Vers des méthodes statistiques pour une meilleure analyse et prédiction des événements et risques extrêmes

Thomas Opitz

BioSP, INRAE, Avignon

Rencontres INRAE-Inria 2023

Nancy, 5-6 juillet 2023

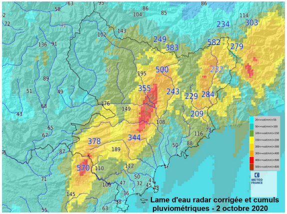
INRAE

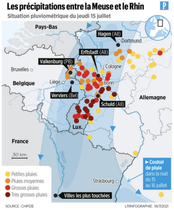
Biostatistique BIO∫Π & Processus Spatiaux

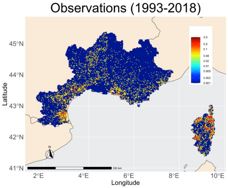
Extreme events

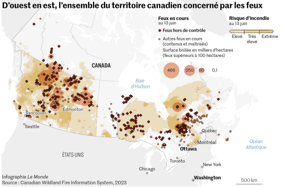
Examples of hydrometeorological hazards.

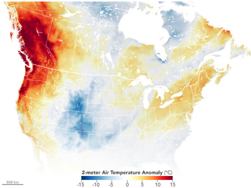
Many research questions: Occurrence probabilities? Short-to-long-term predictability? Stationarity (Climate change)? Compound events?





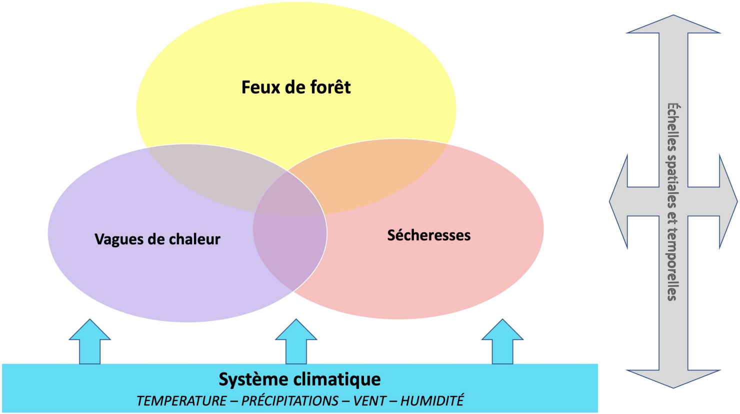






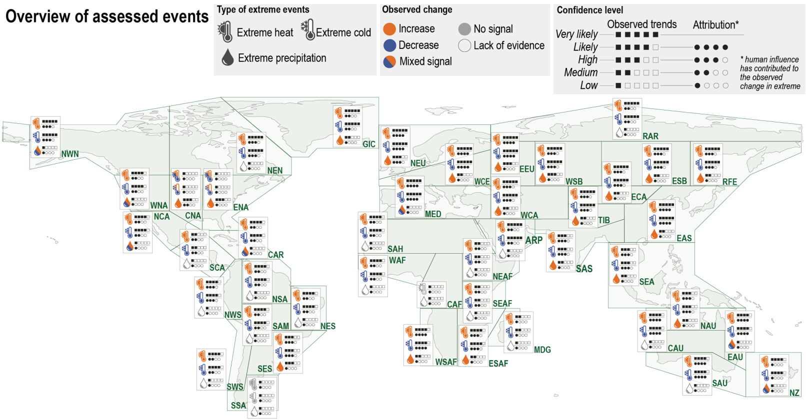
An example of complex compounding of meteorological risk drivers

Standard weather variables are often aggregated into fire-danger indices (FWI etc.)



Extreme events under climate change

IPCC 6th Assessment Report, Chapter 11



Extreme-Value Theory

- **High-Impact Low-Probability** events
- Probabilistic **Extreme-Value Theory:**
  - Regularity assumptions for tails of distributions
  - Probabilistic extrapolation beyond observed extremes
    ($\leftrightarrow$ interpolation in classical spatial statistics)
- Statistical modeling: let the data speak for themselves
  $\Rightarrow$ Bias-variance trade-offs for extreme events
- Include **biophysical and socio-economic drivers** as covariates

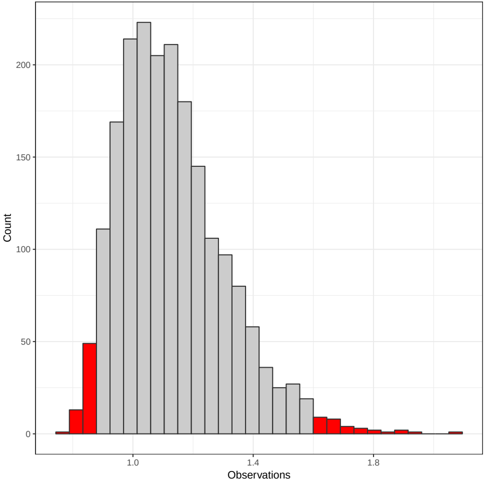
Main purpose

Tools and methods for:

- **Attribution** of extreme impacts to drivers
- **Probabilistic prediction**
- **Stochastic extreme-event generators**

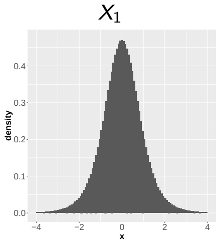
Extreme events are in the tail of the distribution

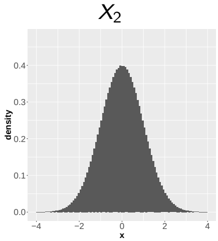
Extreme events are located in the upper or lower **tail of the distribution**.
Without loss of generality, we focus on the upper tail.

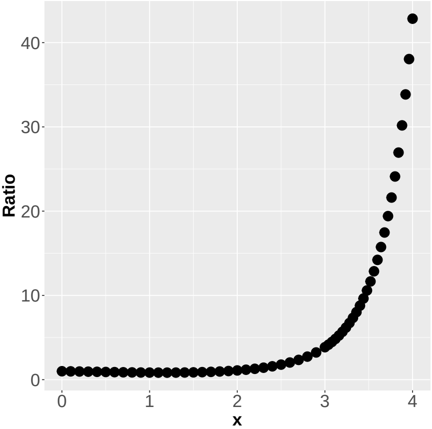


Extreme-event probabilities require specialized models

$X_1 \sim \frac{1}{\sqrt{3/2}} t_6,\ X_2 \sim \mathcal{N}(0,1),\ \mu_1 = \mu_2 = 0,\ \sigma_1^2 = \sigma_2^2 = 1,\ \mathbf{Ratio}(\mathbf{x}) = \frac{\Pr(\mathbf{X_1} > \mathbf{x})}{\Pr(\mathbf{X_2} > \mathbf{x})}$







⚠ **Mean, variance and normal distributions do not well characterize tail events** ⚠

Classical Extreme-Value Theory: Maxima

Fisher–Tippett-Gnedenko Theorem (1928,1943)

Classical asymptotic result for the maximum

Consider a **sample of independent and identically distributed random variables** $X_1, X_2, \ldots, X_n$ and its **maximum** $M_n = \max_{i=1}^n X_i$.

If there exist deterministic sequences $a_n \in \mathbb{R}$ and $b_n > 0$ such that

$$\frac{M_n - a_n}{b_n} \xrightarrow{d} M, \quad n \to \infty,$$

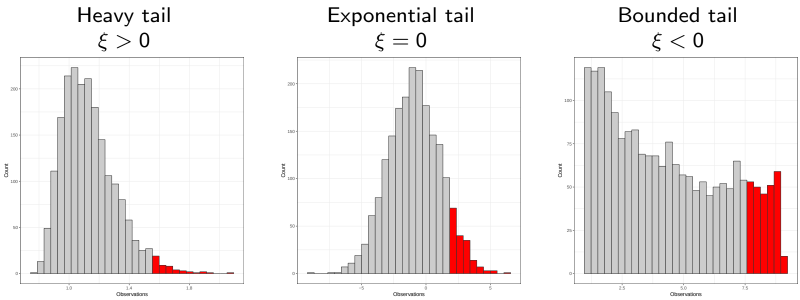
with a nondegenerate limit random variable M, then M has a **max-stable distribution** that is (up to a location-scale transformation),

$$\Pr(M \leq z) = \exp\left(-(1+\xi z)_+^{-1/\xi}\right)$$

with **tail index** ξ.

Tail index ξ determines form of tail

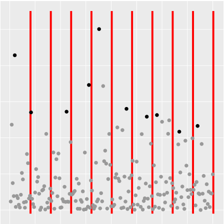
Examples of data histograms for three fundamentally different situations.



Examples of heavy tails (power laws): Wildfires, landslides, precipitation...

Extreme-Value Theory: the trinity of approaches

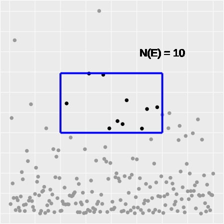
Block maxima



$\Pr(\max_{i=1}^n \tilde{X}_i \leq z)$
$\approx \exp\left(-\Lambda[z,\infty)\right)$

Max-stable distr.

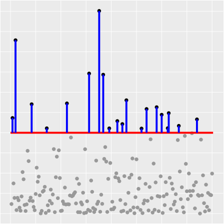
Occurrence counts



$\Pr(N(E) = k)$
$\approx \exp(-\tilde{\Lambda}(E)) \frac{\tilde{\Lambda}(E)^k}{k!}$

Poisson process

Threshold exceedances



$\Pr(\tilde{X}_i - u > y \mid \tilde{X}_i > u)$
$\approx \Lambda[y,\infty)/\Lambda[u,\infty)$

Threshold-stable distr.

Exponent measure Λ **possessing asymptotic stability:**
for any event E and $c > 0$, there are constants $\alpha(c) > 0$, $\beta(c)$ such that

$$c \times \Lambda(E) = \Lambda\left(\frac{E - \beta(c)}{\alpha(c)}\right)$$

ANOVEX project (2023–2024)

ANalysis Of Variability in EXtremes

- 2-year funding from **joint Inria-INRAE call for projects on Environmental Risks**
- Inria-Statify, INRAE-BioSP, INRAE-RECOVER, Avignon Université
- Postdoc 18 months (Chen Yan)

Renaud Barbero (RECOVER)

Stéphane Girard (Statify)

Thomas Opitz (BioSP)

Antoine Usseglio-C. (Avignon Université)

Goals of the ANOVEX project

- **Adapt standard multivariate statistical tools to extreme-event analysis**
  - ANOVA, Principal Component Analysis, regression trees, sensitivity analysis... (based on means, (co)variances and linear operators)
  - Replace Means/variances by **Extreme risk measures** (quantiles, tail means, expectiles)
  - Heavy tails, asymmetry, **non-linear operators** (maximum, exceedance)
- **Applications: hydrometeorological extremes and wildfires under climate change**

First results: ANOVA for extremes

Classical ANOVA

- Test for differences of means μ_j in $j = 1, \ldots, J$ groups
- Test statistic
$$T = \frac{\textbf{Between-Group Variability}}{\textbf{Within-Group Variability}} \overset{H_0}{\sim} F_{df1,df2}$$
- Reject equality of means of $T > F^{-1}_{df1,df2}(\alpha)$

How to test for differences in the few most extremes events of J groups?

$\Rightarrow$ **Use extreme-quantile estimators from Extreme-Value theory!**

ANOVA for heavy-tailed extremes (**ANOVEX**)

- Test for differences in extreme quantiles with null hypothesis

$$\lim_{\alpha \to 1} \frac{q_{j'}(\alpha)}{q_j(\alpha)} \quad \text{for } 1 \leq j < j' \leq J,$$

  where $\Pr(X_j \leq q_j(\alpha)) = \alpha$

- Based on quantile levels $\alpha_\ell = 1 - \ell/n$ with $\ell = 1, \ldots, L$, where n is the sample size
- Replace original observations in ANOVA by **log-quantile estimators** $\log \hat{q}_j(\alpha_\ell)$
- Test statistic

$$T = \frac{\textbf{Same-Quantile Between-Group Variability}}{\textbf{Same-Group Between-Quantile Variability}} \overset{H_0}{\sim} \chi^2_{J-1}$$

Reference:
S. Girard, T. Opitz, A. Usseglio-Carleve. ANOVEX: ANalysis Of Variability for EXtremes. *(Almost submitted...)*

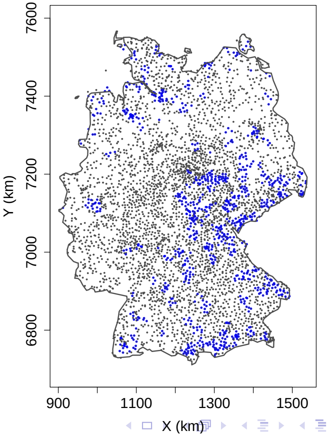
Example: Precipitation extremes (Germany)

Has the distribution of the most extreme precipitation events changed over time?

Here: $J = 6$ decades, for 1901–1960 (few stations) and 1961–2020 (many stations)

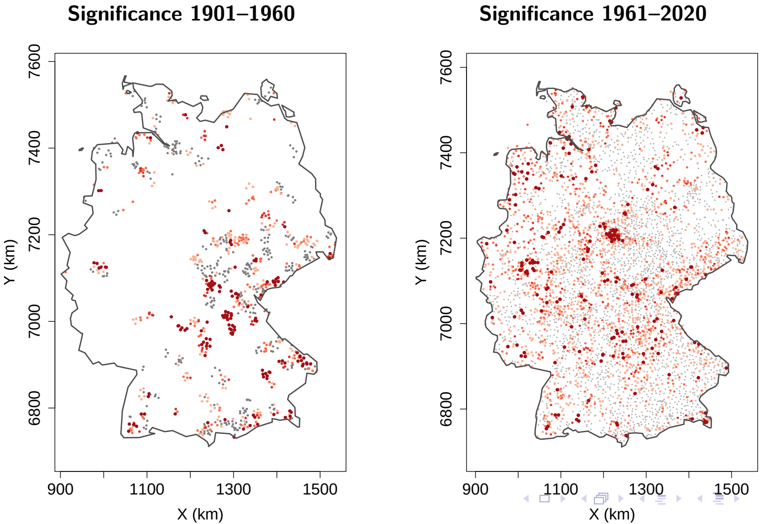
Topography

Rain gauges
(blue: good coverage 1901–1960)

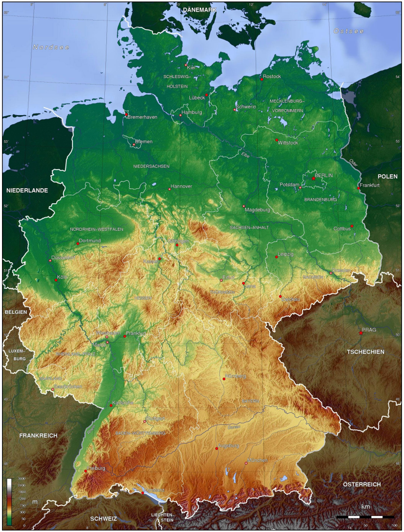


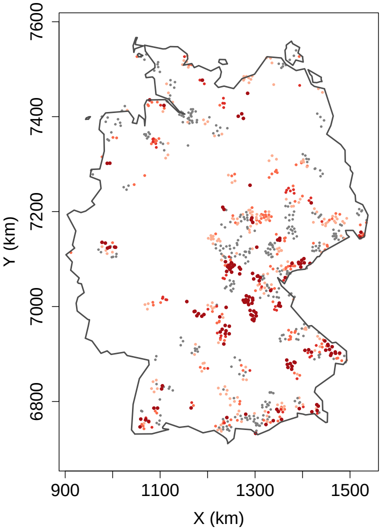
Relatively strong nonstationarities in 1901–1960

- p-values $[0.0, 0.05]$ (dark red); $(0.05, 0.5]$ (lighter red); $(0.5, 1]$ (grey)
- More than 15% of significant gauges for 1901–1960
- Follow-up: Type of nonstationarity? Expert interpretation?

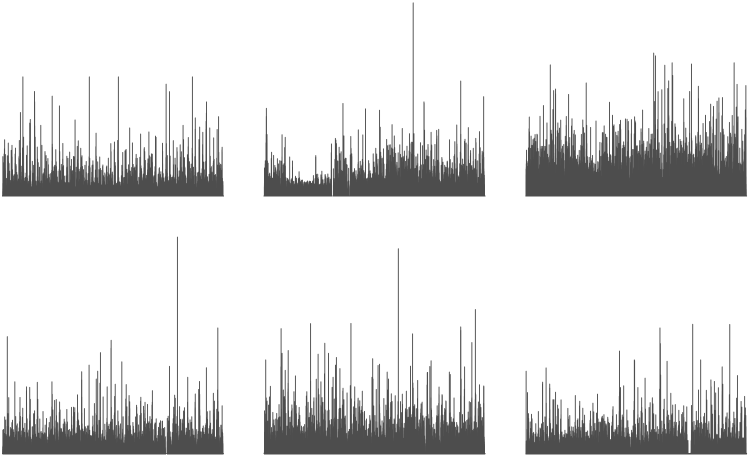


Zoom on 1901–1960



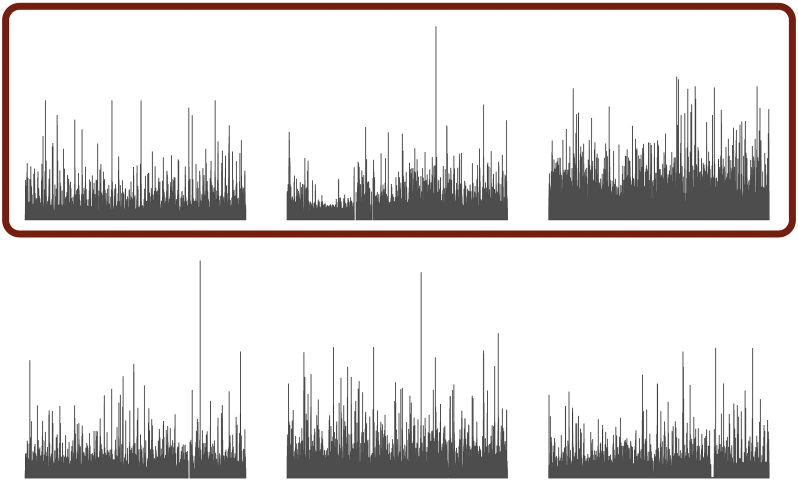


Quiz: Which series are nonstationary?

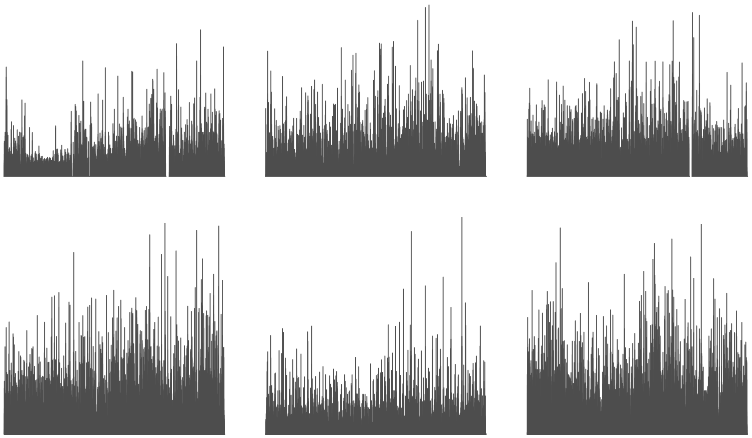


Quiz: Which series are nonstationary? – Solution

More investigation into sources of nonstationarity required (observational vs. physical)

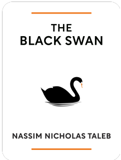


Other examples of nonstationary series



Discussion and outlook

- **Extreme-event data that are both Small and Big**
  - Few independent temporal replicates
  - Recent data with high resolution
    (remote sensing, mobile sensors, participative data, climate models)
  - Need for data fusion and downscaling of extremes
- **Predictability?** Black Swans (Taleb) or Dragon Kings (Sornette)?



- Strong French theoretical community but need more **operational transfer**
- **Some initiatives:**
  - Chair of Geolearning (BioSP-INRAE, Mines-Paris)
  - Coordination of an applied work package in the ANR proposal *EXSTA*
  - RESSTE network of INRAE (Risques, Extrêmes, Statistique Spatiotemporelle)
    `https://reseau-resste.mathnum.inrae.fr/`