Analyzing the dynamics of extreme events with marked point processes

joint work with D. Allard, E. Gabriel and T. Opitz

Antoine Heranval

Postdoctoral researcher at INRAE BioSP
funded by Chaire Geolearning

Natural hazards in 2022 in France

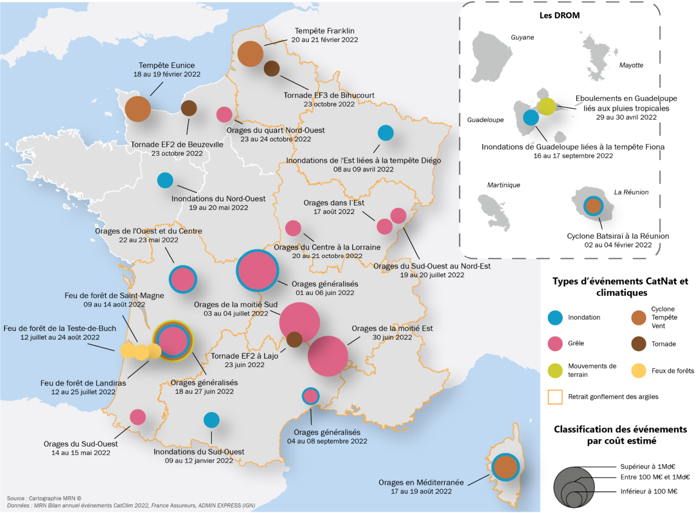


Figure: Source: Mission Risques Naturels

Motivation

- **Context:** Transforming "big data" from climate observation and modeling into actionable knowledge on extreme events and climate change
- **Challenges:** Use conceptual framework and methods from point processes and extreme value theory
- **Objectives:** Understand and model the spatial and temporal (co-)occurrence of extreme events focusing on :
  - event representation,
  - subdivision of geographic space,
  - risk analysis

Marked Point Processes

- A collection of discretely observed stochastic events.
- Each event is characterized by:
  - A **point**: a coordinate in time, space, or space-time.
  - A **mark**: one or more numerical variables (e.g., duration, size of the event or numerical value).
- Examples:
  - Spatial processes: clustered distributions (e.g., urban heat islands) or regular patterns (e.g., wind farm layouts).
  - Temporal processes: clustered events (e.g., rainfall bursts during storms), periodic events (e.g., seasonal floods).
  - Spatio-temporal processes: wildfire occurrences over a landscape with marks representing burned area and duration (Koh et al (2022)), and extreme events in their meteorological drivers (humidity, precipitation, temperature, wind).

Intensity Function

- **Definition:** The intensity function λ represents the average event density over a domain $\mathscr{D}$:

$$\lambda = \frac{\text{Number of events}}{\text{Measure of domain}}$$

  where:
  - Measure could be length, area, volume, or time interval
- **Key Properties:**
  - Generalizes temporal counting to spatial domains
  - Works for 1D (line), 2D (plane), or 3D (volume) spaces
  - Provides occurrence rate of events.
- **Interpretation:**
  - Represents expected number of events per unit measure
  - Allows comparison across different domain sizes

Correlation Function

- **Definition:** The (cumulative) correlation function $K(r)$ measures event distribution relative to a reference point:

$$K(r) = \frac{1}{\lambda} \frac{\mathbb{E}\left[\text{Number of events within radius } r\right]}{|B(0,1)|}$$

Where:

  - $|B(0,1)|$ is the measure of the unit buffer (e.g., length 2 for 1D, area π for 2D)

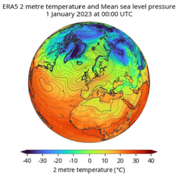
- **Interpretation:**
  - If $K(r)$ grows faster than linear $\rightarrow$ clustering
  - If $K(r)$ grows linearly $\rightarrow$ random distribution
  - If $K(r)$ grows slower $\rightarrow$ event dispersion
- **Extensions:**
  - Marked point processes
  - Cross-correlation between event types, possibly with temporal asymmetry

Point Process of Extreme Events

- Our goal is to analyze extreme events, so the first key question is how to define an extreme event.
- We have data at the scale of individual pixels (0.25° or 1° resolution).
- We choose a threshold for exceedance to identify extreme values.
- We then group the exceeding points using connected component analysis.
- Two numerical applications in mind:
  - One for precipitation events, with a more **temporally** centered approach.
  - One for temperature anomalies, with a more **spatially** centered approach.
- This is still in progress, so please don't hesitate to share your input!

Old Data Analysis

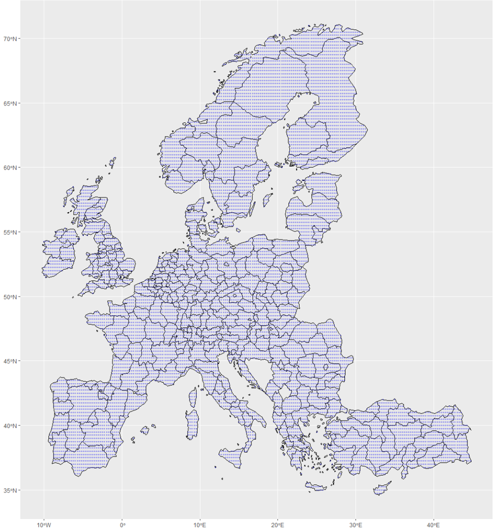
- **Dataset:** ERA5 hourly data on single levels from the Copernicus Data Store.
- **Period:** 2003-01-01 to 2023-12-31.



Variable	**Description**
10m u-component of wind	Eastward wind speed
10m v-component of wind	Northward wind speed
2m temperature	Near-surface air temperature
Mean sea level pressure	Atmospheric pressure at sea level
Sea surface temperature	Ocean surface temperature
Total precipitation	Accumulated rainfall
Evaporation	Water evaporation rate
Total cloud cover	Fractional cloud cover
Convective precipitation	Rainfall from convective systems
Volumetric soil water layer 1	Topsoil moisture
Volumetric soil water layer 4	Deep soil moisture

Regionalisation

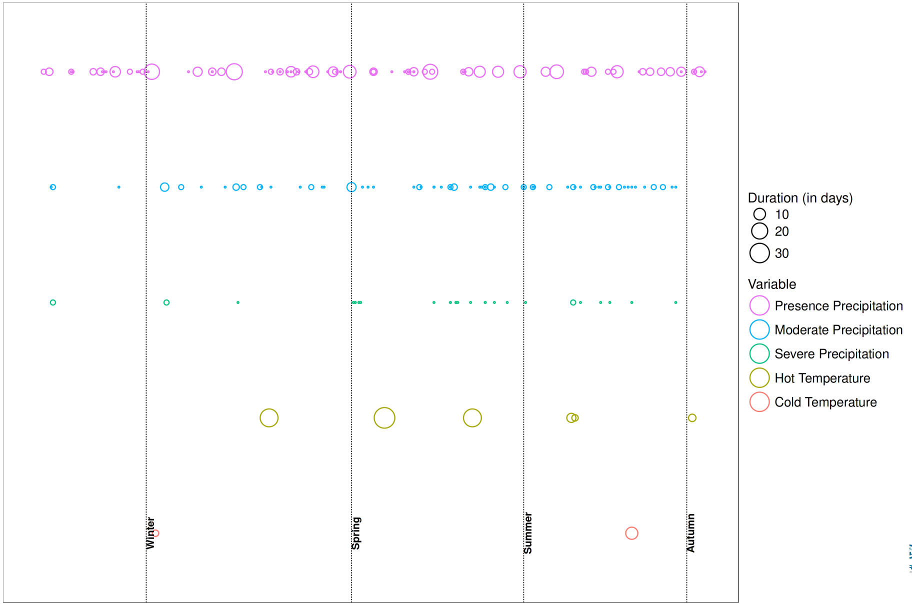
- NUTS level 2 : 332 regions, 11,796 data points.



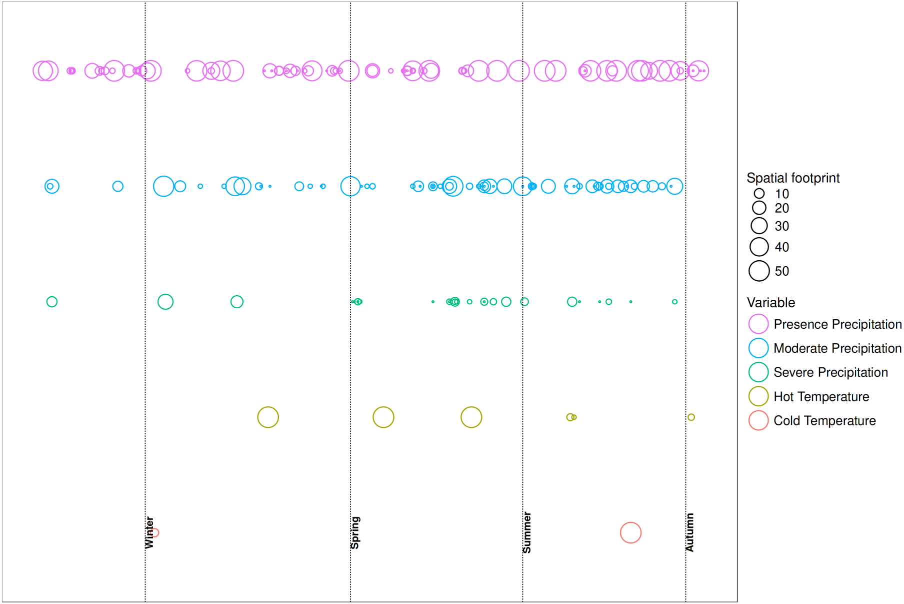
Construction of extreme events

- The variables are transformed into daily measures
- **Precipitation Events:**
  - Presence/absence events (<1mm daily)
  - Moderately events (sum above the 95th percentile with dry<0.2mm)
  - Extreme events (sum above the 99th percentile with dry<0.2mm)
- **Temperature Events:**
  - Seasonal historical mean calculated using ERA5 monthly data from 1940-present
  - Hot events based on **anomalies** (mean daily anomalies above the 95th percentile)
  - Cold events based on **anomalies** (mean daily anomalies below the 5th percentile)
- **Clustering method:** Connected component analysis within each NUTS region.
  - Spatial separation: 50 km.
  - Temporal separation: 3 days (1 day before, 1 day after).

Temporal MPPs for Languedoc-Roussillon in 2022



Idem with spatial footprint



Cross-K Function

- To study multiple point processes, consider a cross-K function
- **Cross K-Function estimation from (Cebrián et al. 2020):**

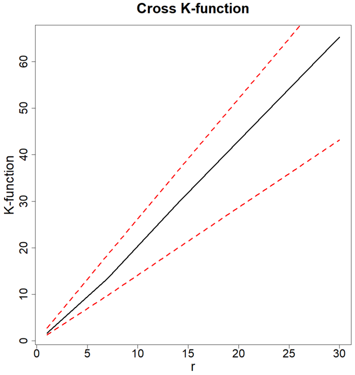
$$\hat{K}_{xy}(r) = \frac{1}{T} \sum_{t_i \in N_x} \sum_{s_j \in N_y} \frac{I(|t_i - s_j| \leq r)}{\lambda_x \lambda_y}$$

- Where:
  - T: Total observation time
  - N_x, N_y: Point sets of processes x and y
  - λ_x, λ_y: Local intensities
  - $0 \leq r < T$: Time interval
- The package `IndTestPP` implements a test based on this method to measure the independence between processes.

Cross-K function for distant NUTS

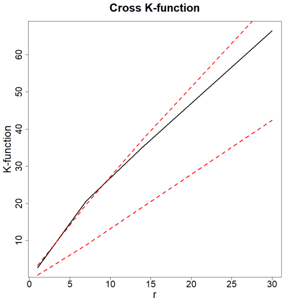
Severe precipitation Languedoc-Roussillon / Northern and Eastern Finland

- If independent, $K(r) = 2r$, up to border effects
- The bounds show 95% confidence interval
- Independence at large distance

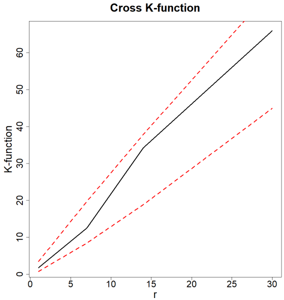


Cross-K function for different variables

Languedoc-Roussillon Severe precipitation / Temperature anomalies



Cold

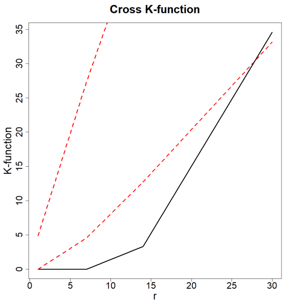


Hot

- Some dependency at short time horizon between cold events and severe precipitation.

Cross-K function for different variables

Finland HOT/COLD anomalies events



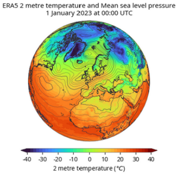
- Negative dependency (ie, repulsion) up to a month for HOT/COLD anomalies

Outlook

- Interesting first analysis, but we aim to go further and develop a complete toolbox for the analysis (initially descriptive) of extreme compound events.
- New data analysis to be able to see more things.
- Improve the work on the Cross-K function to make it symmetric.
- The thresholds (50 km and 3 days) are somewhat arbitrary and could be adapted based on the data.
- Work on extreme value theory and time series analysis could be incorporated, such as the use of extremal indices.

New data analysis

- **Dataset:** ERA5 **daily** data on single levels from the Copernicus Data Store.
- **Period:** 1940-01-01 to 2024-12-31.
- We want to study two periods, 1960-1989 and 1990-2019, to analyze the changes.



Variable	Description
t2m daily max	Max daily temperature at 2m
t2m daily min	Min daily temperature at 2m
t2m daily mean	Mean daily temperature at 2m
Total precip daily sum	Total daily precipitation
10m u component wind daily maximum	Max eastward wind speed at 10m
10m v component wind daily maximum	Max northward wind speed at 10m
CAPE daily max	Max daily Convective Available Potential Energy
Convective precipitation daily sum	Daily precipitation from convective systems

World Spatial Analysis

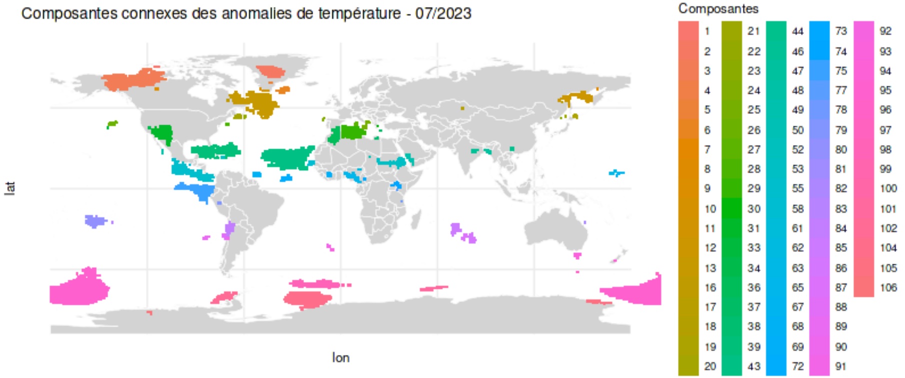
- **Dataset:** ERA5 monthly averaged data on single levels (1940-present).
- **Variables:** 2m temperature.
- **Temporal Resolution:** Monthly (1 017 temporal points over 84 years).
- **Spatial Resolution:** Aggregation at 1° (65 160 points)
- **Anomalies:** Monthly anomalies based on regression quantiles (95)
- We use the `qgam` package to fit a **quantile generalized additive model**
  - A smooth function of `year` using `k = 10` basis functions. This captures long-term trends over time.
  - A cyclic smooth function of `month` using `k = 12` basis functions with cyclic cubic splines. This ensures continuity between December and January to model seasonality properly.

Creation of the points: Patches with terra

Function: `patches` from the `terra` package

- Used to identify and group contiguous areas with the same value in a raster (binary exceedance).
- Here, we use the `8-direction neighborhood`, which includes diagonal connections.
- Average size of patches $350\ 000 km^2$

Anomaly July 2023



Modeling Spatiotemporal Anomalies

- **Representation:** Anomalies defined by spatial location s and time t.
- **Modeling Framework:** Binary variable indicating the presence (1) or absence (0) of an anomaly in each cell.
- **Probabilistic Approach:** Generalized regression model for binary outcomes.

$$Y_i \mid p_i \sim \text{Bernoulli}(p_i)$$

  where p_i represents the probability of observing an anomaly in cell C_i.
- **Key Properties:**
  - $Y_i = 1$ if an anomaly is present, otherwise $Y_i = 0$.
  - Observations in distinct cells are conditionally independent.

INLA Framework

- A natural Bayesian approach is to model the binary outcome using a generalized regression model with a **probit link function**.
- **INLA** (Integrated Nested Laplace Approximation) provides efficient Bayesian inference, offering a faster alternative to MCMC.
- The **SPDE-INLA approach** approximates spatial effects by representing a Gaussian field with a Matérn covariance function as a Gaussian Markov Random Field (GMRF) with a sparse precision matrix.

Bernoulli Model for Anomalies

- We model the presence of an anomaly in each cell C_i as:

$$Y_i \mid p_i \sim \text{Bernoulli}(p_i)$$

- The probit link function relates the probability p_i to the covariates:

$$\Phi^{-1}(p_i) = \beta_0 + f_{\text{YEAR}} + f_{\text{MONTH}} + f_{\text{COV}} + f_{\text{LAT}} + f_{\text{Spatial}}$$

  where Φ denotes the cumulative distribution function of the standard normal distribution.

- Each term $f_{\cdot}$ represents a smooth effect of year, month, covariates, latitude, and spatial structure.
- The parameters follow **hyperpriors** defined in the Bayesian framework via INLA.

Future Directions

- Continue this study and obtain initial results.
- **Long-term Climate Projections and Trends:**
  - Investigate the impacts of climate change on the frequency and intensity of extreme events.
  - Analyze long-term changes in the intensity function of anomalies.
- **Compound Extreme Events Analysis:**
  - Study the joint occurrence of extreme events and develop risk measures using advanced statistical tools.
- **Conditional Extremes Modeling:**
  - Explore multivariate marks associated with threshold exceedances.
- **Ruin Processes in Ecology:**
  - Assess the impact of consecutive extreme events on ecosystems.

Thank You for Your Attention