Identifying coherent homogeneous regions for extreme rainfall.

Philomène Le Gall
Supervised by Clémentine Prieur, Anne-Catherine Favre and Philippe Naveau

28 January 2020

Trajectories
Univ. Grenoble Alpes

LABORATOIRE
JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

Table of contents

1. Introduction
2. Regional Frequency Analysis
3. A threshold-free partitionning
4. Application to Switzerland

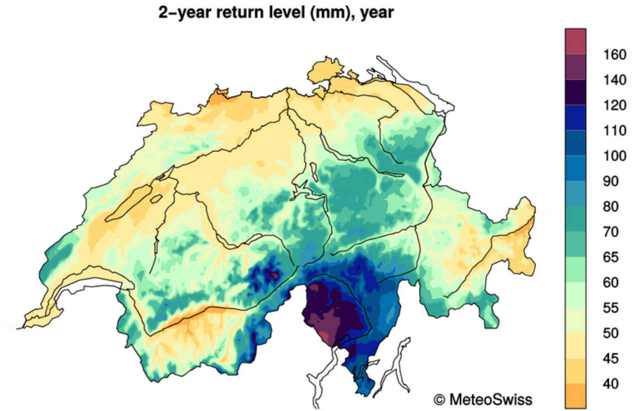
(Flash) flood

- More common natural disaster
- Caused by heavy precipitation (extreme rainfall)

Flood in SW of France

Ex : Maximum observed 412mm/day vs 900mm/year in Grenoble

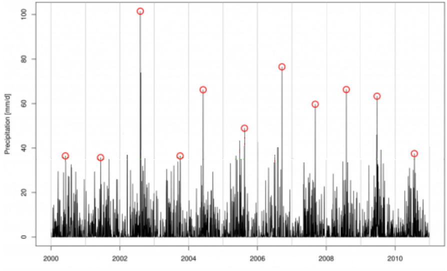
Defining coherent regions for extreme rainfall

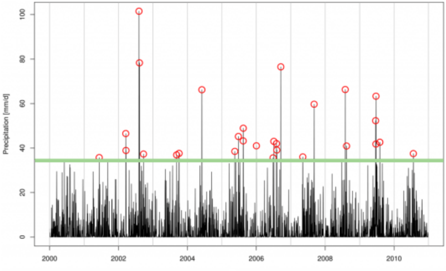


What about longer return period ?

Goal: Estimating **return levels** of unobserved events.

Two approaches: **GEV**(μ, σ, ξ) (maxima) or **GPD**(σ, ξ) (threshold)

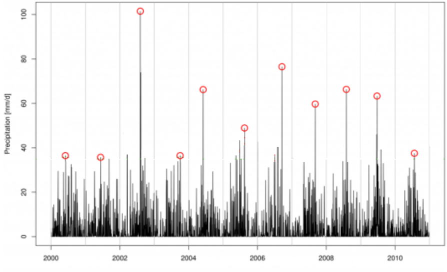


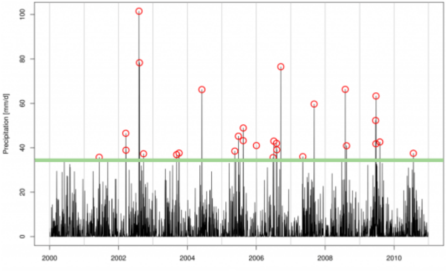


Regional Frequency Analysis Hosking et al. [1985]

Goal: Estimating **return levels** of unobserved events.

Two approaches: **GEV**(μ, σ, ξ) (maxima) or **GPD**(σ, ξ) (threshold)



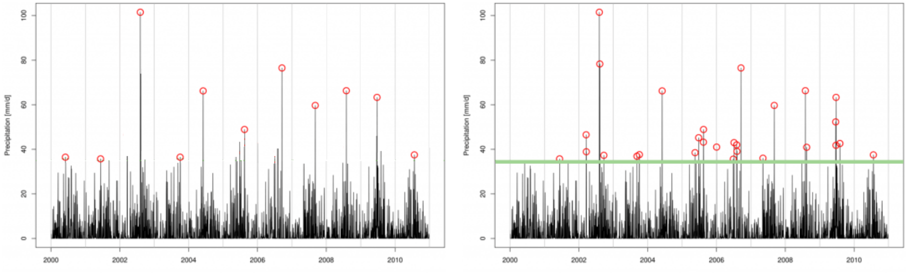


Problems: Asymptotic distributions + lack of information

Regional Frequency Analysis Hosking et al. [1985]

Goal: Estimating **return levels** of unobserved events.

Two approaches: **GEV**(μ, σ, ξ) (maxima) or **GPD**(σ, ξ) (threshold)



Problems: Asymptotic distributions + lack of information

Solution: Regional Frequency Analysis **(RFA)**

Table of contents

1 Introduction

2 Regional Frequency Analysis
- Region of Influence
- Partitionning
- Limits

3 A threshold-free partitionning

4 Application to Switzerland

RFA: What's that ?

Hypothesis: Region is **homogeneous** iff

$$Y_i = \sigma_i Y, \quad \sigma_i \geq 0$$

e.g $Y \sim GEV, GPD, ...$

Goal: Gathering weather stations with same distribution (except scale factor).

RFA: What's that ?

Hypothesis: Region is **homogeneous** iff

$$Y_i = \sigma_i Y, \quad \sigma_i \geq 0$$

e.g $Y \sim GEV, GPD, \ldots$

Goal: Gathering weather stations with same distribution (except scale factor).

Methods available: Region of Influence (**RoI**) vs partitioning.

Regions of influence (RoI) Evin et al. [2016]

Four steps:

Regions of influence **(RoI)** Evin et al. [2016]

Four steps:

1. Choosing "relevant" site-**characteristics** (e.g elevation, coordinates...)

Regions of influence (RoI) Evin et al. [2016]

Four steps:

1. Choosing "relevant" site-**characteristics** (e.g elevation, coordinates...)
2. Gathering stations by their characteristics

Regions of influence (RoI) Evin et al. [2016]

Four steps:

1. Choosing "relevant" site-**characteristics** (e.g elevation, coordinates...)
2. Gathering stations by their characteristics
3. Homogeneity tests

Regions of influence (RoI) Evin et al. [2016]

Four steps:

1. Choosing "relevant" site-**characteristics** (e.g elevation, coordinates...)
2. Gathering stations by their characteristics
3. Homogeneity tests
4. Homogeneous region = **RoI** (Region of Influence)

Regions of influence (RoI) Evin et al. [2016]

Four steps:

1. Choosing "relevant" site-**characteristics** (e.g elevation, coordinates...)
2. Gathering stations by their characteristics
3. Homogeneity tests
4. Homogeneous region = **RoI** (Region of Influence)

Remark

A same station can belong to several RoI.

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model $\rightarrow$ values $>$ 98% quantile

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model $\rightarrow$ values $>$ 98% quantile
2. Estimating first PWM : $\mathbb{E}\left[XF(X)\right]$

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model $\rightarrow$ values $>$ 98% quantile
2. Estimating first PWM : $\mathbb{E}[XF(X)]$
3. Clustering at-site PWM values

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model $\rightarrow$ values $>$ 98% quantile
2. Estimating first PWM : $\mathbb{E}[XF(X)]$
3. Clustering at-site PWM values
4. Homogeneous regions = Clusters

Partitionning into subregions Carreau et al. [2017]

Principle: Gathering stations by measured precipitation.

Four steps:

1. GPD approach/model $\rightarrow$ values $> 98\%$ quantile
2. Estimating first PWM : $\mathbb{E}\left[XF(X)\right]$
3. Clustering at-site PWM values
4. Homogeneous regions = Clusters

Remark

One region per station with this method.

Limits of these approaches

- **RoI**: Choice of the characteristics
- **Partitionning** as in Carreau et al. [2017]: choice of the threshold

Limits of these approaches

- **RoI**: Choice of the characteristics
- **Partitionning** as in Carreau et al. [2017]: choice of the threshold

What's new ?

A threshold-free clustering algorithm.

Table of contents

1. Introduction
2. Regional Frequency Analysis
3. A threshold-free partitionning
   - Probability Weighted Moments
   - Threshold-free model
   - Clustering methods
4. Application to Switzerland

Objective

Identifying homogeneous regions for extreme rainfall.

Methode based on :

- **Partitionning** in homogeneous regions
- **Probability Weighted Moments** (PWM)
- Extended model (no block, no threshold) : **EGPD**

Probability Weighted Moments

Definition

Let $X \underset{c.d.f}{\sim} F$*, a random variable. PWM of order* k *of* X *is defined as*

$$\alpha_k := \mathbb{E}\left[XF(X)^k\right]$$

Probability Weighted Moments

Definition

Let $X \underset{c.d.f}{\sim} F$*, a random variable. PWM of order* k *of* X *is defined as*

$$\alpha_k := \mathbb{E}\left[XF(X)^k\right]$$

Theorem (Guillou et al. [2009])

For $k = 0, 1, 2,$

$$\sqrt{n}\,(\hat{\alpha}_k - \alpha_k) \xrightarrow{d} N(0, \sigma_k^2)$$

where

$$\hat{\alpha}_k := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^k X_{(i)}$$

A threshold-free model Naveau et al. [2016]

Goal : Modeling low, moderate and heavy rainfall intensities (EVT).

A threshold-free model Naveau et al. [2016]

Goal : Modeling low, moderate and heavy rainfall intensities (EVT).

Theorem

The precipitation variable Y can be modeled as

$$Y = \sigma H_{\xi}^{-1}\left[G^{-1}(U)\right], \ \sigma \in \mathbb{R}_{+}^{*}$$

where $U \sim \mathcal{U}[0,1]$, H_{ξ}^{-1} inverse c.d.f of $GPD(1,\xi)$ and G a skewed c.d.f.

A threshold-free model Naveau et al. [2016]

Goal : Modeling low, moderate and heavy rainfall intensities (EVT).

Theorem

The precipitation variable Y can be modeled as

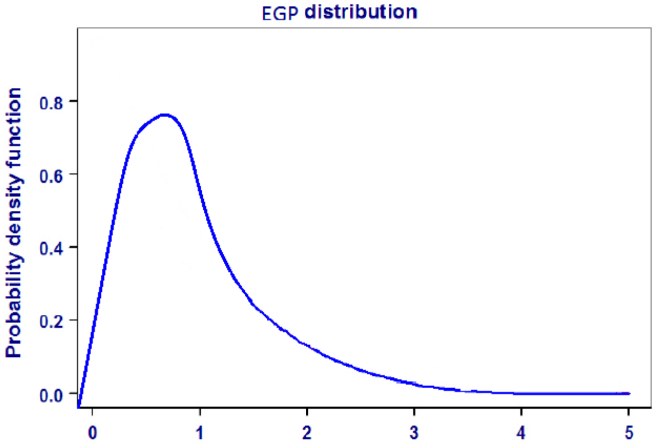
$$Y = \sigma H_{\xi}^{-1}\left[G^{-1}(U)\right], \ \sigma \in \mathbb{R}_{+}^{*}$$

where $U \sim \mathcal{U}[0,1]$, H_{ξ}^{-1} inverse c.d.f of $GPD(1,\xi)$ and G a skewed c.d.f.

Remark

For instance, $G : u \longmapsto u^{\kappa}$ suits.

EGPD density function



EGPD p.d.f

Describing the tail distribution

Definition (Diebolt et al. [2008])

$$R(X) = \frac{3\alpha_2 - \alpha_0}{2\alpha_1 - \alpha_0}$$

Describing the tail distribution

Definition (Diebolt et al. [2008])

$$R(X) = \frac{3\alpha_2 - \alpha_0}{2\alpha_1 - \alpha_0}$$

Theorem (Naveau et al. [2016])

R(EGPD) only depends on κ *and* ξ,

$$R(EGPD) = \frac{3B(3\kappa, 1-\xi) - B(\kappa, 1-\xi)}{2B(2\kappa, 1-\xi) - B(\kappa, 1-\xi)}$$

Method

Reminder: We aim at partitionning space in homogeneous regions for extreme rainfall.

Method

Reminder: We aim at partitionning space in homogeneous regions for extreme rainfall.

1. Sampling (all positive precipitation)

Method

Reminder: We aim at partitionning space in homogeneous regions for extreme rainfall.

1. Sampling (all positive precipitation)
2. Estimating the at-site PWM $\rightarrow \alpha_k$

Method

Reminder: We aim at partitionning space in homogeneous regions for extreme rainfall.

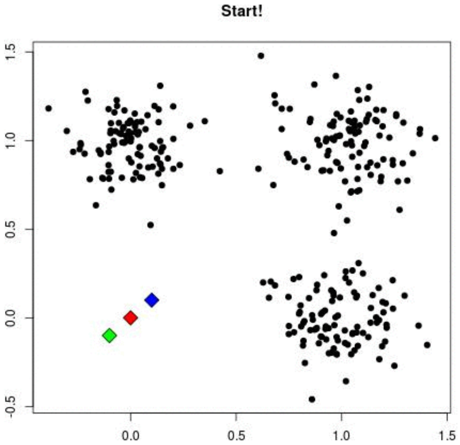
1. Sampling (all positive precipitation)
2. Estimating the at-site PWM $\rightarrow \alpha_k$
3. Estimating the at-site ratio R of PWM $\rightarrow \hat{R} = \dfrac{3\hat{\alpha}_2 - \hat{\alpha}_0}{2\hat{\alpha}_1 - \hat{\alpha}_0}$

Method

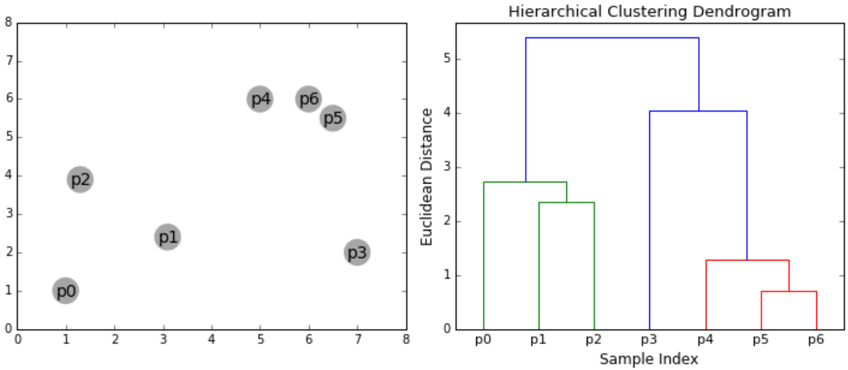
Reminder: We aim at partitionning space in homogeneous regions for extreme rainfall.

1. Sampling (all positive precipitation)
2. Estimating the at-site PWM $\rightarrow \alpha_k$
3. Estimating the at-site ratio R of PWM $\rightarrow \hat{R} = \dfrac{3\hat{\alpha}_2 - \hat{\alpha}_0}{2\hat{\alpha}_1 - \hat{\alpha}_0}$
4. Clustering the at-site ratio estimations: Hierarchical clustering, K-means or K-medoids

K-means Du et al. [2006]



Hierarchical clustering



Battle of clustering methods

	K-means/medoids	**HCA**
Drawbacks	Initialisation dependent	Costly
Advantages	Cheap	Initialisation independent

Remark

K-medoids $\sim$ with centers belonging to the dataset.

Table of contents

1 Introduction

2 Regional Frequency Analysis

3 A threshold-free partitionning

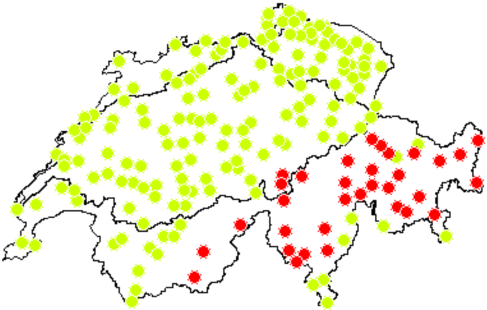
4 Application to Switzerland
- Clustering on marginal distributions
- Clustering on pair-dependence

Application to Switzerland

Dataset: Daily rainfall, 85 years.

Weather stations in Switzerland

Application to Switzerland



Clustered stations by marginal distributions

Limitations

This approach does not take into account:

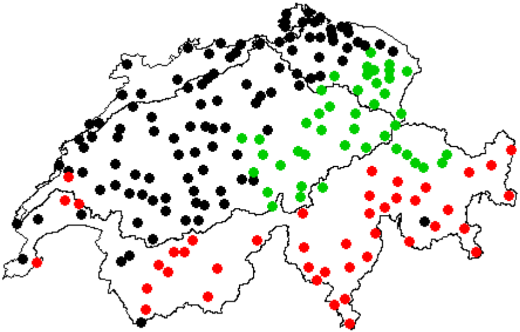
- Spatial dependence
- Temporal dependence
- Non-stationarity

Spatial dependence Bador et al. [2015]

Measure of pair-synchronicity with the **F-madogram**.

Spatial dependence Bador et al. [2015]

Measure of pair-synchronicity with the **F-madogram**.



Clustered stations by spatial dependence

Wanted: ideas

How to combine clustering based on marginal distributions and pair-dependence ?

Wanted: ideas

How to combine clustering based on marginal distributions and pair-dependence ?

Thank you for your attention !

M. Bador, P. Naveau, E. Gilleland, M. Castellà, and T. Arivelo. Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and climate extremes*, 9:17–24, 2015.

J. Carreau, P. Naveau, and L. Neppel. Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research*, 53(5):4407–4426, 2017.

J. Diebolt, A. Guillou, P. Naveau, and P. Ribereau. Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT-Statistical Journal*, 6(1):33–50, 2008.

Q. Du, M. Emelianenko, and L. Ju. Convergence of the lloyd algorithm for computing centroidal voronoi tessellations. *SIAM journal on numerical analysis*, 44(1):102–119, 2006.

G. Evin, J. Blanchet, E. Paquet, F. Garavaglia, and D. Penot. A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology*, 541:1185–1198, 2016.

A. Guillou, P. Naveau, J. Diebolt, and P. Ribereau. Return level bounds for discrete and continuous random variables. *test*, 18(3):584, 2009.

J. Hosking, J. Wallis, and E. Wood. Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. 27:251–261, 08 1985.

P. Naveau, R. Huser, P. Ribereau, and A. Hannart. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769, 2016.