Generalized Peaks-Over-Thresholds Modelling

Anthony Davison

Joint with
Raphaël de Fondeville

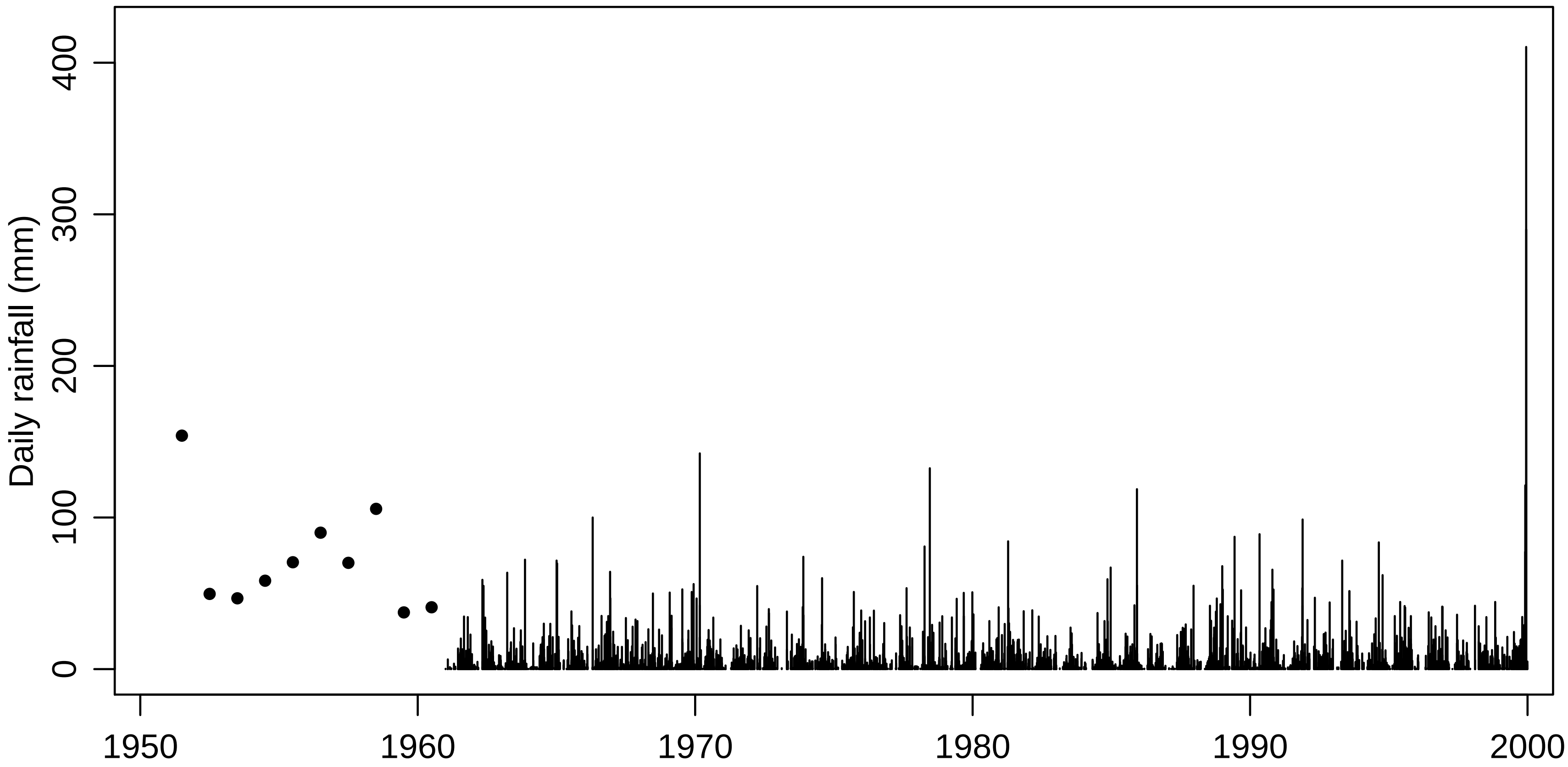
Funding: Swiss National Science Foundation

▷ Motivation
Functional setting
Application
Closing

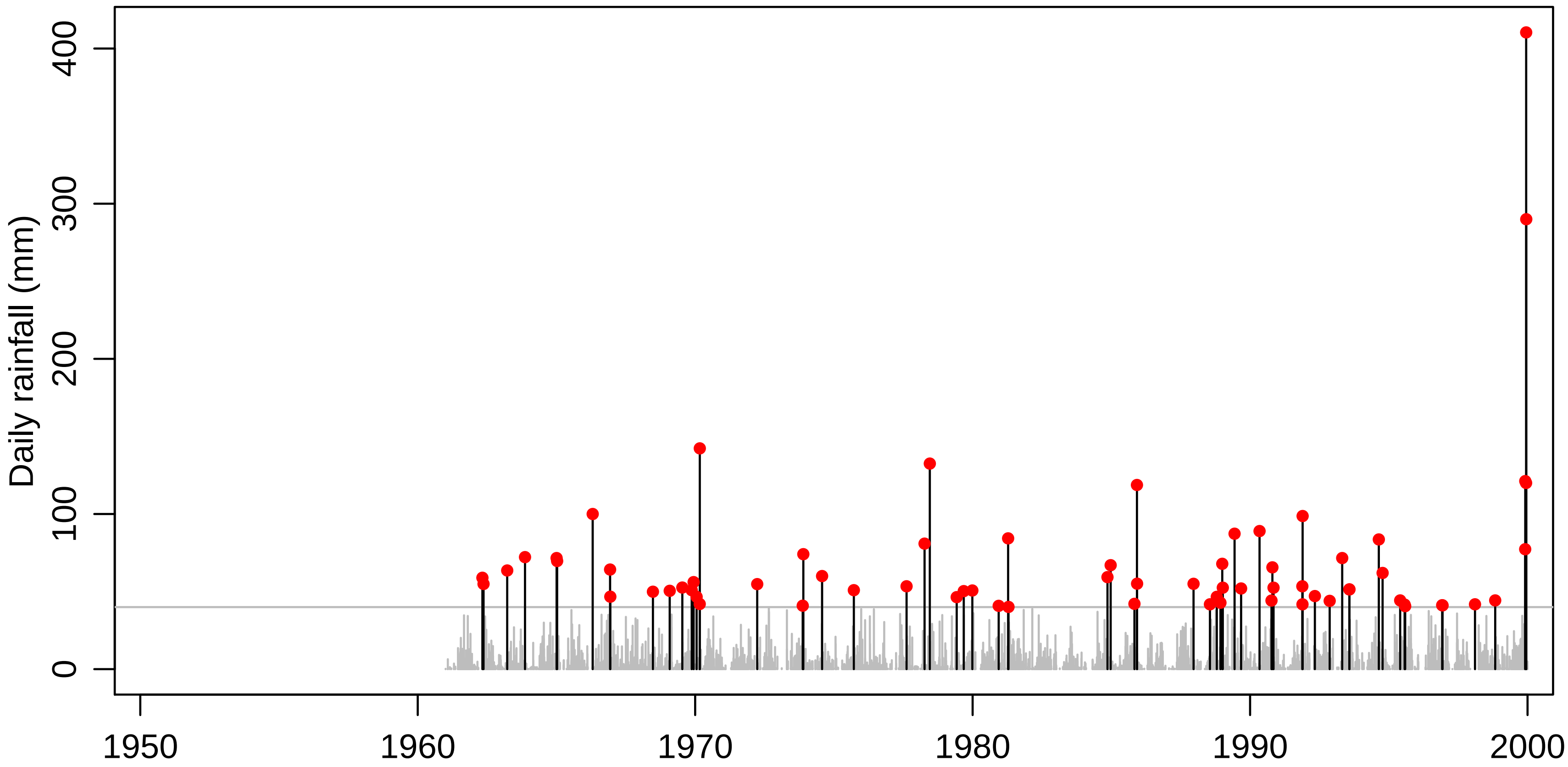
Motivation

- Two broad approaches to dealing with extremes of a scalar time series
  - block maxima
  - exceedances over thresholds
- Block maxima traditionally used in applications, often modelled using **generalized extreme-value (GEV) distribution**
- Threshold exceedances also widely used:
  - simpler and more flexible, involve fitting the **generalized Pareto distribution (GPD)** to threshold exceedances
  - in principle allow more detailed modelling of events—in practice often just retain 'peaks' of clusters of exceedances, hence **Peaks Over Thresholds (POT)** analysis
- Use of POT originated in hydrology ('partial duration series') in early 1970s, accompanying probability theory developed by Balkema, de Haan, Pickands in mid-1970s, widespread statistical use since ~ 1990

Vargas rainfall series



Vargas partial duration series



Univariate extreme-value statistics

- For a scalar random variable X, under mild conditions

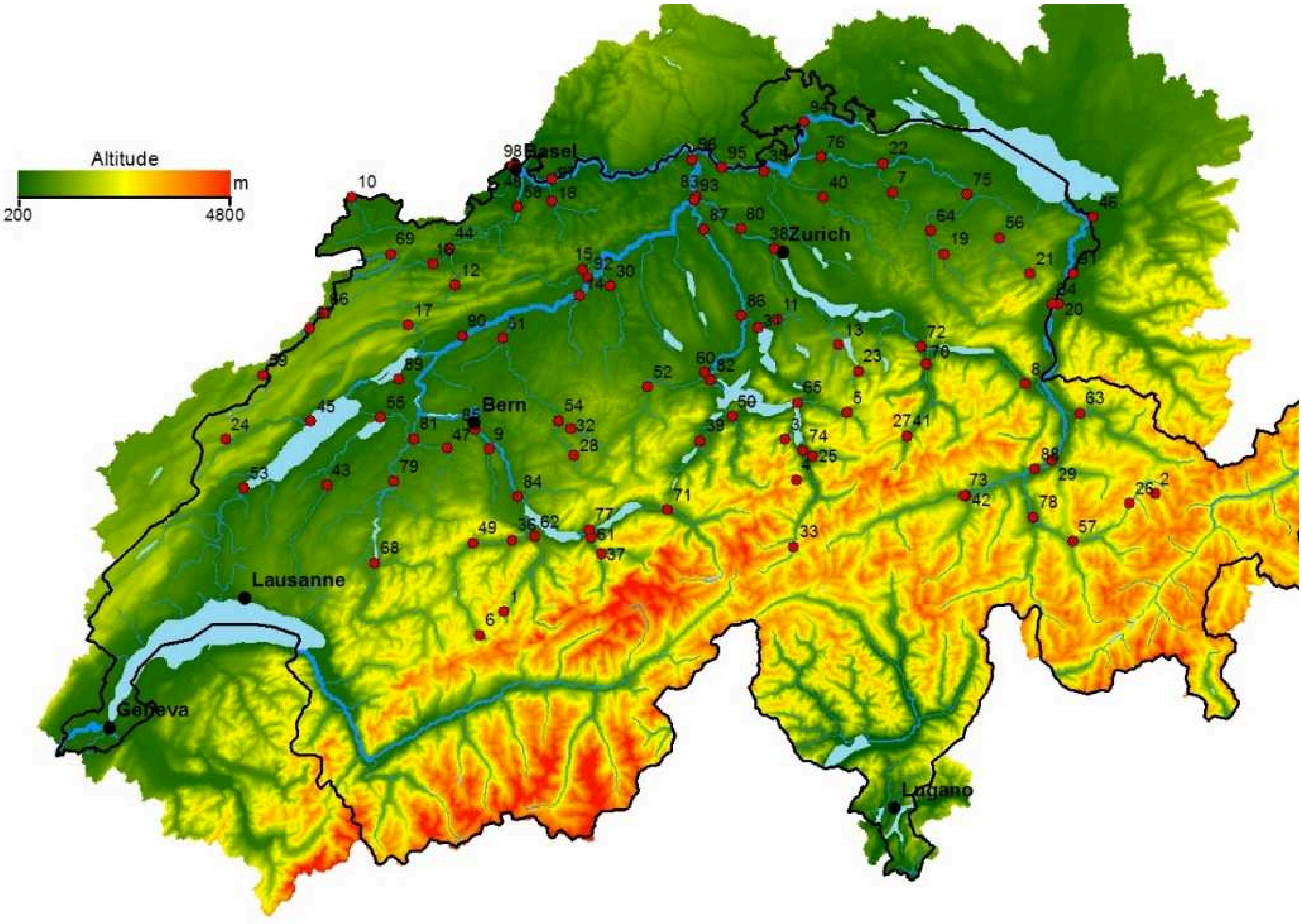
$$\mathrm{P}(X > x \mid X - b_n > 0) \to \left(1 + \xi \frac{x - b_n}{a_n}\right)_+^{-1/\xi}, \quad n \to \infty,$$

  where b_n is a sequence of increasingly large thresholds. Equivalently,
  - the conditional distribution of exceedances over a high threshold can be approximated by a **generalized Pareto distribution**.
  - the exceedances of a random sample above a high threshold can be modelled by a **Poisson process**.
- These results form the basis of statistical methods for analysis of rare events, applied as approximations to exceedances over some threshold u:
  - we set $b_n = u$, retain only the n_u exceedances $\{X_j - u : X_j - u > 0\}$,
  - fit the GPD/Poisson process to the exceedances, and then
  - extrapolate (crossing fingers . . .).
- Here an 'event' is scalar: what if it is more complex?

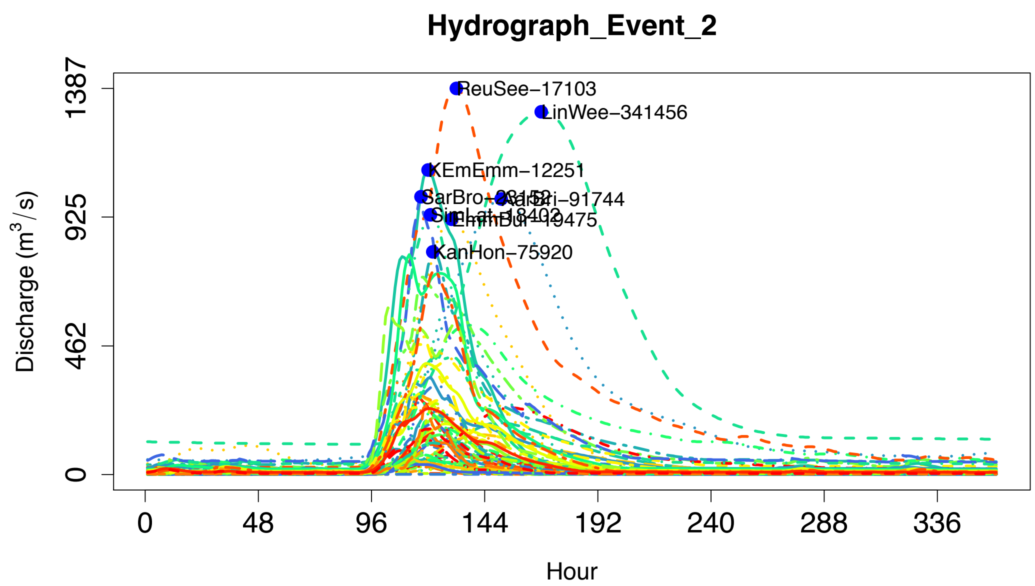
EXAR project

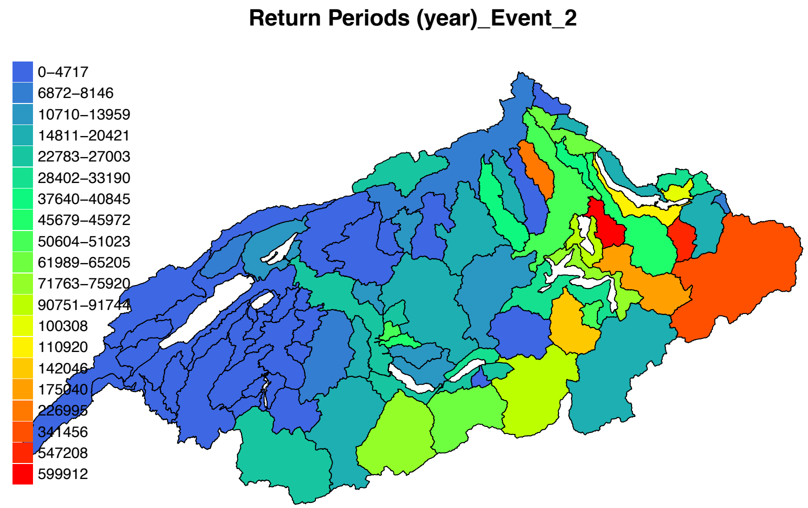
- Goal: estimate flood risk on Aare river basin up to 2050, taking into account climate change.
- Probabilities needed for events on river network with annual probabilities 10^{-4} (Swiss nuclear power plants are on the riverbanks).
- Must assess combined flooding risk, using data including ~ 80 river flow time series of length at most 80 years, rainfall observations, climate simulations, . . . ,
- Basic task: **extrapolation** (unreasonably?) far outside the range of any observations.

Aare river basin



Large event in terms of subcatchment discharges





Motivation
▷ Functional setting
Application
Closing

Functional setting

Risk functional

- We represent the phenomenon of interest by a stochastic process X defined on a domain S and taking values in the space of continuous functions $\mathcal{C}(S)$, and define an **$\mathcal{R}$-exceedance** as an event

$$\mathcal{R}(X) \geq u,$$

where $\mathcal{R}$ is a scalar monotone increasing **risk functional** and $u \geq 0$ is a threshold.

- Examples of $\mathcal{R}(X)$:
  - $\sup_{s \in S} X(s)$ for events that exceed u somewhere (anywhere) in S,
  - $X(s_0)$ for risks at a specific location s_0,
  - $\sum_t \int_S X_t(s)\,\mathrm{d}s$ for the total size of the event over some period,
  - $\left\{\int_S X(s)^2\,\mathrm{d}(s)\right\}^{1/2}$ when the risk is determined by 'energy',
  - ...

Connections

- Previous work:
  - Opitz (2013, PhD thesis) ‘radial aggregation function’
  - Ferreira and de Haan (2014, Bernoulli): propose $\sup_{s \in S} X(s)$, but in practice X is observed only on a finite subset of S
  - Dombry and Ribatet (2015, Statistics and its Interface): homogeneous ‘cost functionals’
  - de Fondeville and Davison (2018, Biometrika): precursor to the present work
- Lacunae:
  - need wide range of functionals defined in terms of the original process, not on some transformed version
  - should to cover all types of tail behaviour, i.e., no restrictions on ξ
- Our goal: overcome these difficulties.
- For (lots of!) details, see de Fondeville and Davison (2022+, series B)

Which functions?

- All defined on compact subset S of $\mathbb{R}^p$
- 'Standardized functions':
  - $\mathcal{F}$ denotes space of real-valued continuous functions on S, equipped with norm $\|\cdot\|$;
  - $\mathcal{F}_+ \subset \mathcal{F}$ denotes non-negative functions not everywhere zero.
- Transformed functions depend on scale and location functions a and b and on shape parameter ξ, and so does their domain.
- Let ξ be a scalar shape parameter, let $a \equiv a(s) > 0$ and $b \equiv b(s)$ be continuous functions defined for $s \in S$, and define

$$\mathcal{F}^{\xi,a,b} = \begin{cases} \mathcal{F}_+ - (b - \xi^{-1}a), & \xi > 0, \\ \mathcal{F}, & \xi = 0, \\ (b - \xi^{-1}a) - \mathcal{F}_+, & \xi < 0. \end{cases}$$

Illustration

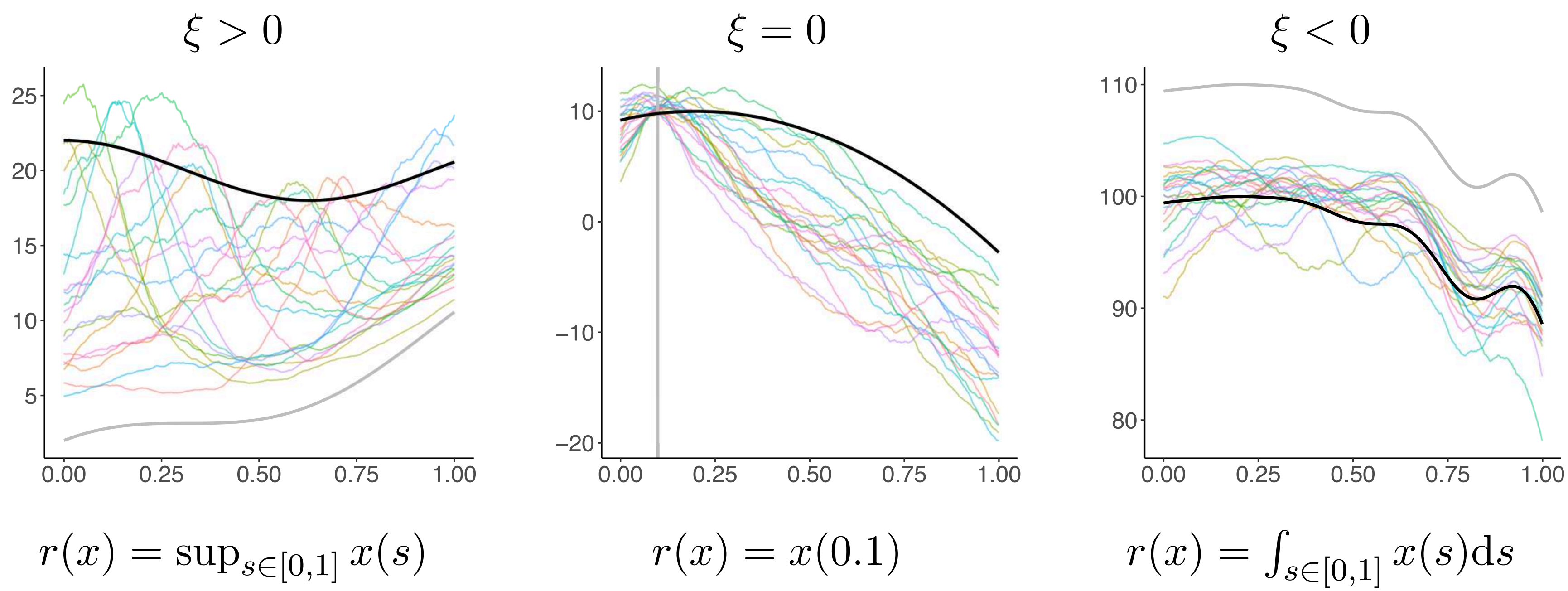


Fig. 2. Realisations of generalized r-Pareto processes on $S = [0,1]$, with location function b (heavy black line) and lower/upper bound $b - \xi^{-1}a$ of $\mathcal{F}^{\xi,a,b}$ (heavy grey line). Left: realisations x with shape parameter $\xi > 0$ and risk functional $r(x) = \sup_{s\in S} x(s)$, with the parts of the xs below the threshold function b shown as dotted lines. Middle: realisations with $\xi = 0$ and risk functional $r(x) = x(0.1)$ representing evaluation at $s = 0.1$ (vertical line). Right: realisations x with $\xi < 0$ and risk functional $r(x) = \int_{s\in S} x(s)\mathrm{d}s$.

Assumptions

- X is a stochastic process with sample paths in $\mathcal{F}$ for which there exist a real number ξ and sequences $\{a_n\}_{n=1}^{\infty} > 0$ and $\{b_n\}_{n=1}^{\infty}$ of continuous functions such that there is pointwise convergence to the GPD at each point of S:

$$\lim_{n\to\infty} n\mathrm{P}\left\{\frac{X(s)-b_n(s)}{a_n(s)} > x\right\} = \begin{cases} (1+\xi x)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x), & \xi = 0. \end{cases}$$

- We assume the existence of a boundedly finite non-zero measure Λ on the space of non-negative and non-zero functions over S such that

$$\left.\begin{array}{ll} \lim_{n\to\infty} n\mathrm{P}\left[\left\{1+\xi\left(\frac{X-b_n}{a_n}\right)\right\}_+^{1/\xi} \in \cdot\right], & \xi \neq 0, \\ \lim_{n\to\infty} n\mathrm{P}\left\{\exp\left(\frac{X-b_n}{a_n}\right) \in \cdot\right\}, & \xi = 0, \end{array}\right\} = \Lambda(\cdot),$$

  where $\{a\}_+ = \max\{a(s), 0\}$ is a function of s; then we say that $X \in \mathrm{GRV}(\xi, a_n, b_n, \Lambda)$.
- The limiting measure Λ is homogeneous of order -1.
- We also assume that there exist real numbers a_n' and a continuous strictly positive function A on S such that $a_n(s) \approx a_n' A(s)$ for large n.

- A risk function is **valid** for X if the $\mathcal{R}$-exceedances of the limiting process have finite positive measure. Holds in fairly wide generality.
- If $u \geq 0$, the risk functional $\mathcal{R}$ is valid for $X \in \mathrm{GRV}(\xi, a_n, b_n, \Lambda)$ and the assumptions above hold, then

$$\mathrm{P}\left[\frac{X - b_n}{\mathcal{R}(a_n)} \in \cdot \;\middle|\; \mathcal{R}\left\{\frac{X - b_n}{\mathcal{R}(a_n)}\right\} \geqslant u\right] \to \mathrm{P}(P \in \cdot\,), \quad n \to \infty,$$

  where P is a generalized $\mathcal{R}$-Pareto process with tail index ξ, scale function A, location function zero and measure Λ.
- For given ξ and positive function $a \equiv a(s)$, let $A = a/\mathcal{R}(a)$ and let

$$\mathcal{A}_{\mathcal{R}} = \begin{cases} \left\{y \in \mathcal{F}_+ : \mathcal{R}\left(A\dfrac{y^\xi - 1}{\xi}\right) \geqslant 0\right\}, & \xi \neq 0, \\ \left\{y \in \mathcal{F}_+ : \mathcal{R}\left(A \log y\right) \geqslant 0\right\}, & \xi = 0, \end{cases}$$

  contain possible sample paths for P.

Generalized $\mathcal{R}$-Pareto processes

- The **generalized $\mathcal{R}$-Pareto process** P associated to the measure Λ and tail index ξ takes values in $\{x \in \mathcal{F}^{\xi,a,b} : \mathcal{R}\{(x-b)/\mathcal{R}(a)\} \geqslant 0\}$ and is defined as

$$P = \begin{cases} a(Y_{\mathcal{R}}^{\xi} - 1)/\xi + b, & \xi \neq 0, \\ a \log Y_{\mathcal{R}} + b, & \xi = 0, \end{cases}$$

  where $Y_{\mathcal{R}}$ is a stochastic process on $\mathcal{A}_{\mathcal{R}}$ with probability measure $\Lambda(\cdot)/\Lambda(\mathcal{A}_{\mathcal{R}})$.

- **Polar decomposition** for $Y_{\mathcal{R}}$ is

$$Y_{\mathcal{R}} \stackrel{\mathrm{D}}{=} RW \;\Big|\; \mathcal{R}[A\xi^{-1}\{(RW)^{\xi} - 1\}] \geq 0,$$

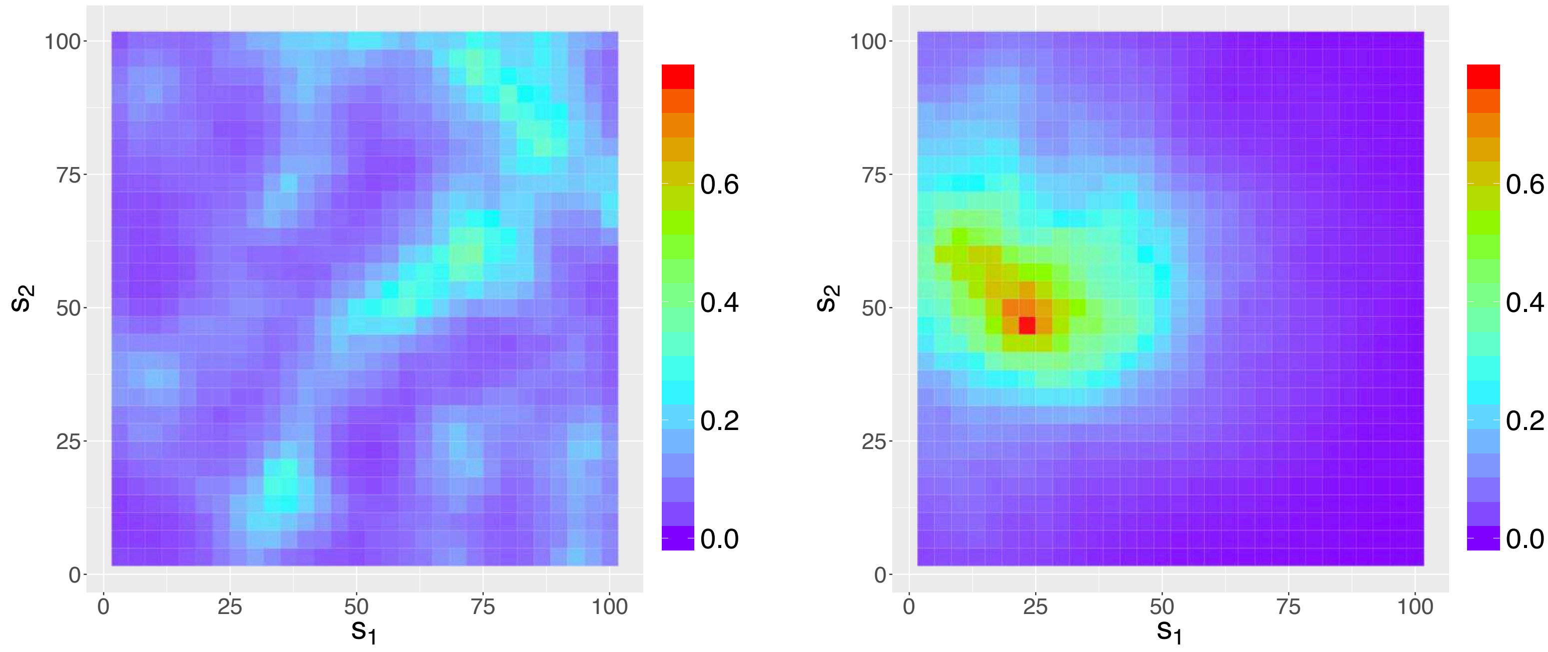
  where R and W are independent, the scalar R is unit Pareto and W is a stochastic process on S with values in $\mathcal{S} = \{y \in \mathcal{F}_+ : \|y\|_1 = 1\}$ with probability measure

$$\sigma_0(\cdot) = \frac{\Lambda\{y \in \mathcal{F}_+ : y/\|y\|_1 \in \cdot\,, \|y\|_1 \geq 1\}}{\Lambda\{y \in \mathcal{F}_+ : \|y\|_1 \geq 1\}},$$

  where $\|\cdot\|_1$ denotes the 1-norm on $\mathcal{F}_+$.

- Polar decomposition allows simulation of $Y_{\mathcal{R}}$, even with fixed risk levels.

Simulations



$$\gamma(h) = 0.5\left[1 - \exp\{-(\|h\|/15)^{1.8}\}\right] \qquad \gamma(h) = (\|h\|/30)^{1.8}$$

Fig. 3. Simulated generalized r-Pareto processes with $r(X) = \int_S X(s)ds = 100$ for two semi-variogram functions $\gamma(h)$. Left: bounded power-exponential semi-variogram function. Right: unbounded power variogram.

Properties

- For any location $s_0 \in S$, the generalized $\mathcal{R}$-Pareto process has a **GP distribution** above a sufficiently high threshold function $u_0 \geqslant 0$:

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

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

- If the risk functional is **linear**, the $\mathcal{R}$-exceedance distribution of P is also generalized Pareto:

$$\mathrm{P}\left\{\mathcal{R}(P) \geqslant r\right\} = \left\{1 + \xi \frac{r - \mathcal{R}(b)}{\mathcal{R}(a)}\right\}^{-1/\xi}, \quad r \geqslant \mathcal{R}(b).$$

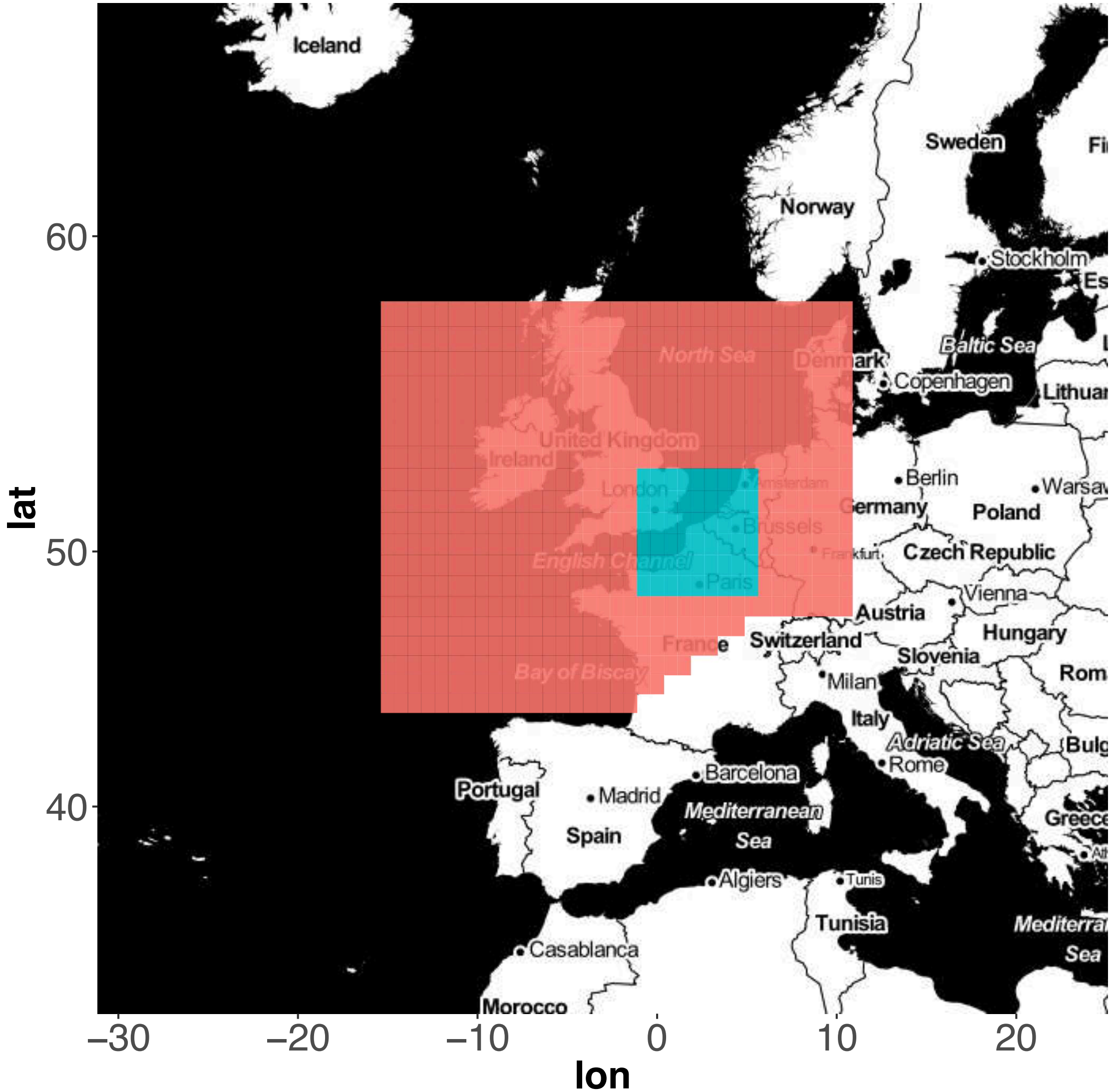
- These are natural, and can be used for model-checking.

Motivation
Functional setting
▷ Application
Closing

Application

Extreme European wind storms

Area of Study



Storm Daria

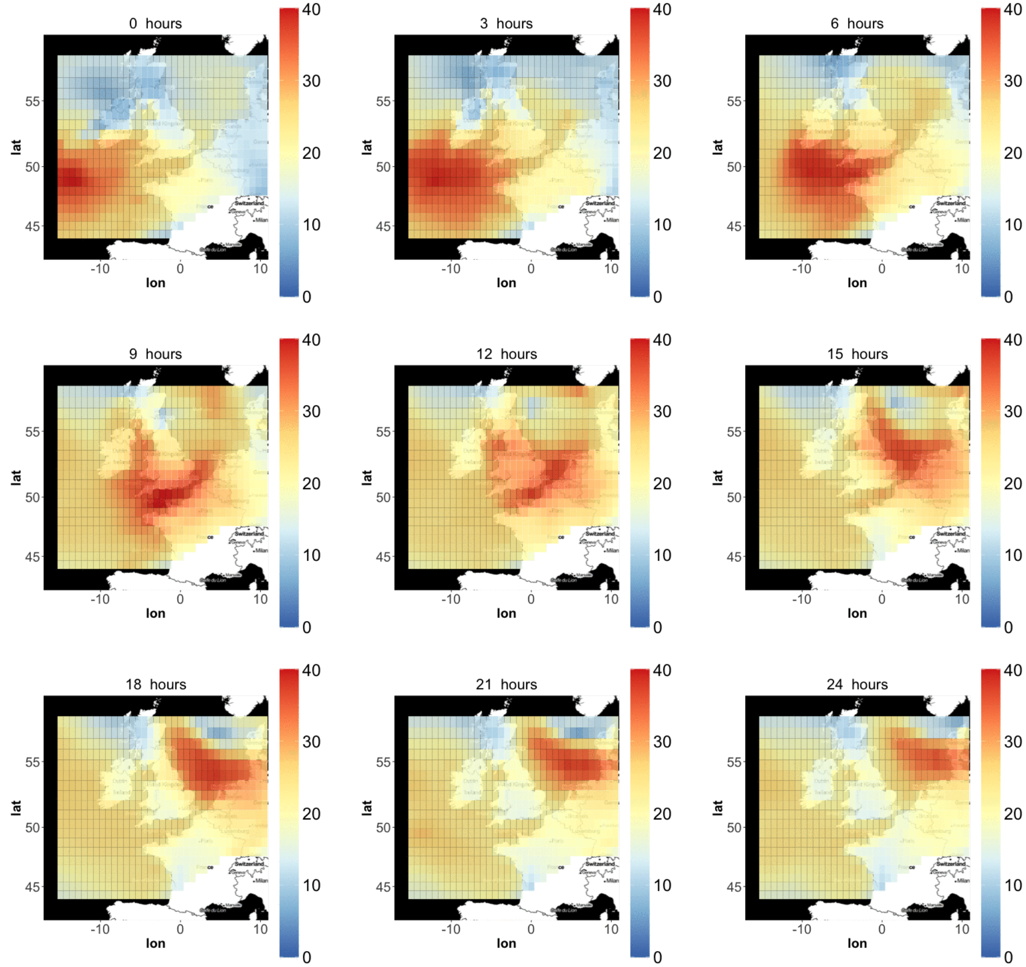


Fig. 4. Maximum speed (ms^{-1}) over the past 3 hours of the wind gusts sustained for at least 3s from ERA-Interim reanalysis during the peak of windstorm Daria, which swept over Europe during January 1990.

Extreme European wind storms

- Risk estimation usually based on historical catalogues: realistic but limited.
- EV modelling has numerous advantages, but must model complexity for risks of interest.
- Base model on maximum speed of windgusts sustained for at least 3s every 3 hours for the period 1979 to 2016 from ERA-Interim reanalysis model.
- Storms are defined as an exceedance of the the spatial mean over a region with dense human infrastructure (in green),

$$\mathcal{R}(X)(t) = |E_{\mathrm{ABLP}}|^{-1} \int_{E_{\mathrm{ABLP}}} X(s,t) ds, \quad t \in T,$$

  containing Amsterdam, Brussels, London and Paris.
- Study period T is months October–March for 1979–2014.
- A time frame of 24h is centered on the maximum of the spatial mean.
- We use the 0.96 quantile as threshold, yielding 63 events with which to fit a generalized r-Pareto process.

Comparison with XWS catalogue

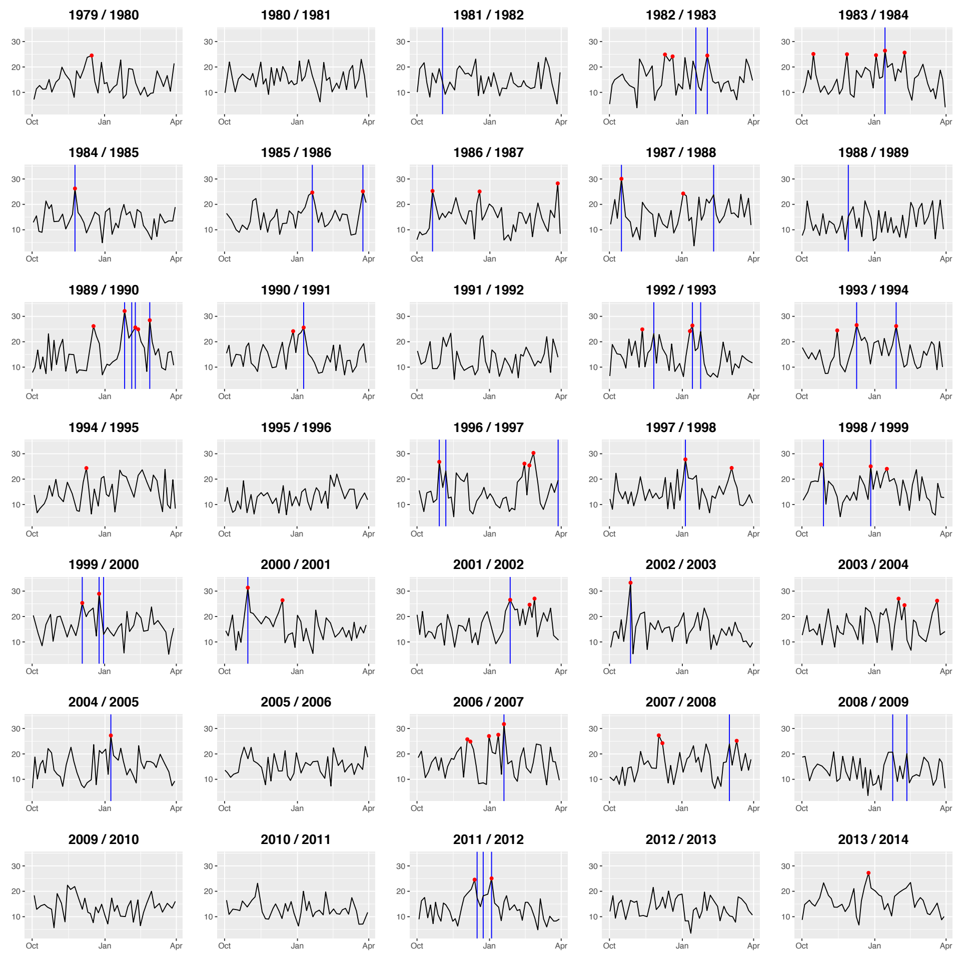


Fig. 6. Declustered risk functional $r(X)(t) = |E_{\mathrm{ABLP}}|^{-1} \int_{E_{\mathrm{ABLP}}} x(s,t)\,\mathrm{d}s$ (ms^{-1}), computed on the ERA–Interim data set for each winter. r-exceedances above the empirical 0.96 quantile are represented by red dots and windstorm starting dates from the XWS catalogue are represented by blue vertical lines.

Modelling $\mathcal{R}$-exceedances

- Generalized $\mathcal{R}$-Pareto processes describe the limiting tail behaviour of **conditional** $\mathcal{R}$-exceedances.
- To apply them to the chosen exceedances in practice, we write

$$\mathrm{P}\left(X \in \cdot\right) = \mathrm{P}\left\{\mathcal{R}(X - b_n) > 0\right\} \times \mathrm{P}\left\{X \in \cdot \mid \mathcal{R}(X - b_n) > 0\right\},$$

  which we approximate by

$$\mathrm{P}\left(X \in \cdot\right) \approx \mathrm{P}\left\{\mathcal{R}(X - b_n) > 0\right\} \times \mathrm{P}\left(P \in \cdot\ \right).$$

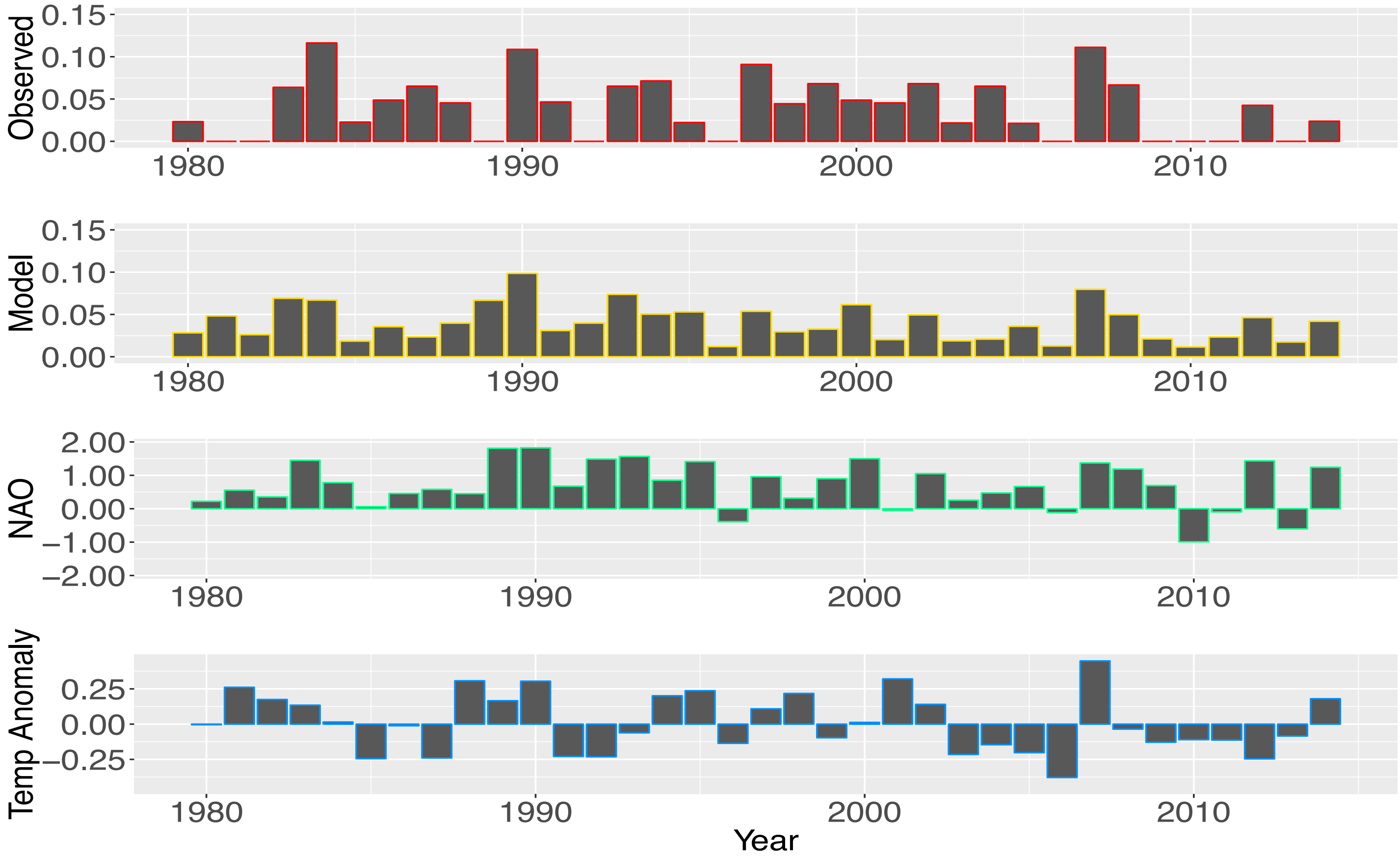
- We use binary modelling for the first term, including dependence on covariates ...

Probability of exceedance

- Here we focus on the first term,

$$\mathrm{P}\,(X \in \cdot) \approx \mathrm{P}\,\{\mathcal{R}(X - b_n) > 0\} \times \mathrm{P}\,(P \in \cdot\,)\,.$$

- Use logistic regression, with covariates NAO and a temperature anomaly:



Extreme wind storms

- For the second term,

$$\mathrm{P}\left(X \in \cdot\right) \approx \mathrm{P}\left\{\mathcal{R}(X - b_n) > 0\right\} \times \mathrm{P}\left(P \in \cdot\,\right).$$

- We must estimate ξ, a and b: we use quantiles to obtain $\widehat{b}$ and then independence likelihood to obtain $\widehat{\xi} = -0.15_{0.01}$ and $\widehat{a}$:

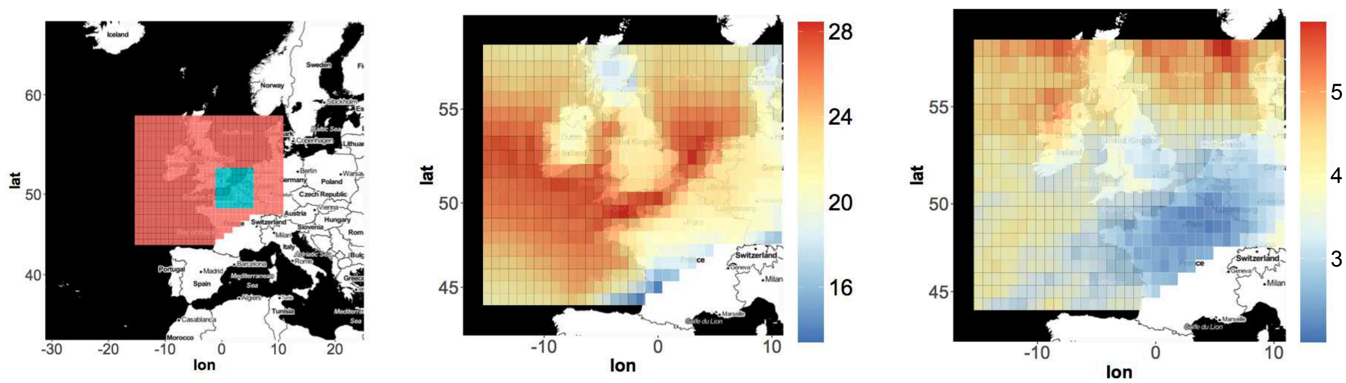


Fig. 5. Left: Study region E (coloured cells) for modelling extreme windstorms over Europe. Mountainous regions were removed to avoid the systematic bias of the reanalysis model. The green cells show the region E_{ABLP} containing Amsterdam, Brussels, London and Paris. Estimated location and scale functions b_n (middle) and a_n (right) (both in ms^{-1}) of the generalized r-Pareto process for modelling extreme windstorms over Europe.

Marginal model fit

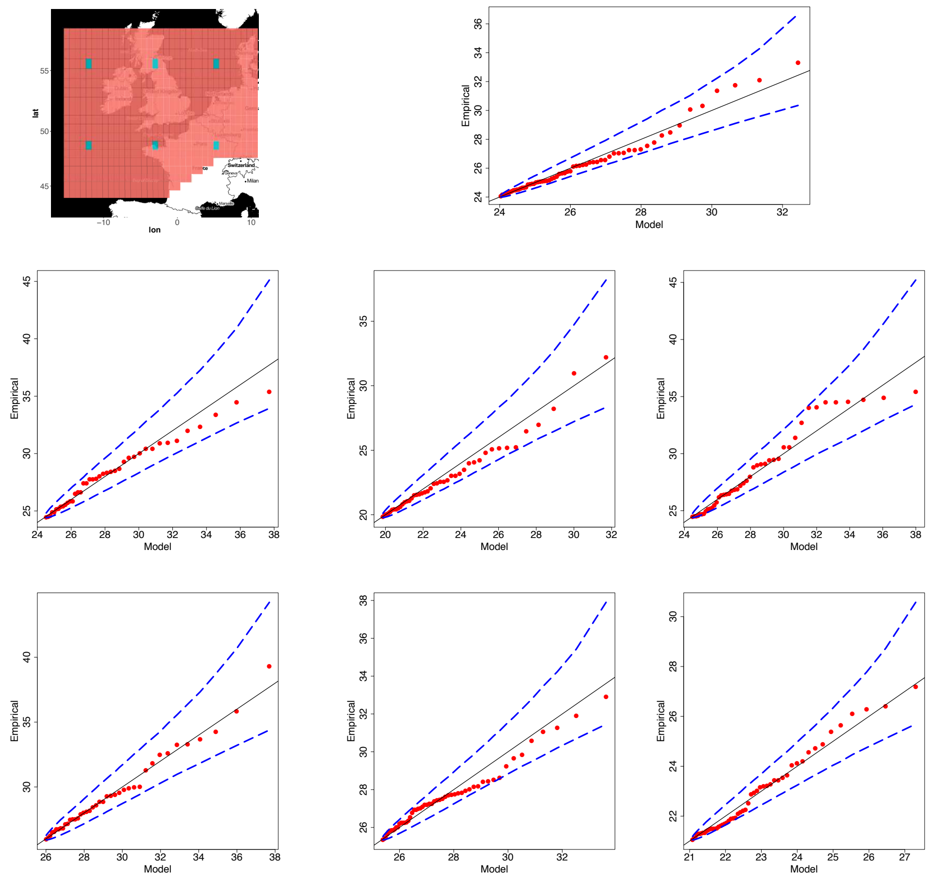


Fig. 8. Model assessment for the windstorm data. The lower six panels show QQ-plots of the local tail distributions for the locations represented by the green cells in the map at the upper left. The thresholds correspond to the local 0.675 quantiles of the r-exceedances, yielding 184 excesses for each cell. The upper right panel QQ-plot is for exceedances of $r(x)$ above the threshold $u_n = 24$ ms^{-1} modelled by a generalized Pareto distribution with scale $\widehat{a}'_n$ and tail index $\widehat{\xi} = -0.15$. The blue dashed lines correspond to pointwise 95% confidence intervals.

Dependence model

- Many possibilities ...
- We model the angular component W using log-Gaussian random functions and Whittle–Matérn semi-variogram

$$\gamma(s, s', t, t') = \kappa \left\{1 - \|h\|^{\nu} K_{\nu}(\|h\|)\right\}, \quad \kappa, \nu > 0,$$

where K_ν is the modified Bessel function of the second kind of order $\nu(=1)$, and the 'distance' between space-time coordinates (s, t) and $s', t')$ is

$$\|h\| = \|h(s, s', t, t')\| = \left\{ \left\| \frac{\Omega(s' - s) - V(t' - t)}{\tau_s} \right\|_2^2 + \left| \frac{t' - t}{\tau_t} \right|^2 \right\}^{1/2},$$

for $s, s' \in E$ and $t, t' \in [0, 24]$, with

- scale parameters $\tau_s, \tau_t > 0$ for the space and time dependence,
- a wind vector $V \in \mathbb{R}^2$ that models the average displacement of the storm in a three-hour period, and
- an anisotropy matrix

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

Model fit

Semi-variogram parameter estimates obtained by least squares applied to the extremogram and using the gradient score. The standard errors (subscripts) are obtained using a block jackknife.

	κ	τ_s(km)	τ_t(h)	a	$\eta(^\circ)$	V_1(km.h)$^{-1}$	V_2(km.h^{-1})
LS	3.5	614	23.8	1.41	-4.12	51.3	14.4
GS	$2.85_{0.01}$	$337_{11.6}$	$9.6_{2.8}$	$1.32_{0.01}$	$21.2_{0.1}$	$50.4_{2.9}$	$12.5_{1.7}$

- Broad agreement on dependence strength at long distances, not on anisotropy:
  - least squares fits long-range north-east anisotropy
  - gradient score fits short-range south-east anisotropy
- Spatial variation of dependence should in principle be modelled.
- Similar estimates of wind vector.

Comparison of extremograms

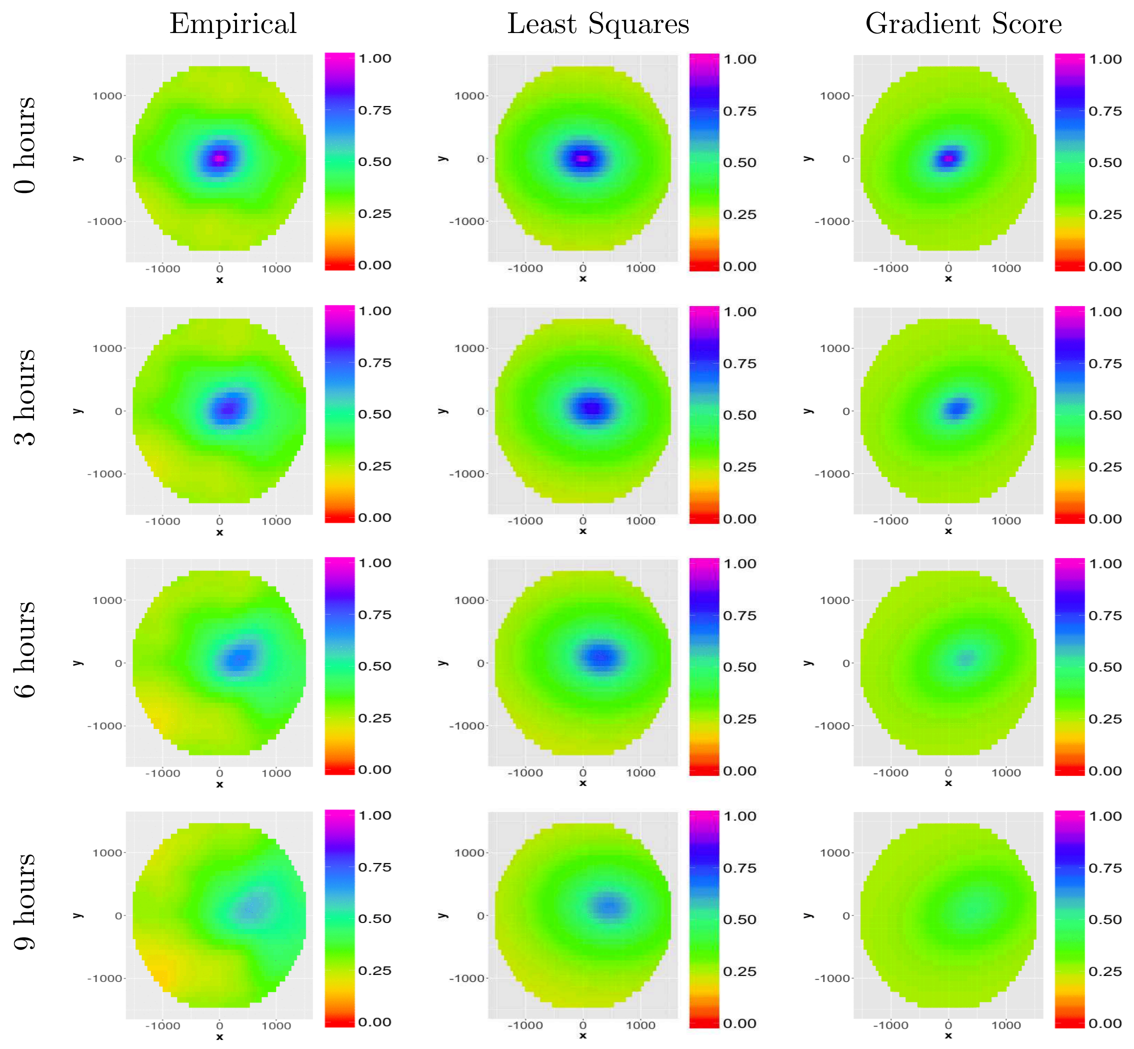


Fig. 9. Extremograms as functions of distance (km): empirical estimates (left), fitted values obtained using the parameters from least squares (middle) and gradient scoring (right) estimates. Each row represents a 3-hour time step.

A simulated storm

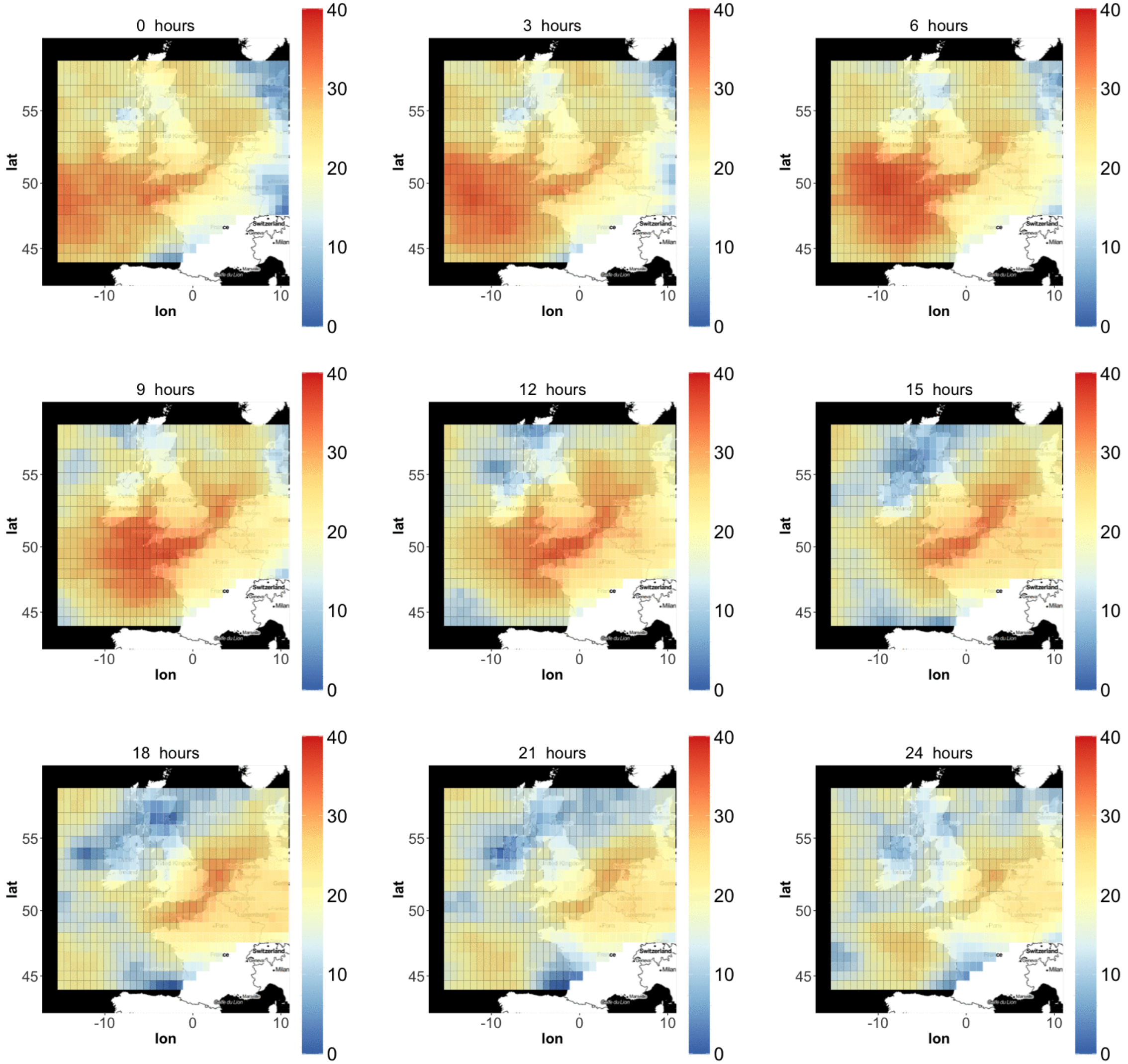


Fig. 10. Simulated maximum speed (ms^{-1}) over the past $3h$ hours of wind gusts sustained for at least 3s. The storm has an intensity $r(x) = 29.1$ ms^{-1}.

Motivation
Functional setting
Application
▷ Closing

Closing

Comments

- Peaks over threshold inference applies to complex settings.
- Need suitably defined risk functionals, many possibilities.
- Can fit such models using (more or less) standard tools
- Model-checking possible, using simulation from fitted models and other techniques.
- For climate change applications, would include causal covariates, maybe fitted using ML techniques?
- Promising general approach for complex problems, but has limitations:
  - common ξ throughout S,
  - scaling functions should satisfy $a_n(s) \approx a'_n A(s)$ for large n,
  - limiting processes are asymptotically dependent (might over-estimate rsome risks),
  - use of parametric models not sufficiently flexible?

Some reading

de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika* **105**, 575–592.

de Fondeville, R. and Davison, A. C. (2022) Functional peaks-over-threshold analysis. *Journal of the Royal Statistical Society, series B* (to appear).

Dombry, C. and Ribatet, M. (2015) Functional regular variations, Pareto processes and peaks over threshold. *Statistics and its Interface* **8**, 9–17.

Engelke, S., de Fondeville, R. and Oesting, M. (2019) Extremal behavior of aggregated data with an application to downscaling. *Biometrika* **106**, 127–144.

Ferreira, A. and de Haan, L. (2014) The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20**, 1717–1737.

Opitz, T. (2013) *Extrêmes multivariés et spatiaux: approches spectrales et modèles elliptiques*. Ph.D. thesis, Université de Montpellier II.