

For insurers and regulators, the most damaging windstorms impact areas with dense human infrastructure. Thus, we define a storm as an exceedance of the 24-hour temporal maxima of the spatial mean in a small region including Paris, London, Brussels, and Amsterdam, i.e.,

$$
\gamma(X) = \max_{t \in [0, 24]} \mathbb{P}(\mathcal{S}_{\text{sp}}(X(t)) > \omega),
$$

where $\mathcal{S}_{\text{sp}}(X(t))$ is the spatial mean in a small region including Paris, London, Brussels, and Amsterdam, i.e.,

$$
\mathcal{S}_{\text{sp}}(X(t)) = \frac{1}{n} \sum_{i=1}^{n} X_i(t),
$$

with $\omega$, $\gamma > 0$, and a wind vector $V \in \mathbb{R}^2$, and an anisotropy matrix

$$
\Omega = \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}, \quad \ell \geq 2.
$$

Figure 2 shows a simulation of a storm with an intensity of $\gamma(X) = 20$ km h$^{-1}$. For comparison, the storm Daria which occurred during the winter 1999-2000, for which the estimated insured loss is around 90 billion dollars, had an intensity equal to 32 km h$^{-1}$.

4.2 Windstorm frequency

We choose the threshold $\omega = 20$ km h$^{-1}$ to obtain a total of 61 r-exceedances, yielding an average of 1.7 storms per year. We use a logistic regression to model $\mathbb{P}(\gamma(X) > \omega)$ with the North Atlantic Oscillation index and temperature anomaly index as covariates, both were found to be highly significant. See Figure 2.

4.3 Marginal model

The marginal tail distribution is specified by the tail index $\xi = -1.22$ and the functions $a_n$ and $h_n$ displayed in Figure 3. Goodness of fit diagnostics can be found in Figure 4.

5. Acknowledgement

This work was supported by the Swiss National Science Foundation.

References


Figure 3: Estimated scale $a_n$ (left) and location $h_n$ (right) functions.

Figure 4: Q-plots for the marginal tail distribution of four locations (top row and bottom left) represented by the green points on the map. The bottom right frame shows the q-q plot for the distribution of the exceedances of $\gamma(X)$ above the threshold $\omega = 21$ km h$^{-1}$.