Quantification of devastating climate events under climate change through novel multivariate bias correction methods

Séminaire Géolearning

13/11/2023, MinesParis

Grégoire Jacquemin

Supervisors : Mathieu Vrac, Denis Allard & Xavier Freulon
Hosting laboratories : Centre de géosciences, Mines Paris & LSCE
Financing organism: Mines Paris, via la Chaire Geolearning INRAE / Mines Paris

LSCE

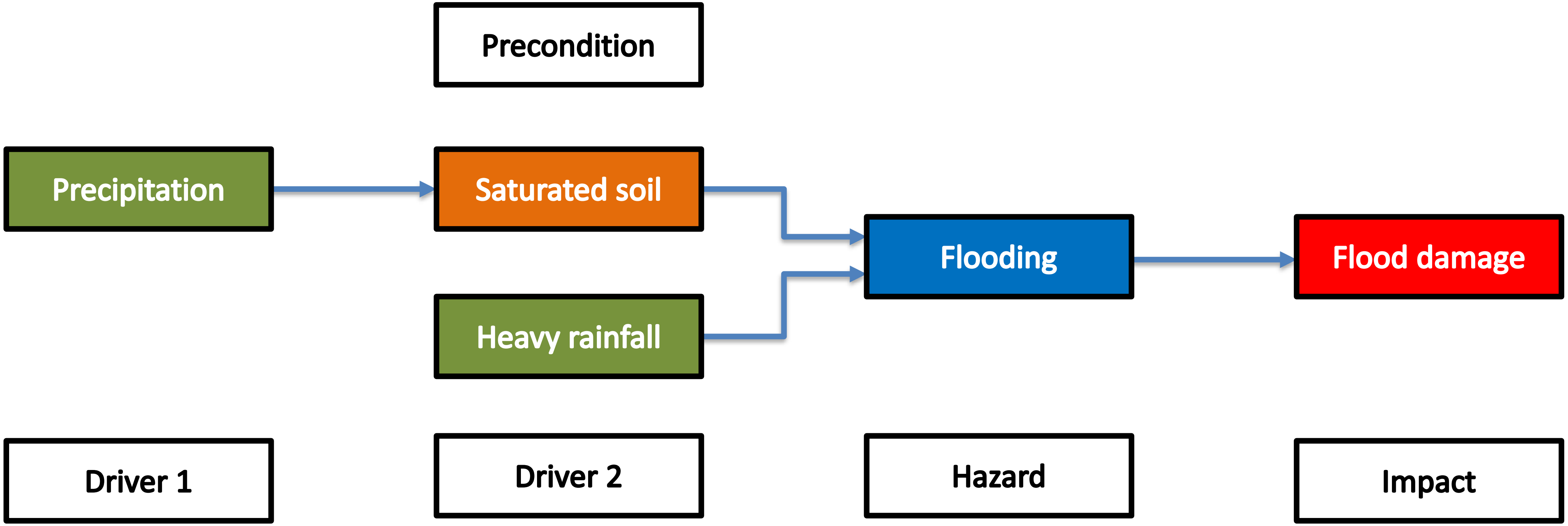
INRAE

GEOLEARNING
CHAIRE /// Data Science for the Environment

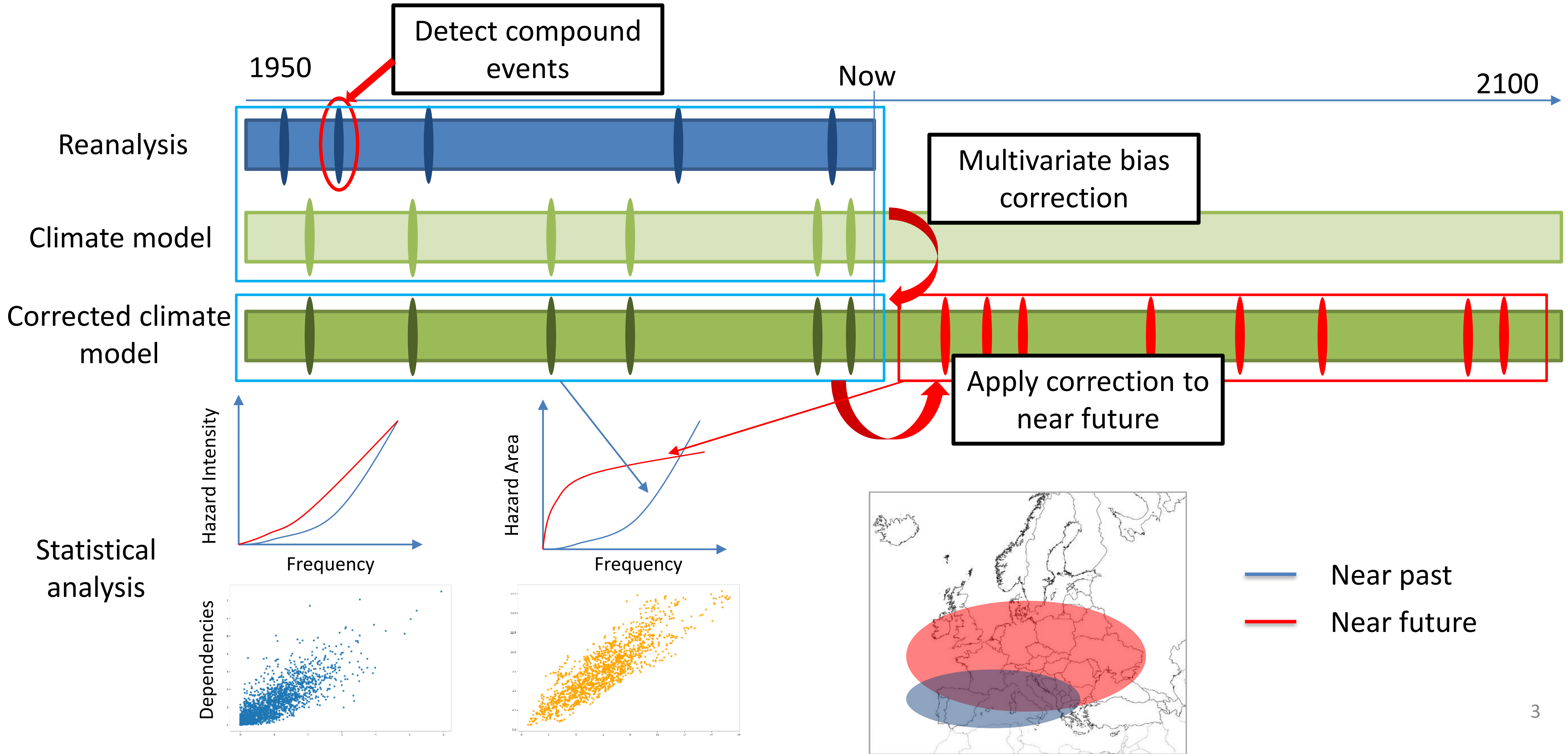
MINES PARIS | PSL

Compound events

A combination of multiple drivers and/or hazards that contributes to societal or environmental risk (Zscheischler et al., 2020)



PHD Objective



Modelling the dependence

- With climate change, or in the simulations, the marginals and the dependence structure can change.

- Multivariate bias correction is probably necessary to affect both the marginals and the dependence.

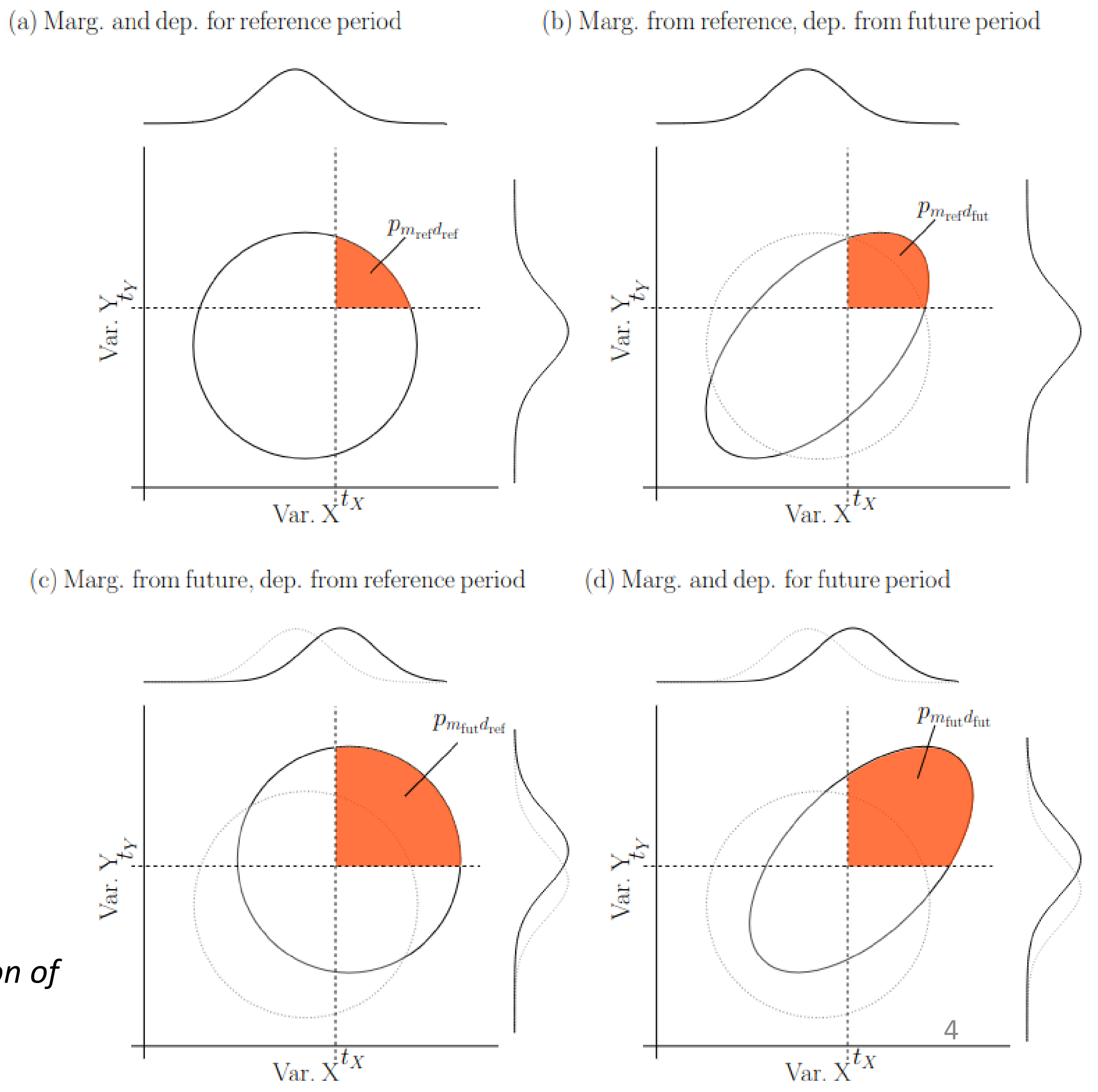


Figure from B. François and M. Vrac, *Time of emergence of compound events: contribution of univariate and dependence properties*

Multivariate extreme value theory

- X_N and Y_N are two samples with bivariate cumulative distribution function **F**, supposedly in the domain of attraction of a bivariate extreme value cumulative distribution function **G**.

- Sklar's theorem (1959) : Any multivariate cumulative distribution function **F** can be expressed in terms of its margins F_i and a copula **C** : $\boldsymbol{F}(x_1, \ldots, x_d) = \boldsymbol{C}(F_1(x_1), \ldots, F_d(x_d))$ with **C** : $[0,1]^d \rightarrow [0,1]$

- Therefore, we get weak convergence of the marginal distribution functions and the copula function

Copulas and uniform margins

- This allows us to propose the following approach :

1. Propose a univariate extreme model for the marginals
2. Reduce to uniform margins
3. Determine the copula

- X_N and Y_N must be i.i.d. and extreme, and (X_N, Y_N) must be extreme in some sense

Three potential events

- July 2021 Belgian/German flooding

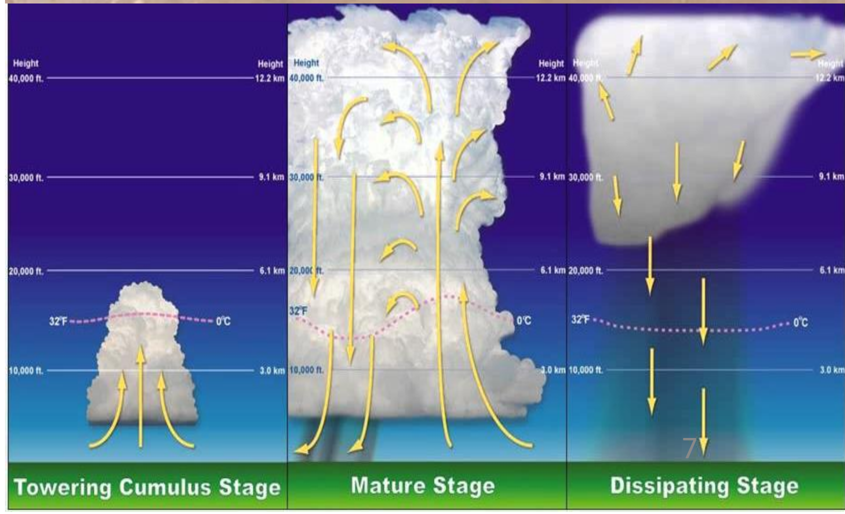
(Preconditioned event)

- May/June 2016 French flooding

(Spatially compound event)

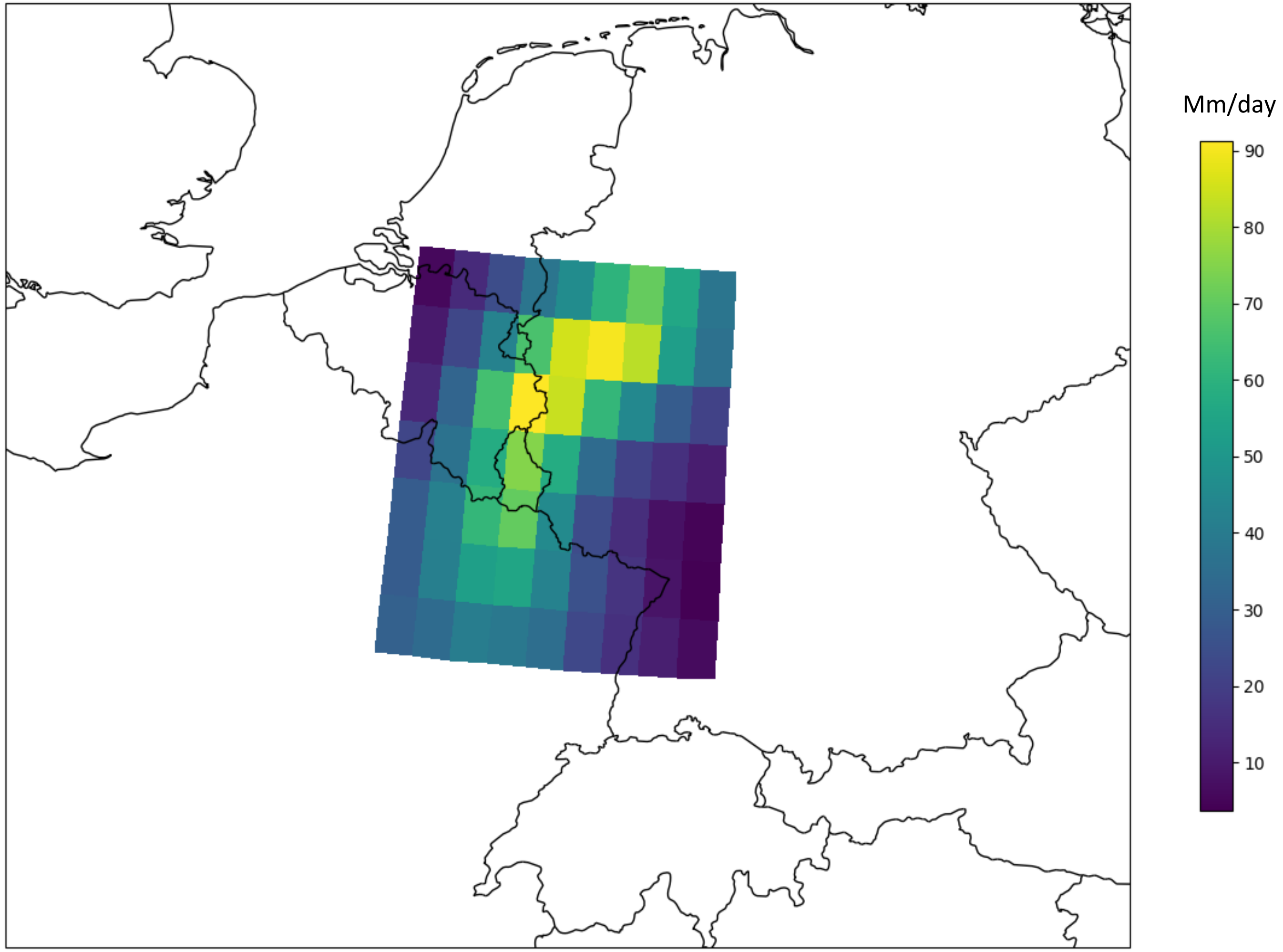
- Convective cells

(Multivariate compound event)



14th July 2021 flooding

- Data from ERA5 reanalyzes
- Total Precipitation (TP) : daily precipitation (mm/day)
- Antecedent Precipitation Index (API) : $API_j = \sum_{i=1}^{i=N} k^{i-1} * TP_{j-i}$ with k=0.9 and N=30



Data Selection

Data must be i.i.d. and extreme

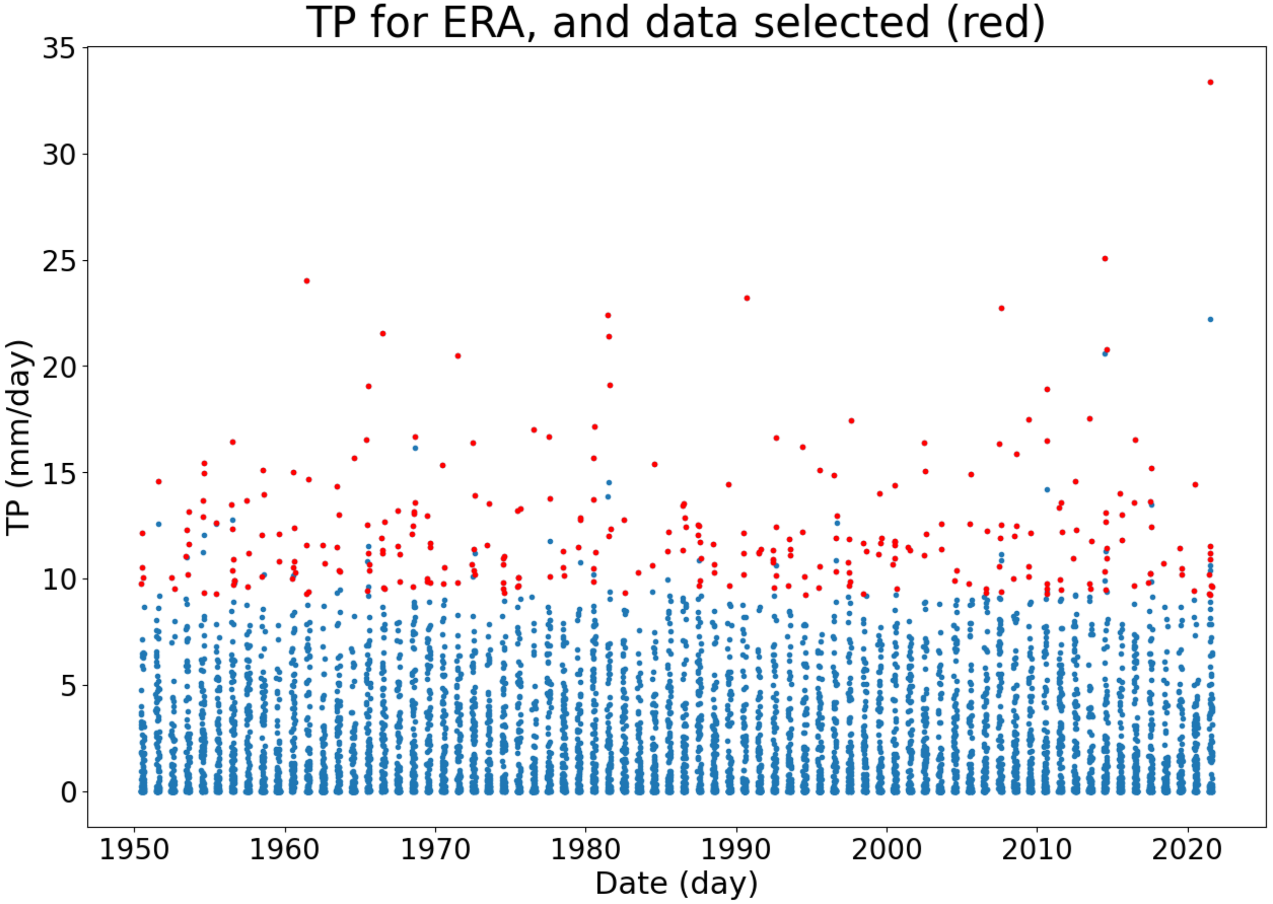
Data selection for 1D

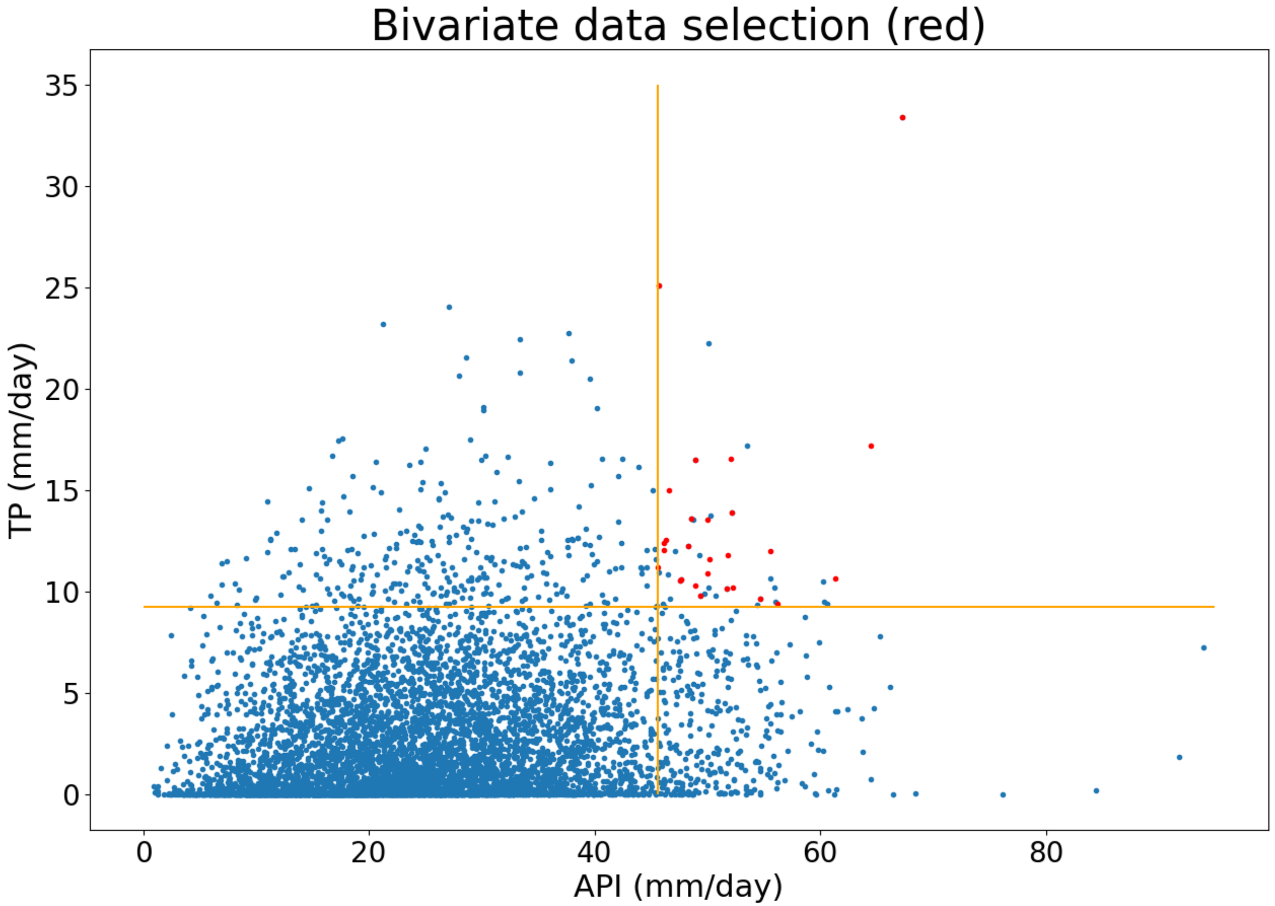
- For TP : select points above the 95th quantile, separated by at least 2 days
- For API : select points above the 95th quantile, separated by at least D days, with : $\rho\left(API_j, API_{j+D}\right) < 0{,}1$ (D = 20)

Data selection for 2D

- Select (TP_i, API_j) with $TP_i >$ $Q95_{TP}, API_j > Q95_{API}$ and $i - 5 \leq j \leq i$
- Then select couples separated by at least D days, according to the highest TP value

Data selection





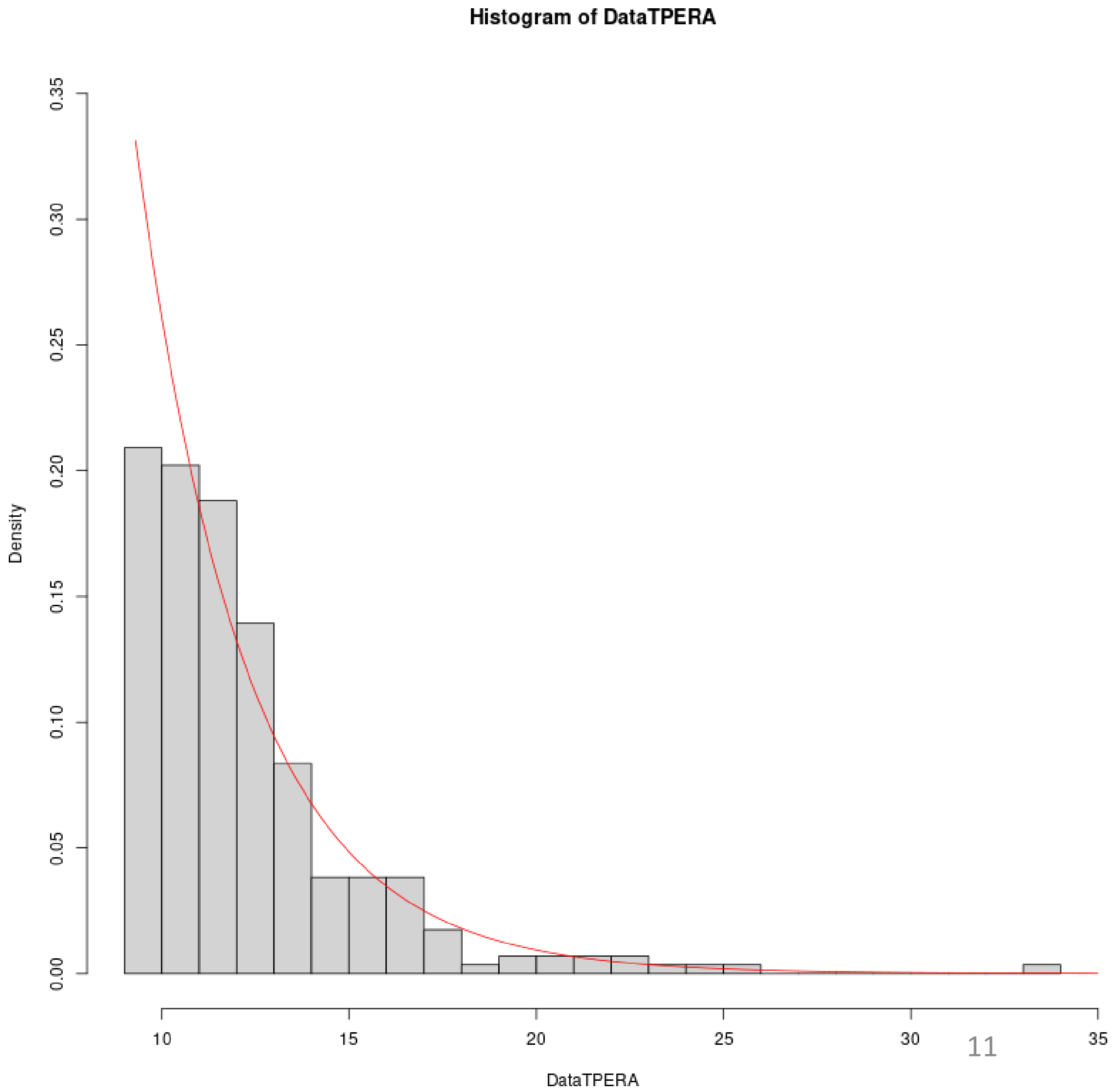
Generalized Pareto Distribution model

With the univariate data selection, we can use a Generalized Pareto Distribution (GPD) model :

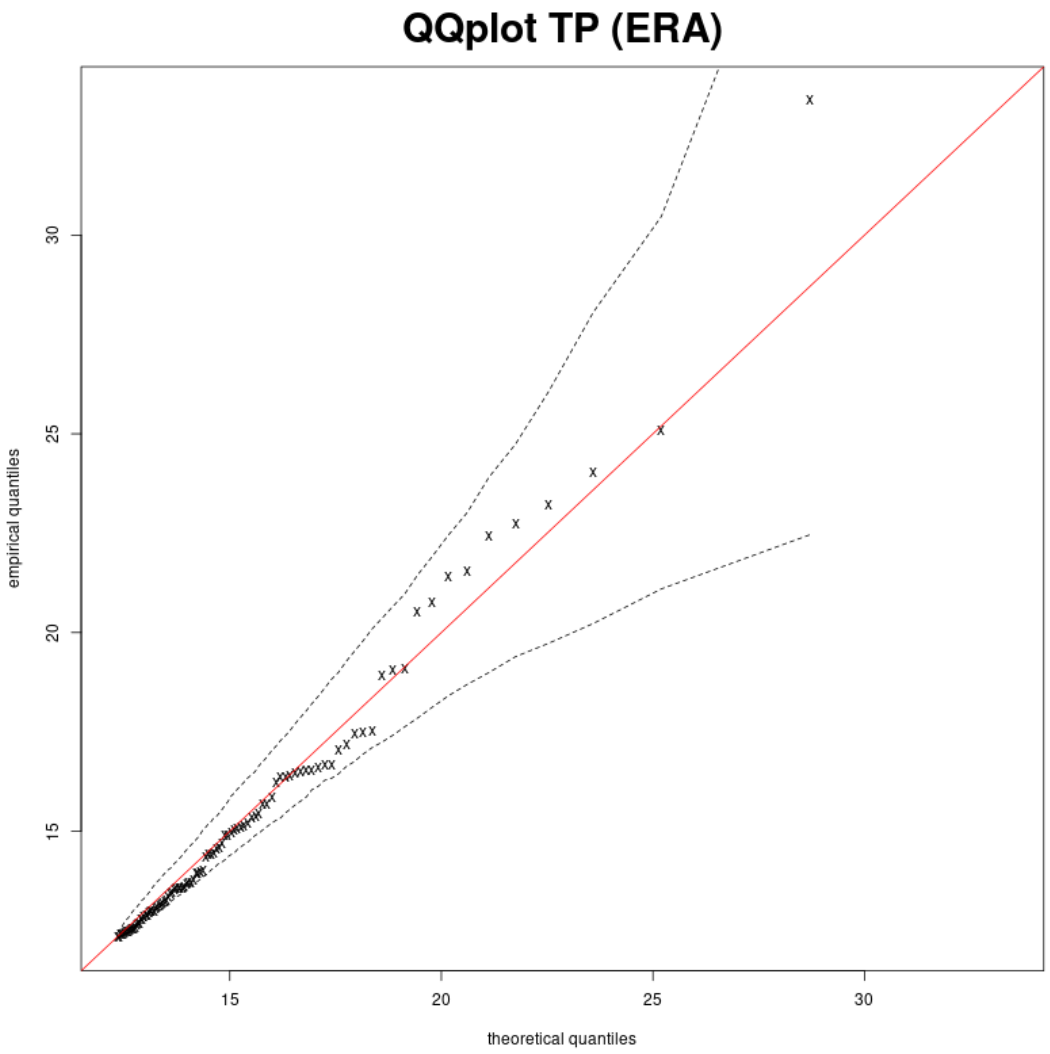
$$F(x) = 1 - (1 + \xi x)^{\frac{-1}{\xi}}$$

with $x \geq 0$ and $\xi \neq 0$

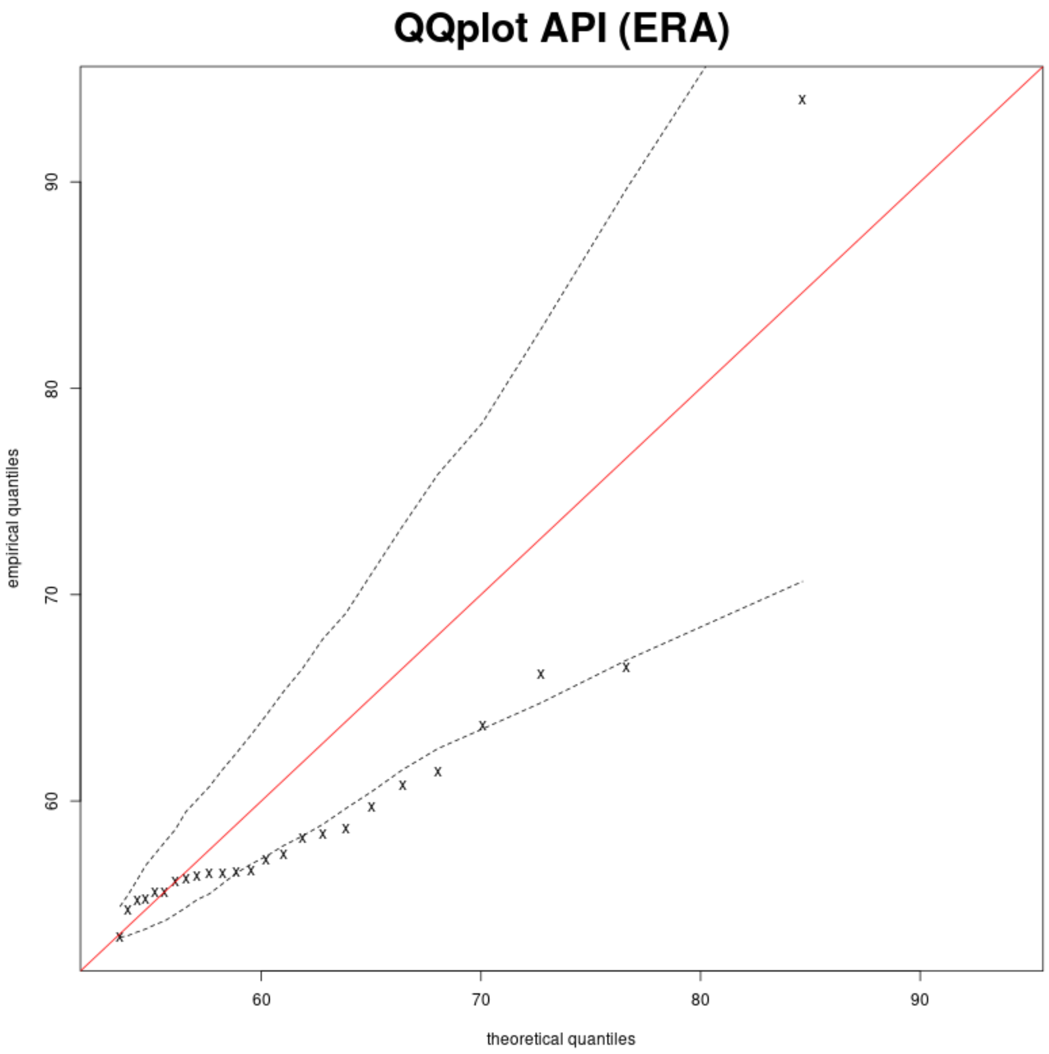
Parameters are estimated through maximum likelihood method



Quantile plots of GPD adjustment



$\hat{\xi} = 0{,}016\ (0{,}060)$

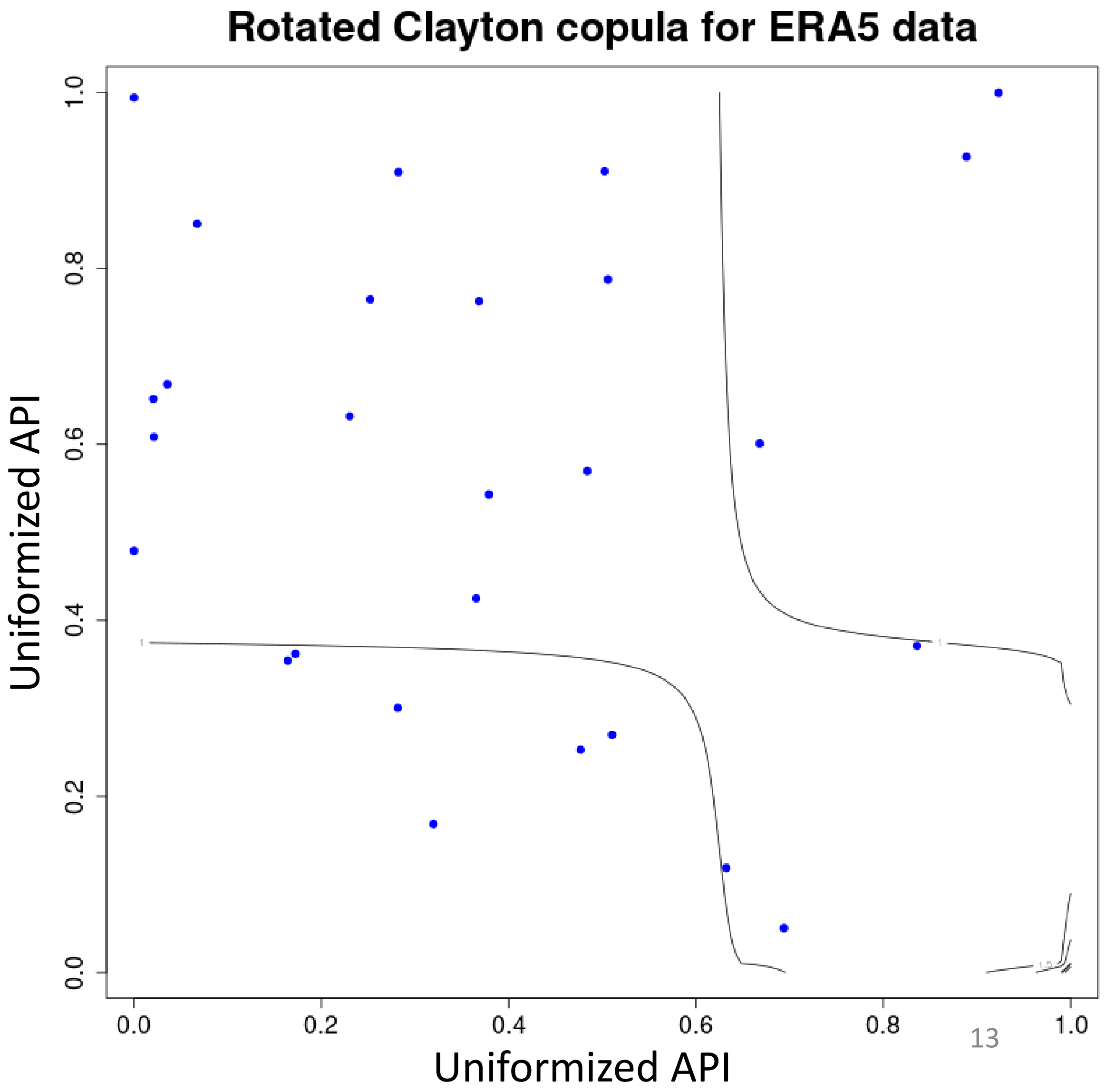


$\hat{\xi} = -0{,}063\ (0{,}130)$

Copula model estimation

Use maximum likelihood to estimate the parameters of all the copulas from the selection : Gaussian, student, Archimedean

Then select the best copula according to the Akaike Information Criteria (AIC)



Coefficients of extremal dependence

For (U, V) uniform r.v., we define χ : $\chi = \lim_{u \to 1} P(V > u | U > u)$

and $P(V > u | U > u) \approx 2 - \frac{\log C(u,u)}{\log u}$

Similarly, we can define $\bar{\chi}$: $\bar{\chi} = \lim_{u \to 1} \frac{2\log(1-u)}{\log \bar{C}(u,u)} - 1$

With $\bar{C}(u, v) = 1 - u - v + C(u, v)$

Here, we have : (χ, $\bar{\chi}$) = (0, -0.019) ---> asymptotic independence, close to total independence

Return periods

- Univariate return period = inverse of the probability to exceed a determined threshold :

$$T(x_{14.07}) = \frac{1/n}{1 - P(X \leq x_{14.07})}$$

- When describing a bivariate event by a joint exceedance (AND), the return period is defined by :

$$T_B(TP_{14.07}, API_{14.07}) = \frac{1/n}{1 - U_{TP} - U_{API} + C(U_{TP}, U_{API})}$$

with $U_X = F(x_{14.07})$ and C the copula

CDF-t correction

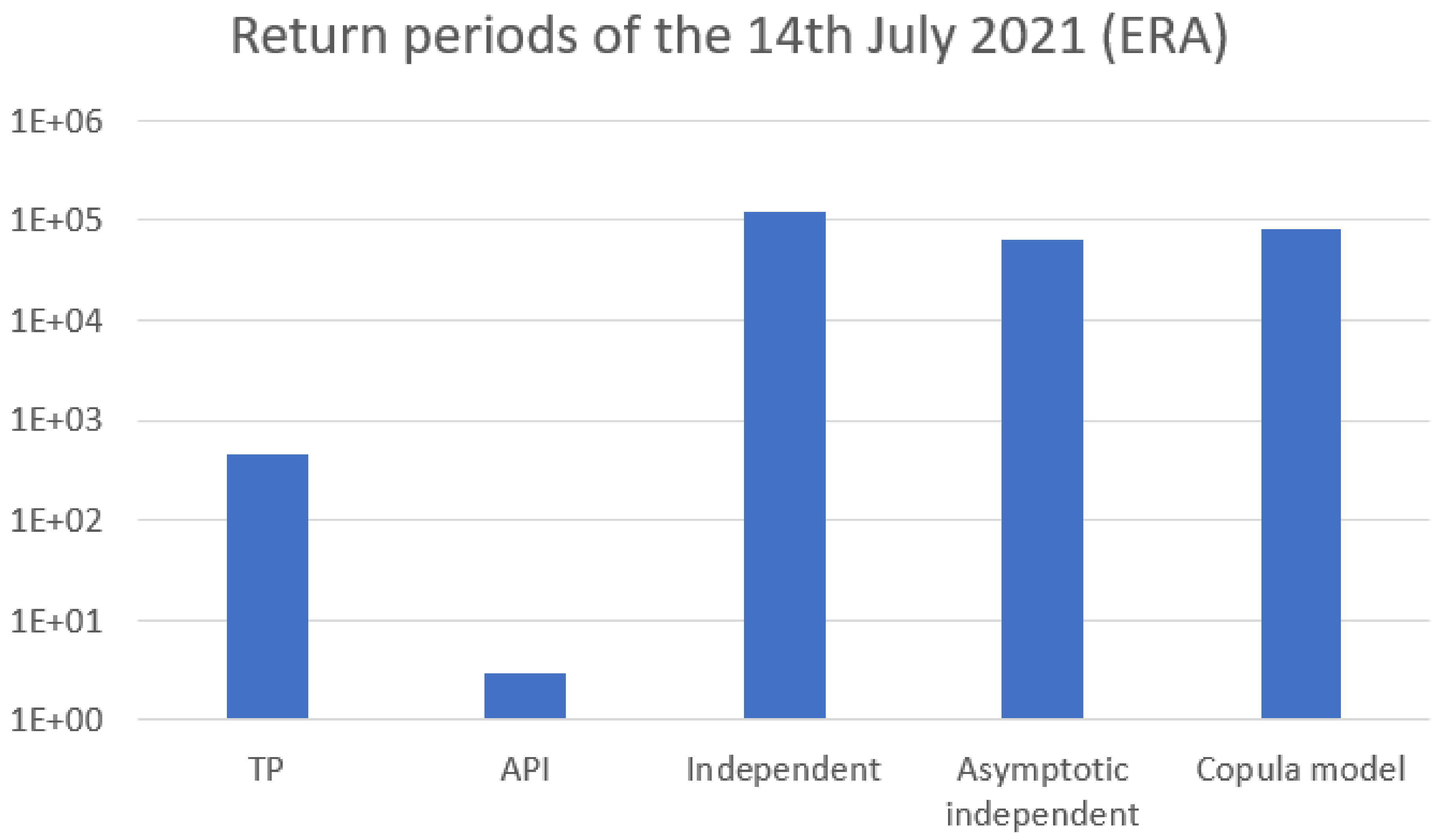
- We apply the same statistical treatment to CMIP-6 Historic data (1950-2021) and CMIP-6 Projection data (2022-2100)
- For the moment, we have considered only the IPSL model, low resolution (ssp585)

CDF-t	Historic	Projection
Model (CMIP-6)	$F_{CMIPHist}$ ↓ T	$F_{CMIPProj}$ ↓ T
Reference (ERA5)	F_{ERA}	$F_{Corrected}$

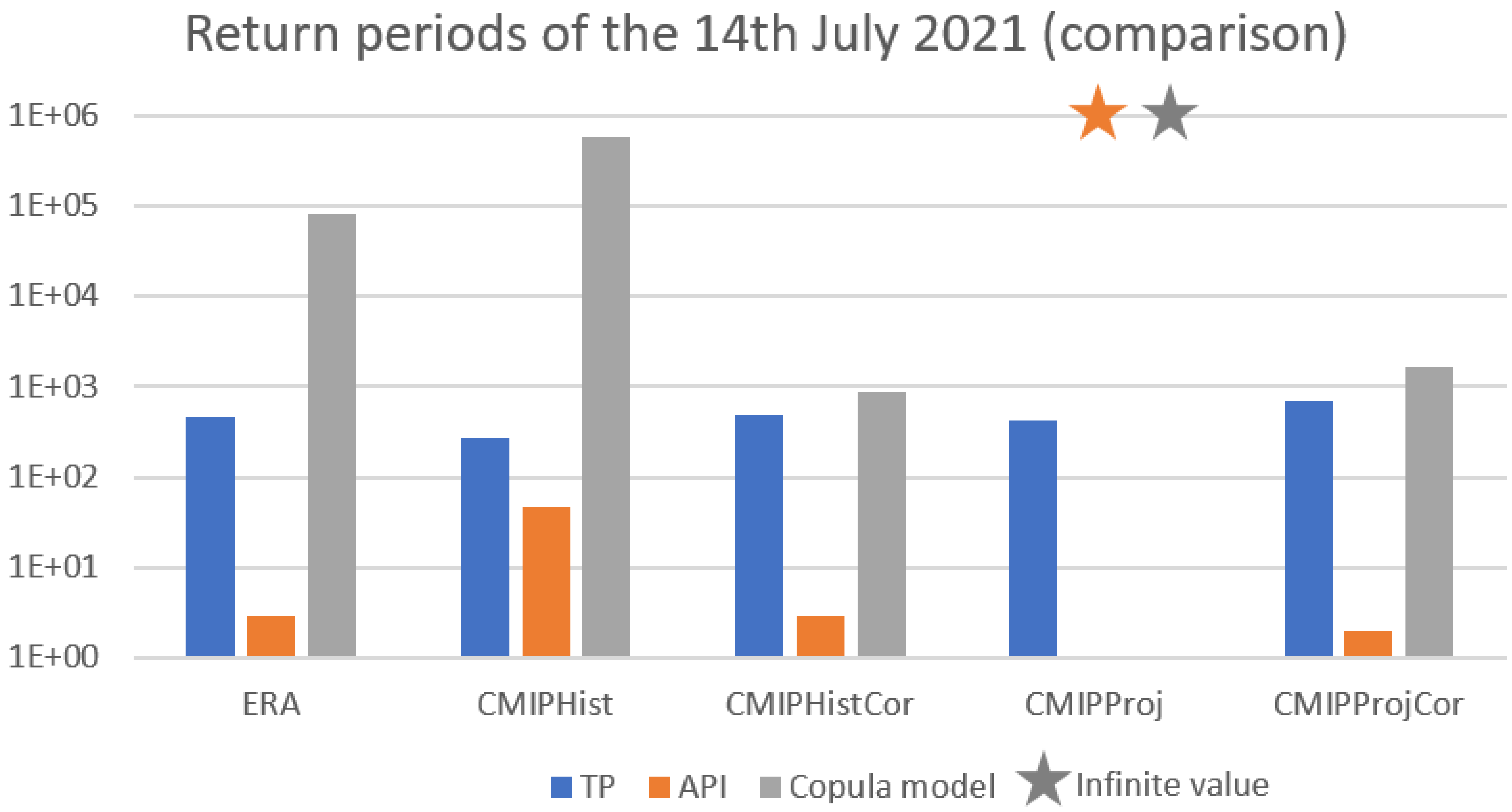
$$F_{Corrected}(x) = F_{ERA}(F^{-1}{}_{CMIPHist}\left(F_{CMIPProj}(x)\right))$$

- We get the corrected CDF, and then we perform a quantile-quantile correction between the corrected CDF and the projection data

Return periods results



Return periods results

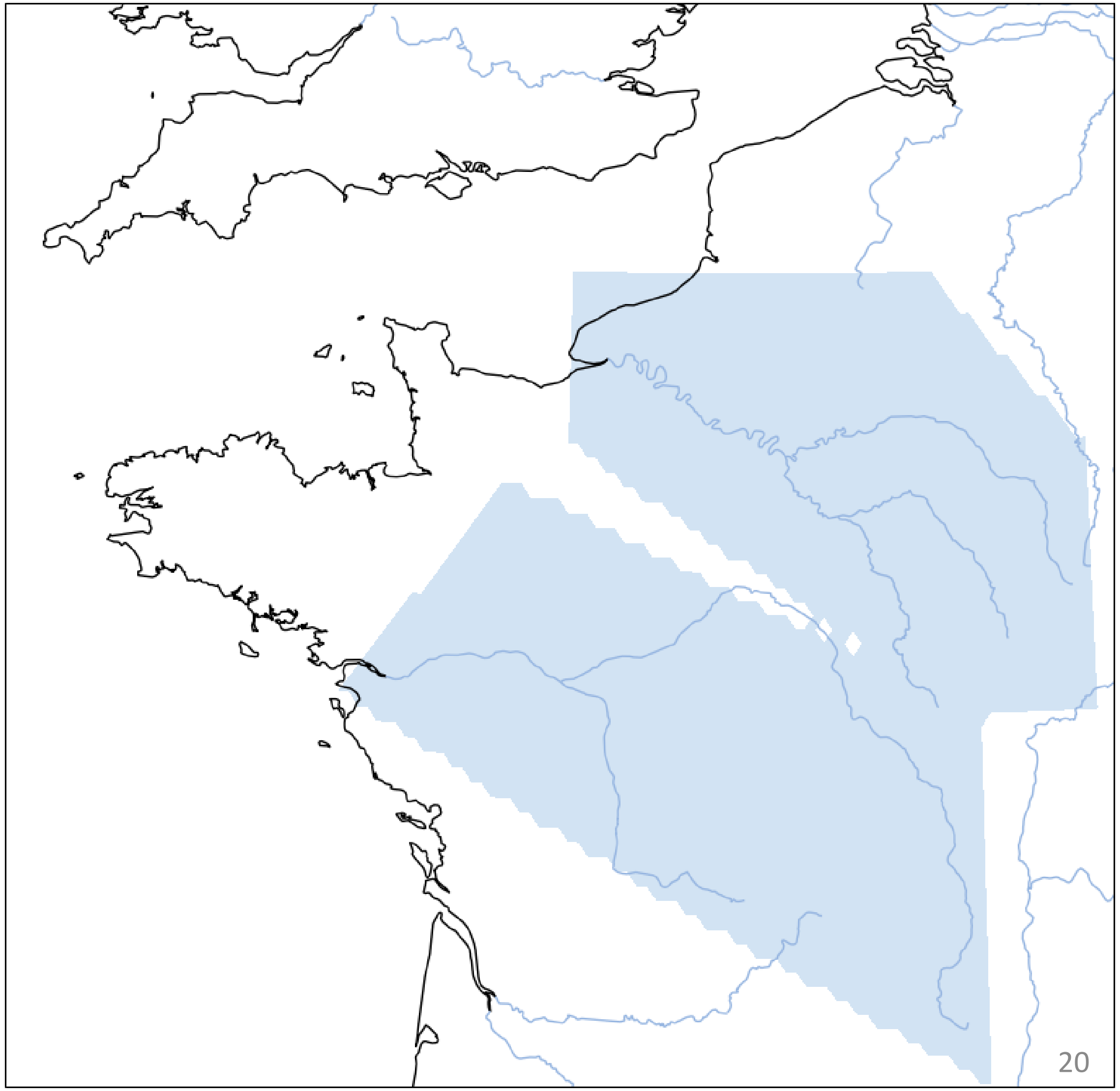


May/June French flooding

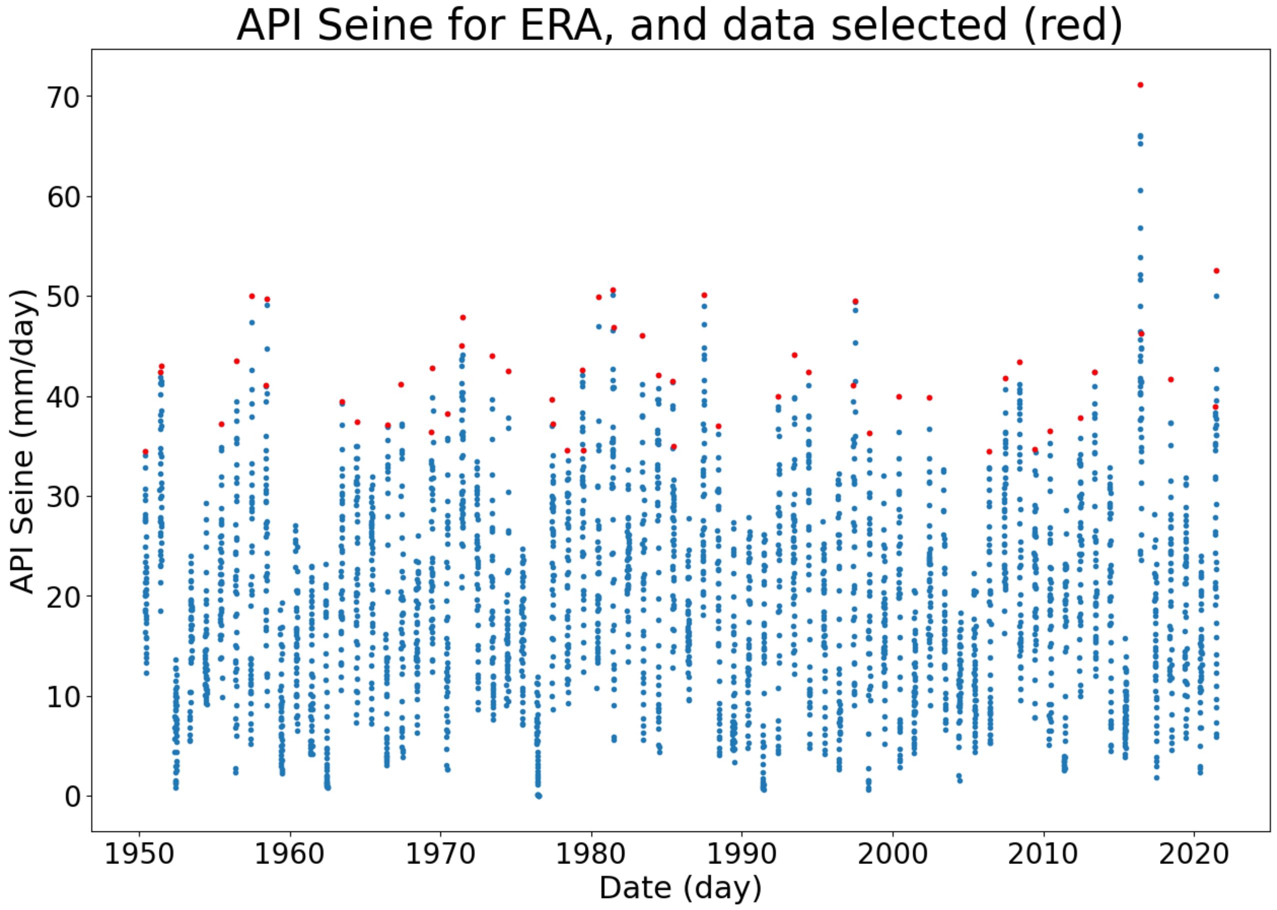
- Spatial daily precipitation means over the Seine and the Loire watersheds

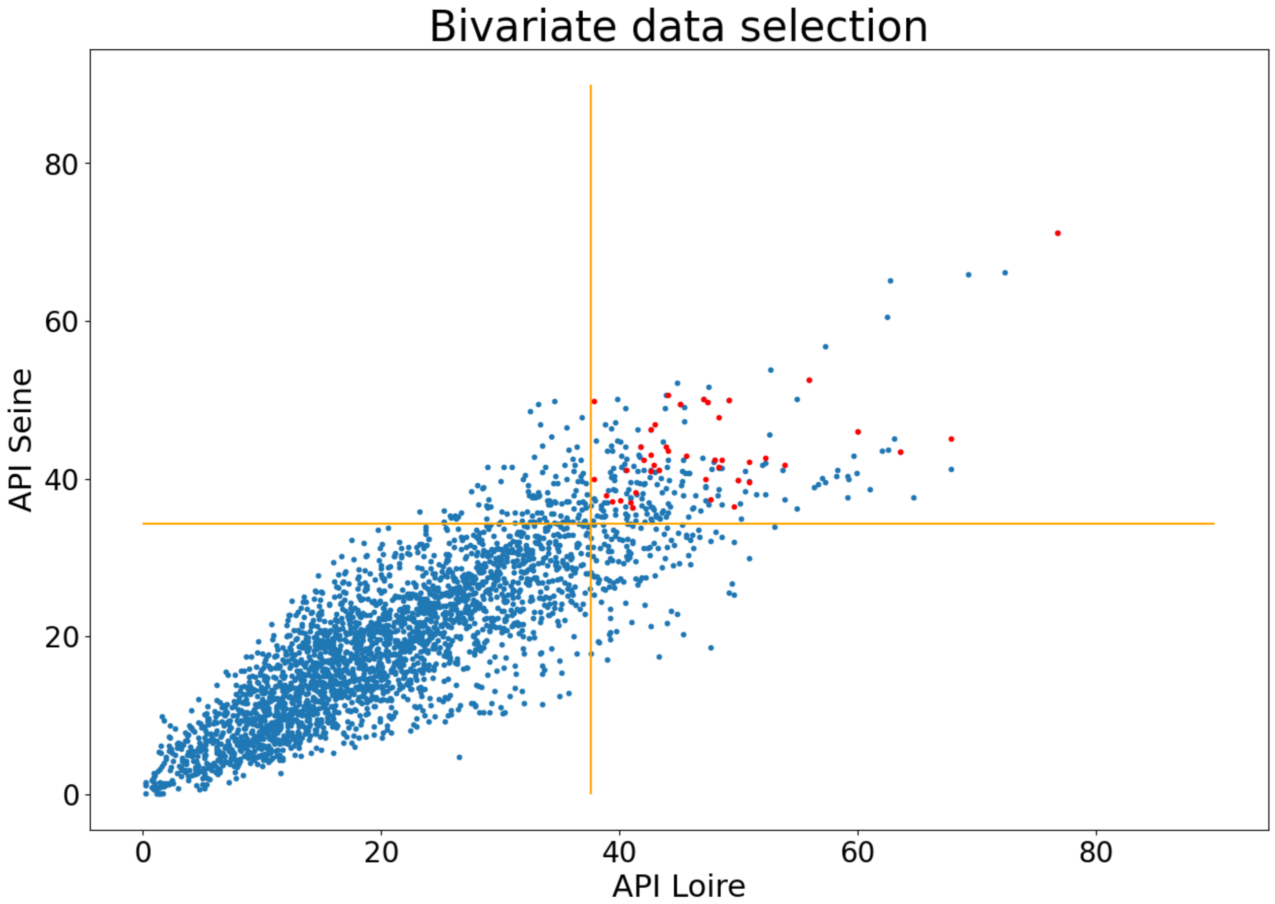
- API : $API_j = \sum_{i=1}^{i=N} k^{i-1} * TP_{j-i}$ with k=0.9 and N=20

- Same methodology (Data selection, GPD model, copula …)

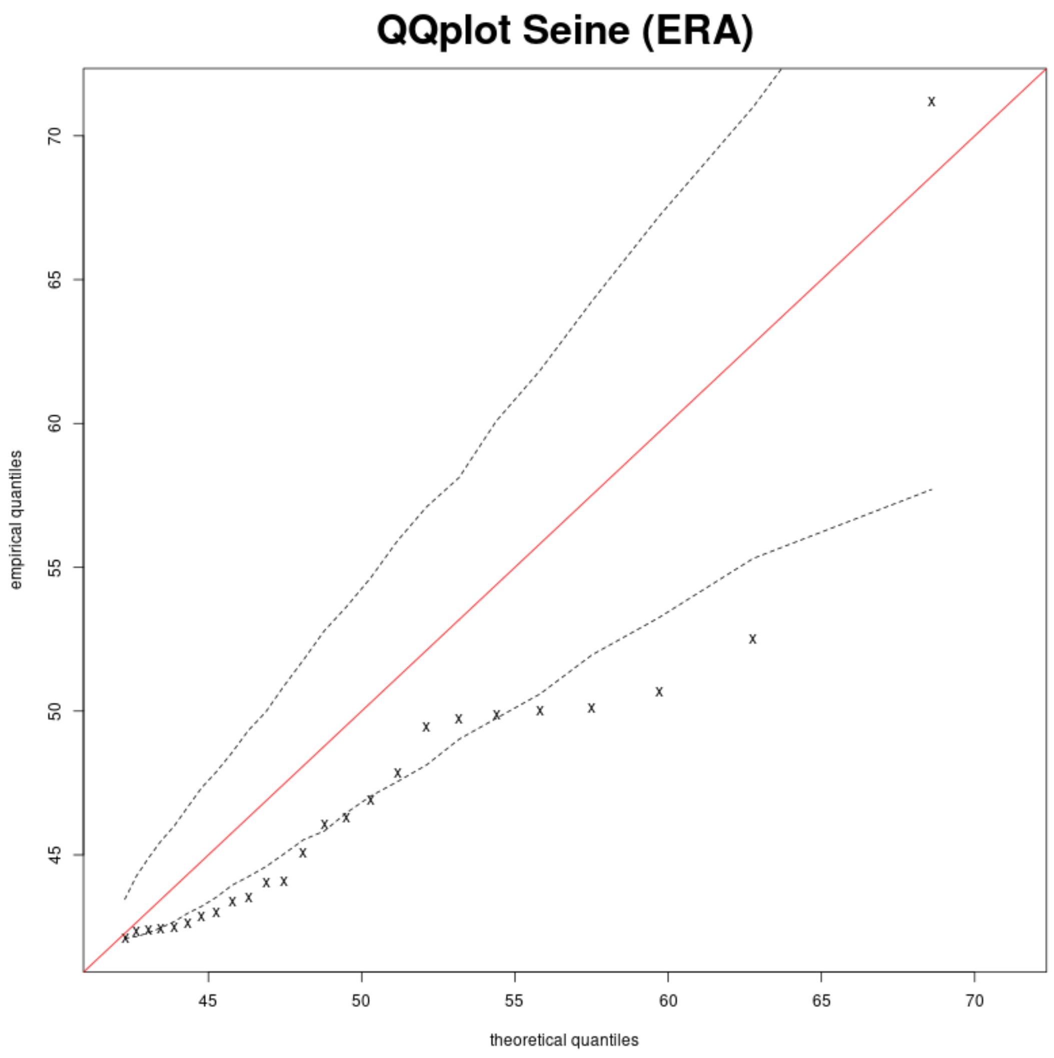


Data selection

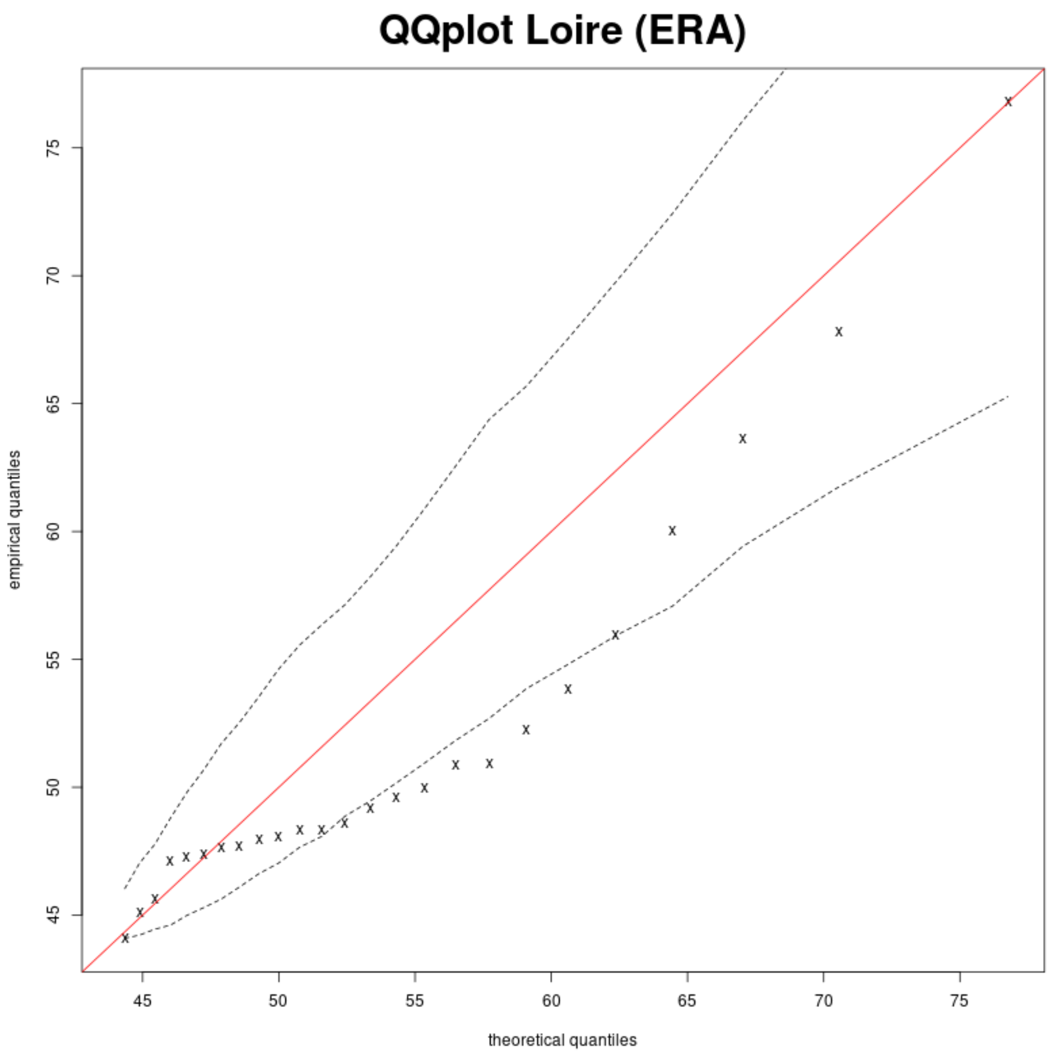




Quantile plots of GPD adjustment

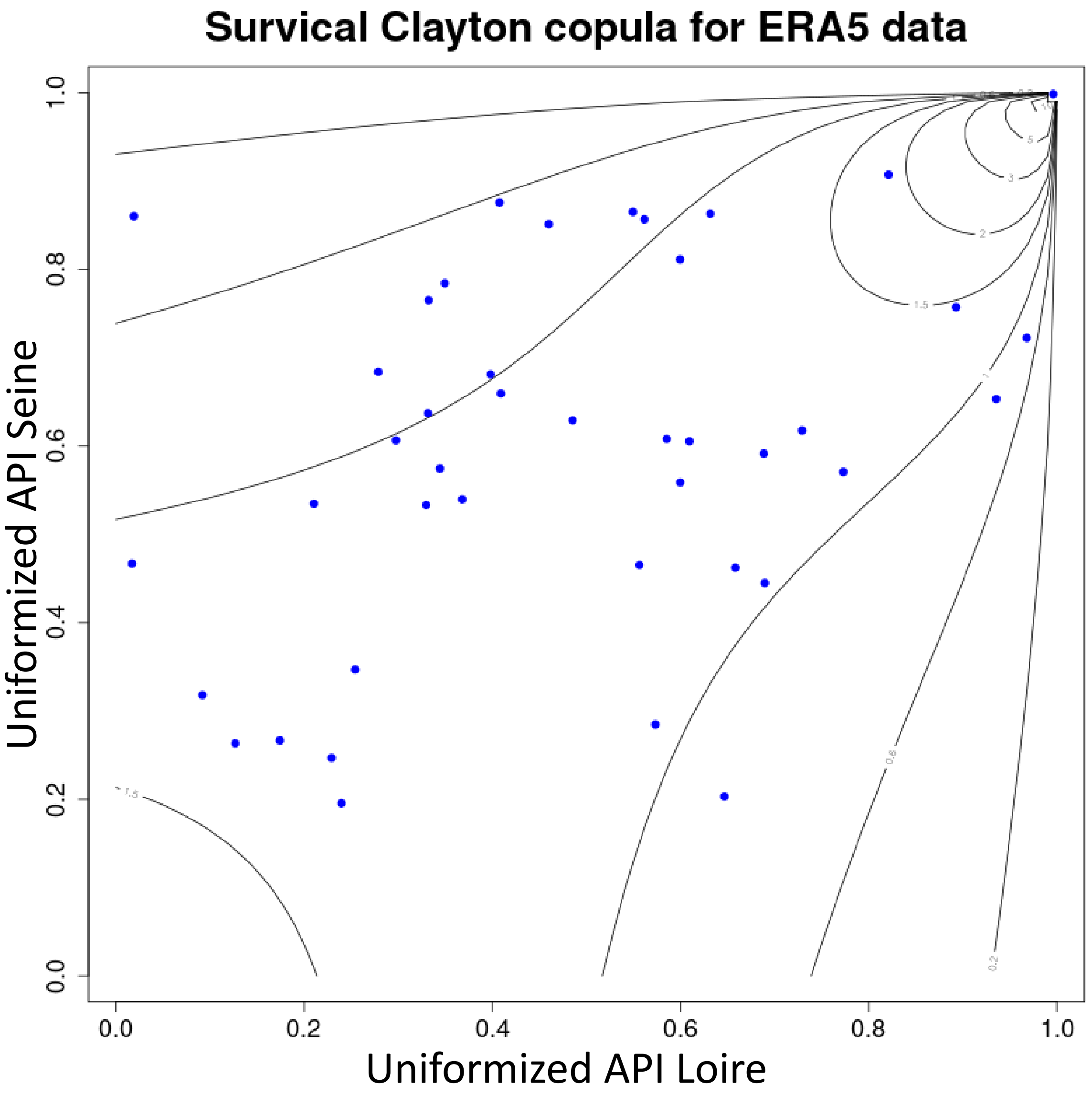


$\hat{\xi} = -0{,}164\ (0{,}116)$



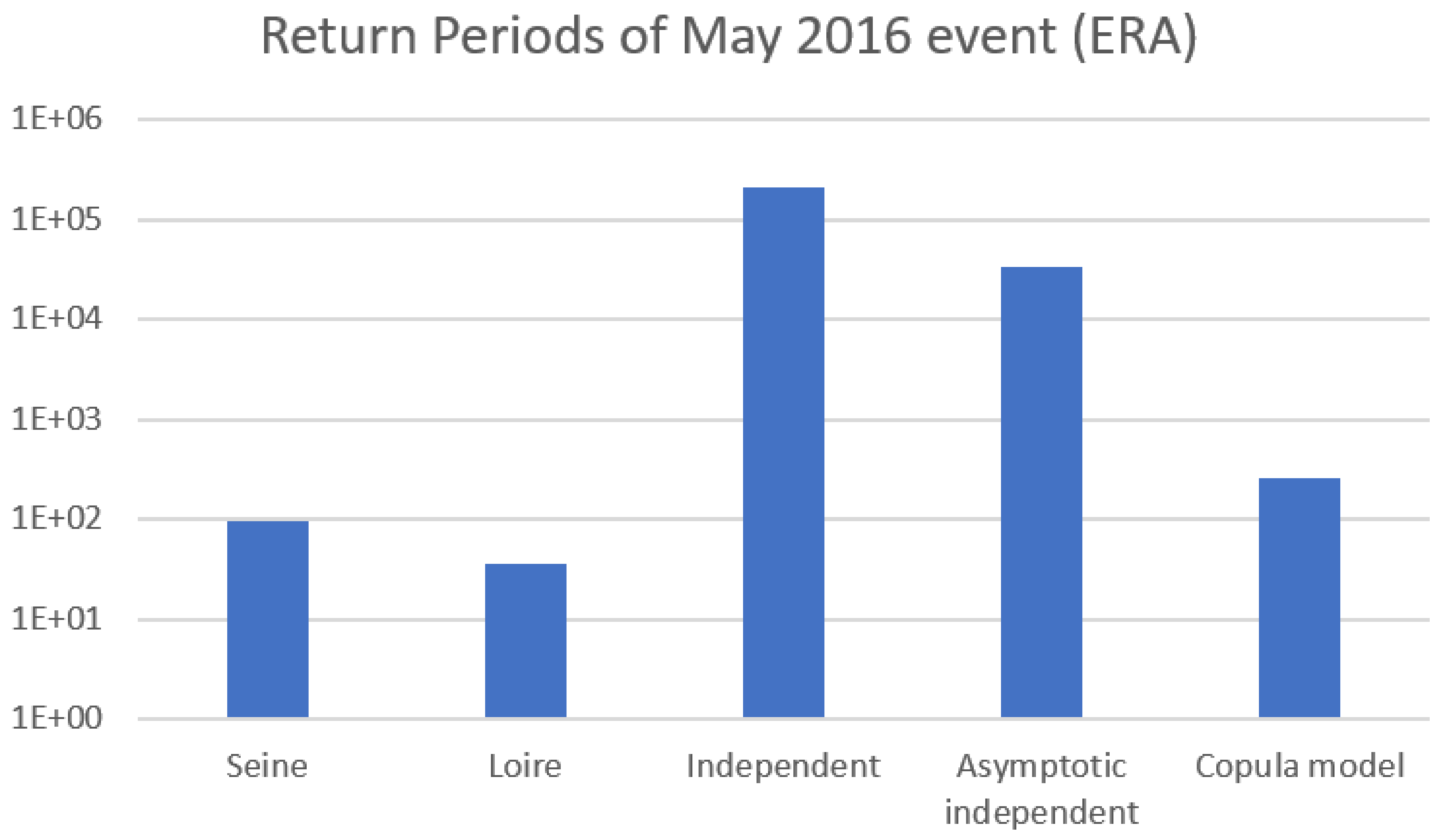
$\hat{\xi} = -0{,}252\ (0{,}109)$

Copula model

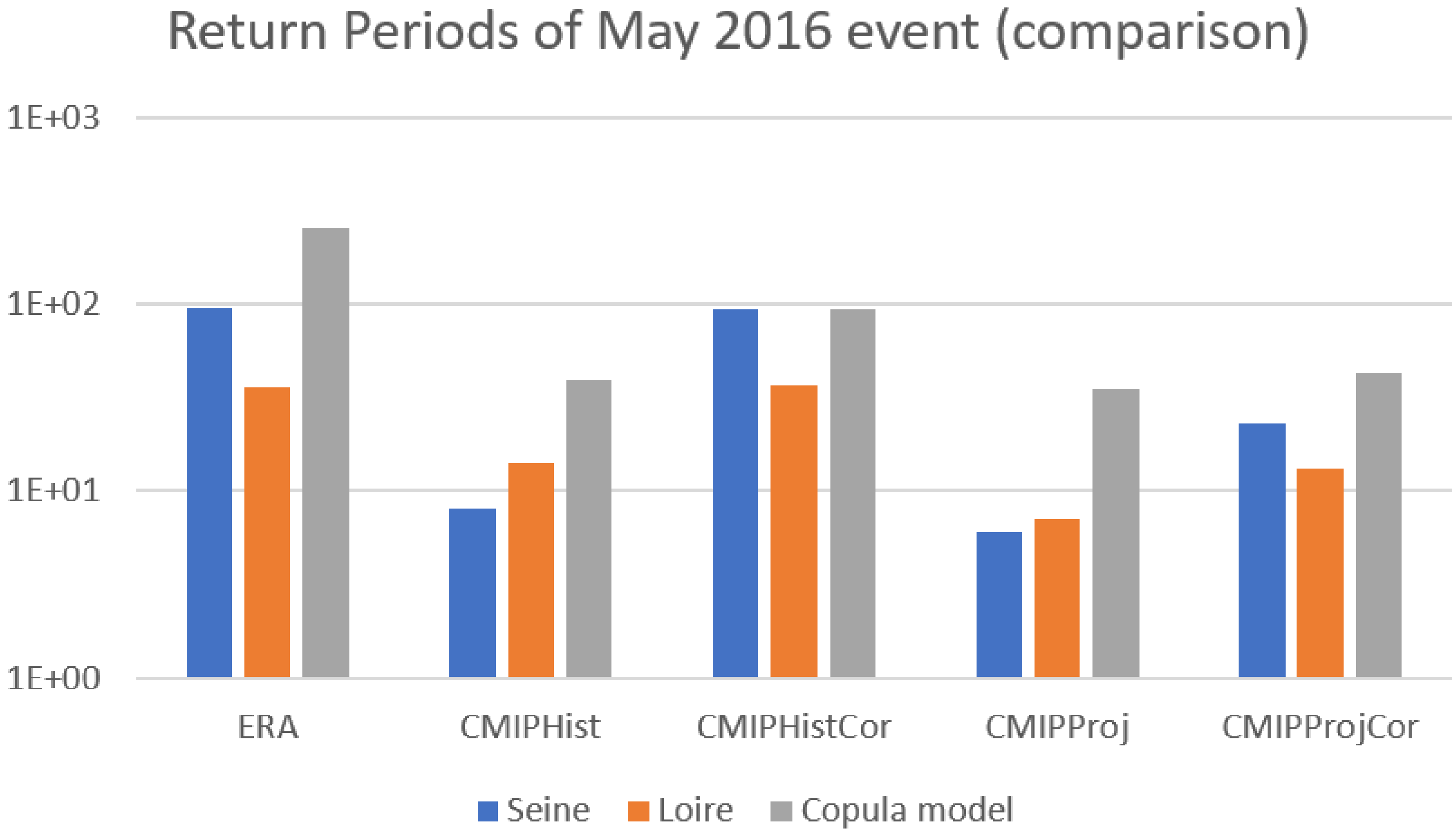


Data sets	Copula	χ	$\bar{\chi}$
ERA	Survival Clayton	0,43	1
Historic	Survival Clayton	0,295	1
Projection	Joe	0,425	1
Projection corrected	Student	0,441	1

Return periods results



Return periods results



Next steps

Next months :

- Multivariate bias correction
- Scale up framework to include more CMIP-6 simulations
- Paper

Next years :

- Apply treatment to convective cells