Characterization of homogeneous regions for regional peaks-over-threshold modeling of heavy precipitation

Julie Carreau*, Philippe Naveau† & Luc Neppel*

* Hydrosciences Montpellier, † LSCE, France

Boulder, April 25th
Precipitation data in the French Mediterranean area

332 raingauge stations, 10-57 years of observations, at most 10% NA

High return levels of precipitation for risk assessment
Regional approach combined with peaks-over-threshold

Generalized Pareto (GP) distributed excesses above a threshold:

\[ Y_i \sim G(\sigma_i, \xi_i) \quad i \in \{1, \ldots, M\} \]

Scaling factor and normalized variable: \( m_i \) and \( Z_i = \frac{Y_i}{m_i} \)

Homogeneity assumption: \( Z_i \sim G(\sigma, \xi) \) is identically distributed

Scale invariance and constant shape parameter:

\[ \frac{\sigma_i}{m_i} = \sigma \quad \text{and} \quad \xi_i = \xi \quad \forall i \]

Normalized observations from all the sites contribute to estimate \( \xi \)

Variability in the region is driven by \( m_i \)
**Objective**: propose a characterization of homogeneous regions in the regional peaks-over-threshold framework

Distribution of the excesses follows a GP:

\[ Y_i \sim G(\sigma(x_i), \xi) \]

\( \xi \) a constant regional shape parameter

\( \sigma(x_i) \) varies smoothly as a function of covariates

Regional approach for peaks-over-threshold for a single homogeneous region
Estimation based on **probability weighted moments** (PWM)

Scaling factor: \( \mu_0(x_i) = \mathbb{E}[Y_i] = \frac{\sigma(x_i)}{1-\xi} \)

Normalized variable GP distributed with a single parameter:

\[
Z_i = \frac{Y_i}{\mu_0(x_i)} \sim G(1-\xi, \xi) \text{ identically distributed across sites}
\]

Second PWM of the normalized variable:

\( \mu_1 = \frac{1-\xi}{4-2\xi} \)

\( \mu_1 \) can be estimated with U-Statistics Furrer & Naveau (2007)

GP parameter estimates:

\[
\hat{\xi} = \frac{1-4\hat{\mu}_1}{1-2\hat{\mu}_1} \text{ from all } Z_i, \ i \in \{1, \ldots, M\}
\]

\[
\hat{\sigma}(x_i) = \hat{\mu}_0(x_i)(1 - \hat{\xi})
\]
How to define regions with constant shape parameter?

$C_i \in \{1, \ldots, N_{\text{reg}}\}$ the homogeneous region label for site $i$

$$Y_i \sim G(\sigma(x_i), \xi_{C_i})$$

$\xi_j$ with $j \in \{1, \ldots, N_{\text{reg}}\}$ are the regional shape parameters

Normalized variable:

$$Z_i = \frac{Y_i}{\mu_0(x_i)} \sim G(1 - \xi_{C_i}, \xi_{C_i})$$ identically distributed within a region

Sites $i$ and $j$ belong to the same region if and only if:

$$\xi_{C_i} = \xi_{C_j} \text{ and } \mu_1(x_i) = \mu_1(x_j)$$
Algorithm steps

(1) estimate $\mu_0(x_i)$ for each site $i$

(2) compute the normalized variable $Z_i = \frac{Y_i}{\mu_0(x_i)}$

(3) estimate $\mu_1(x_i)$ for each site $i$

(4) cluster the gauged sites with K-Means

(5) assign regions to the ungauged sites with KNN

(6) estimate GP parameters for each region with PWM estimators

$$\hat{\xi}_j = \frac{1 - 4\hat{\mu}_1}{1 - 2\hat{\mu}_1} \text{ from all } Z_i \text{ such that } C_i = j$$

$$\hat{\sigma}(x_i) = \hat{\mu}_0(x_i)(1 - \hat{\xi}_{C_i})$$
Simulated data example

Sample

Step 1: estimate $\mu_0(\cdot)$

$$x_i = i, \; i = 1, \ldots 1000 \quad n_i = 100 \; \text{GP sample}$$

$$N_{reg} = 4 \; \text{with} \; \xi_1 = 0.3, \; \xi_2 = 0.2, \; \xi_3 = 0.1 \; \text{and} \; \xi_4 = 0$$
Normalized observations and clustering

Step 2: Normalized observations

Step 3-4: estimate $\mu_1(\cdot) +$ clustering
GP parameter estimates and confidence intervals

\[ \hat{\xi}_j = \frac{1-4\hat{\mu}_1}{1-2\hat{\mu}_1} \text{ from all } Z_i \text{ such that } C_i = j \]

\[ \hat{\sigma}(x_i) = \hat{\mu}_0(x_i)(1 - \hat{\xi}_{C_i}) \]

Parametric bootstrap to estimate the sampling distribution (1000 replications)
Precipitation data in the French Mediterranean area

Threshold: 98% quantile of precipitation intensities

Average number of excesses per year
Partitioning into homogeneous regions

Local shape parameter estimate

Partitioning into three regions

Step 5: assign a region to each grid point with KNN
Local shape parameter estimate

Partitioning into four regions
Local shape parameter estimate

Partitioning into six regions
Scale parameter estimate

Local estimate

Difference in scale parameter estimates

\[ \hat{\sigma}(x_i) = \hat{\mu}_0(x_i)(1 - \hat{\xi}_C_i) \]

the difference at 95% of the grid boxes is at most 5 mm
Sampling distribution of the GP parameters at one grid box

Shape parameter

Scale parameter

Return level 95% confidence bands
Conclusion and perspectives

Regional approach for peaks-over-threshold

Estimators based on probability weighted moments

Characterization of homogenous regions stemming straightforwardly

Application with regions of influence

Application on challenging data sets: sparser network with shorter observation period