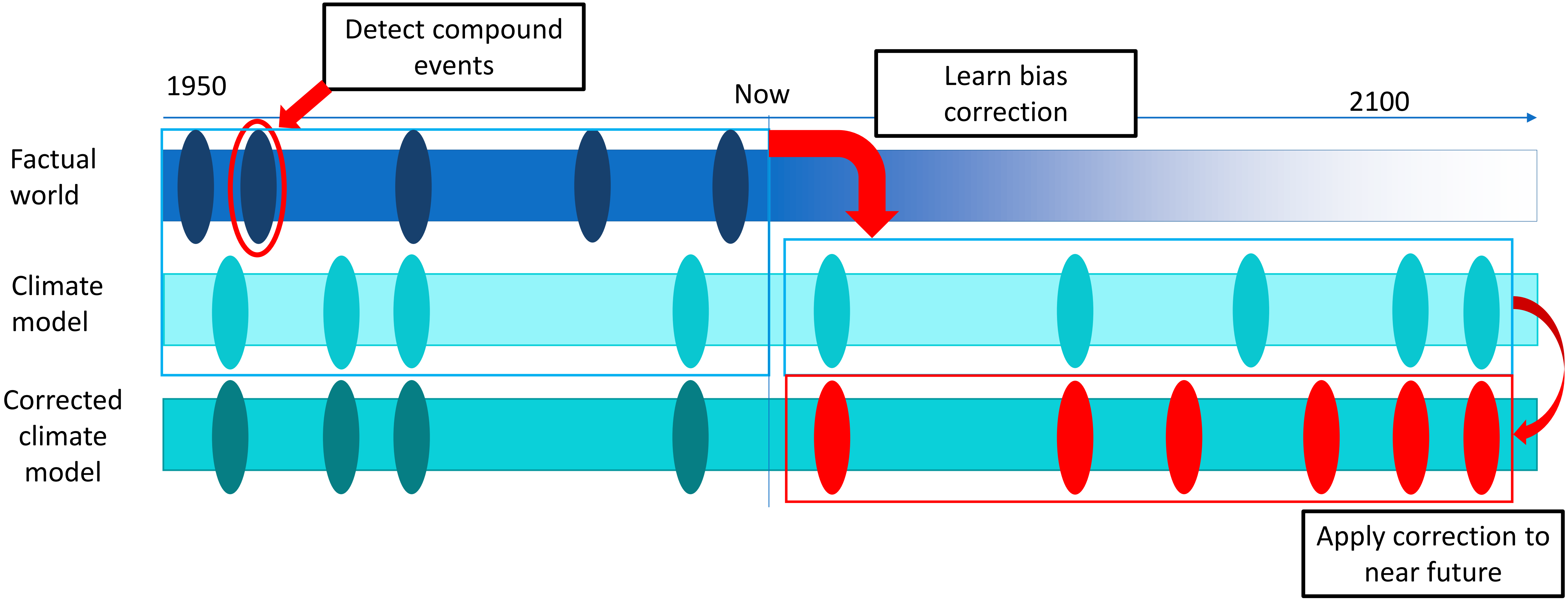
Quantification of devastating climate events under climate change through novel multivariate bias correction methods

Grégoire Jacquemin

Centre de géosciences, Mines Paris-PSL University & LSCE

Supervisors: Mathieu Vrac (LSCE, CNRS) , Denis Allard (BioSP, INRAE) & Xavier Freulon (CG, Mines Paris-PSL)

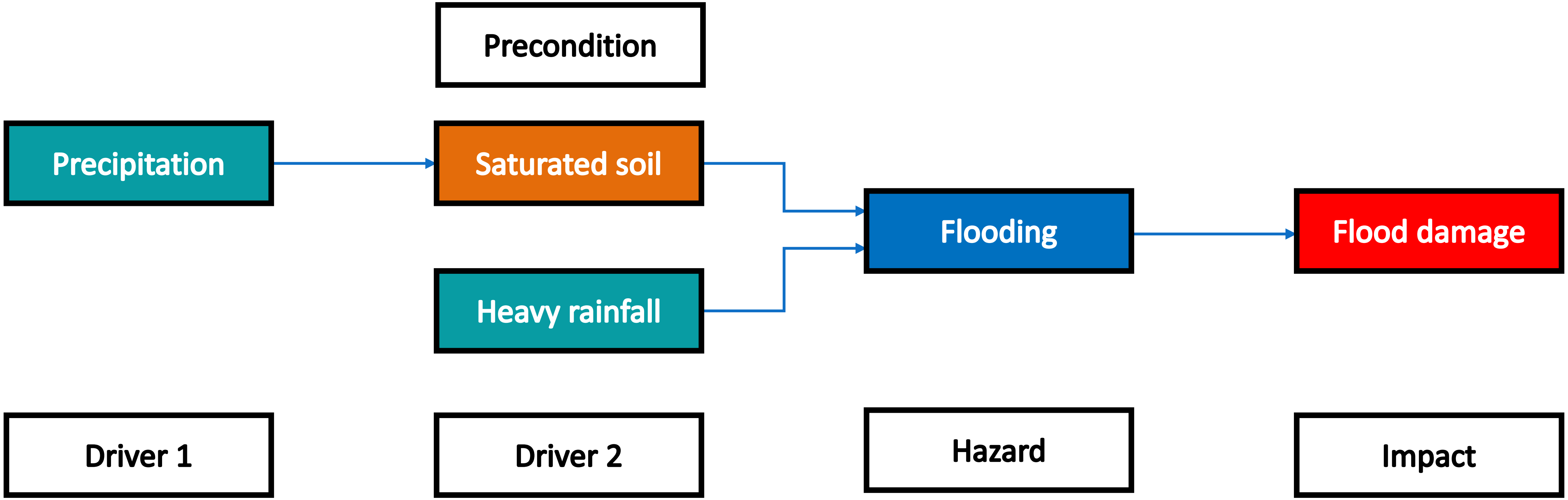
PHD objective



GEOLEARNING
CHAIRE /// Data Science for the Environment

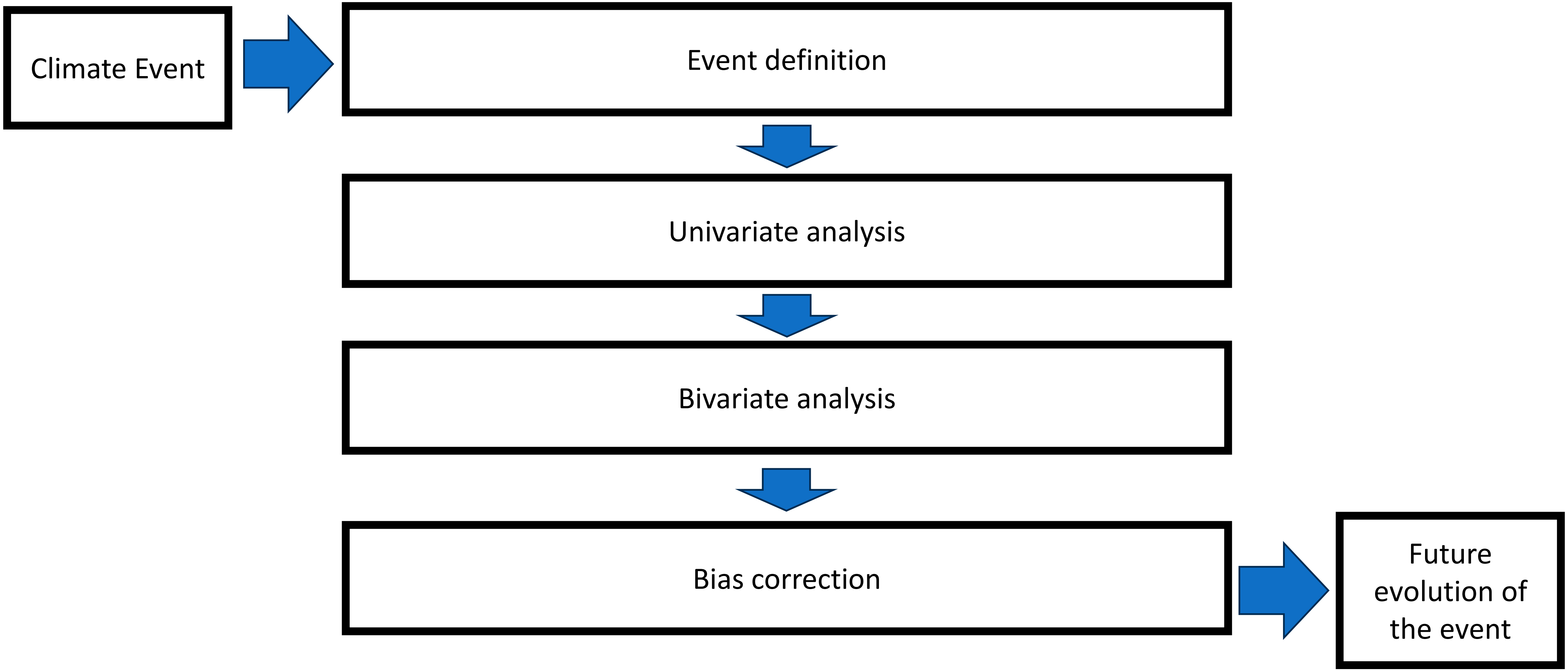
Compound events

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

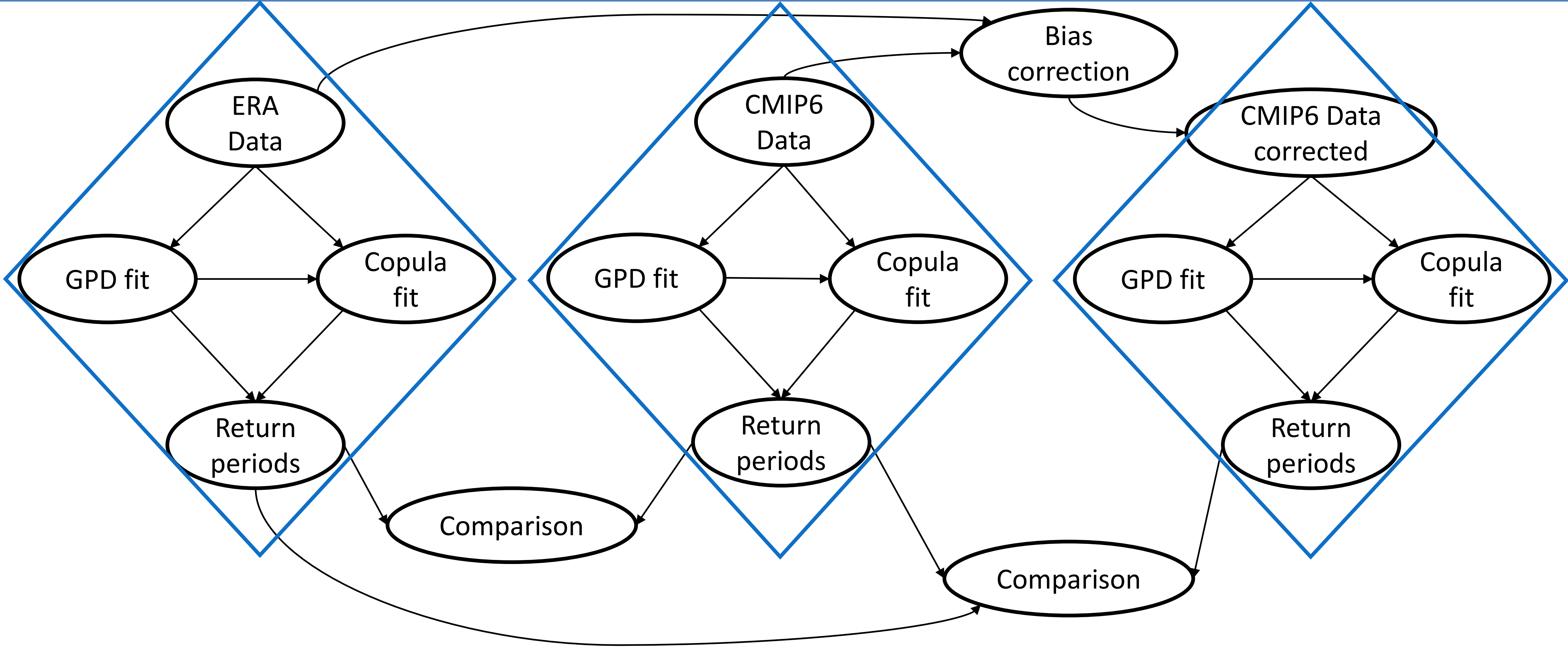


GEOLEARNING
CHAIRE /// Data Science for the Environment

Flowchart of statistical analysis



Method employed to assess events evolution



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 correct both the marginals and the dependence.

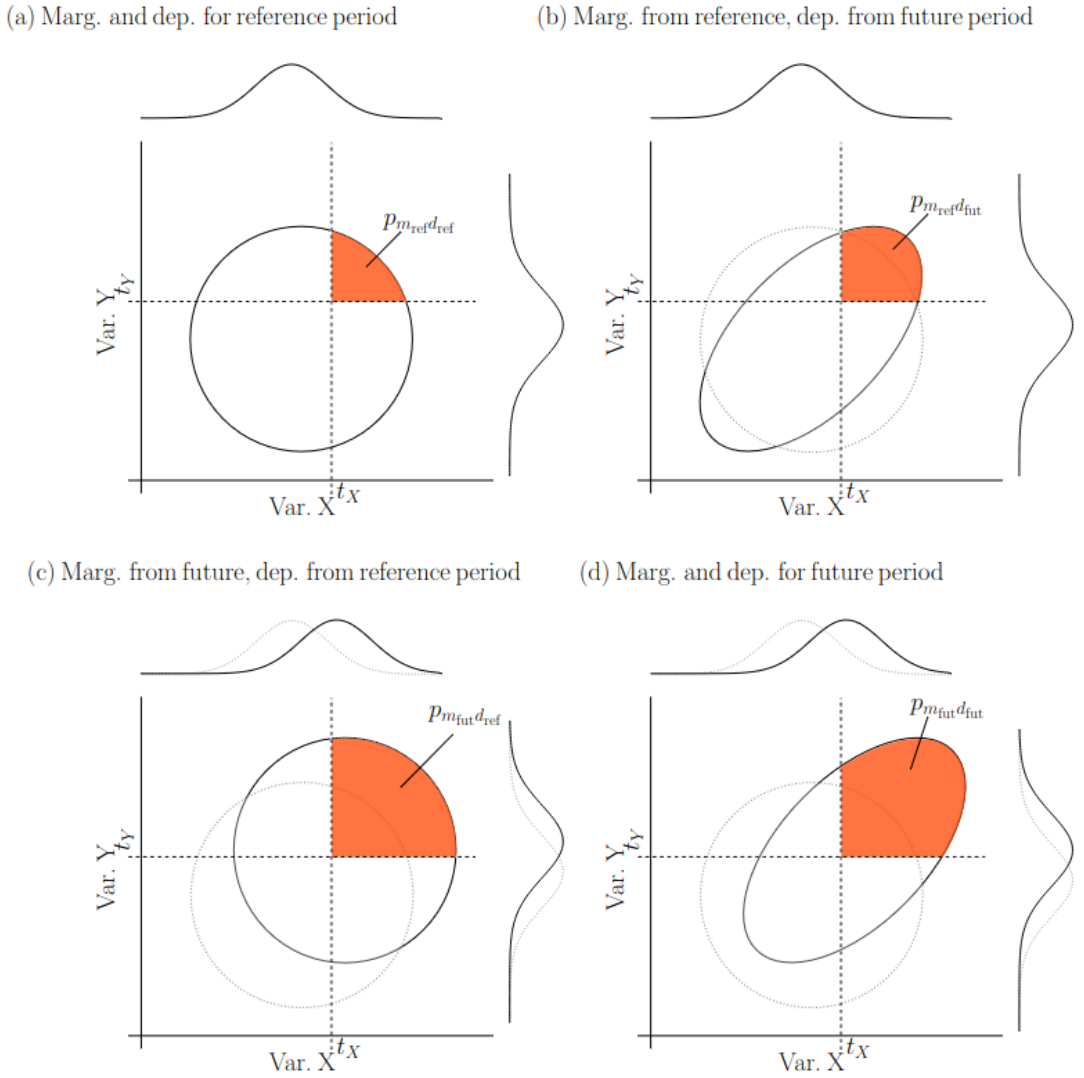


Figure from François, B., & Vrac, M. (2023). Time of emergence of compound events: contribution of univariate and dependence properties. Natural Hazards and Earth System Sciences, 23(1), 21-44.

Extreme value theory

- Following *Beirlant et al., (2006)* and *Bousquet et al., (2021)*, *we* introduce some EVT elements
- Let X and Y be two independent stationary processes with n realizations, and F their joint cumulative distribution function (cdf)
- We note F_X and F_Y the marginal cdf of X and Y respectively
- $\mathtt{M_n = max((X_{\mathit{i}}, Y_{\mathit{i}}))}$ for $1 \leq i \leq n$, component-wise

F is in the domain of attraction of a **multivariate extreme value distribution** G, written $F \in D(G)$, if there exist sequences $(\mathbf{a}_n)_n > 0$ and $(\mathbf{b}_n)_n$ in $\mathbf{R}^2$, and a nondegenerate distribution G, such that:

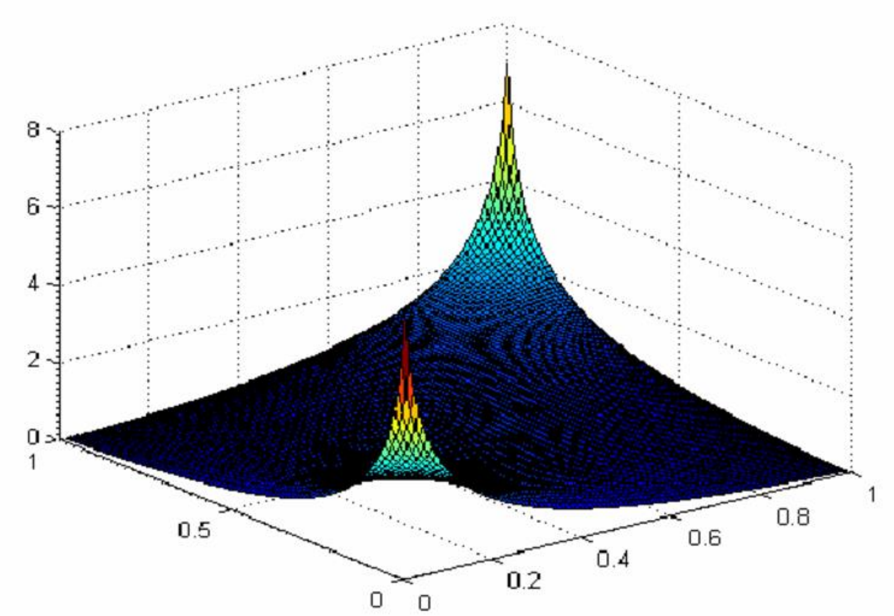
$$\mathbb{P}(\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n) \leq \mathbf{x}) \to G(\mathbf{x}), \quad n \to +\infty$$

Copulas

(Sklar (1959)) Let F be the multivariate cumulative distribution function of a random vector of dimension 2: $\mathbf{X} = (X, Y)$. Then there exists a function $C : \mathbf{R}^2 \to [0,1]$ called a **copula** defined for all $(x, y) \in \mathbf{R}^2$:

$$F(x, y) = C(F_X(x), F_Y(y)).$$

If F_X and F_Y are continuous, the copula C is unique.



Gaussian copula density. Picture from Abid, Fathi & Naifar, Nader. (2008). THE APPLICATION OF COPULAS IN PRICING DEPENDENT CREDIT DERIVATIVES INSTRUMENTS. Journal of Applied Economic Sciences. 3.

Estimate the multivariate distribution

- **<u>Theorem</u>**: *(Deheuvels (1984), Galambos (1987))* F is in the domain of attraction of G if and only if:
  - all the margins of F are in the domain of attraction of the margins of G respectively
  - and the copula of F is in the domain of attraction of the copula of G

- 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

Two events

- July 2021 Belgian/German flooding
  (Preconditioned event)
  (Mohr et al., 2022)

- May/June 2016 French flooding
  (Spatially compound event)
  (van Oldenborgh et al., 2023)

July event

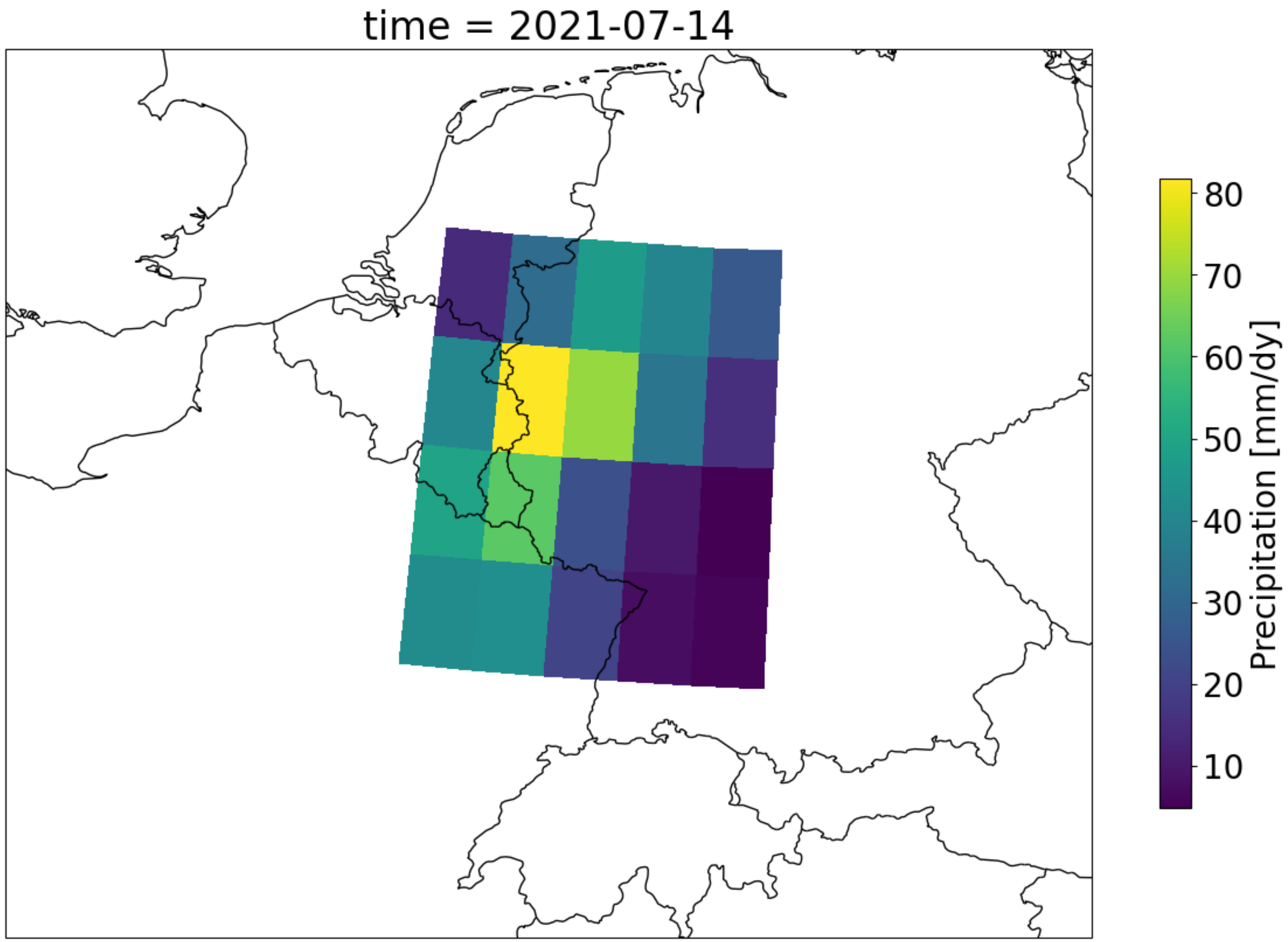
- Data from ERA5, 1°x1° grid over June, July, August

- 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 *(Linsley et al., 1975)*



Data Selection algorithm for July event

Data selection for 1D

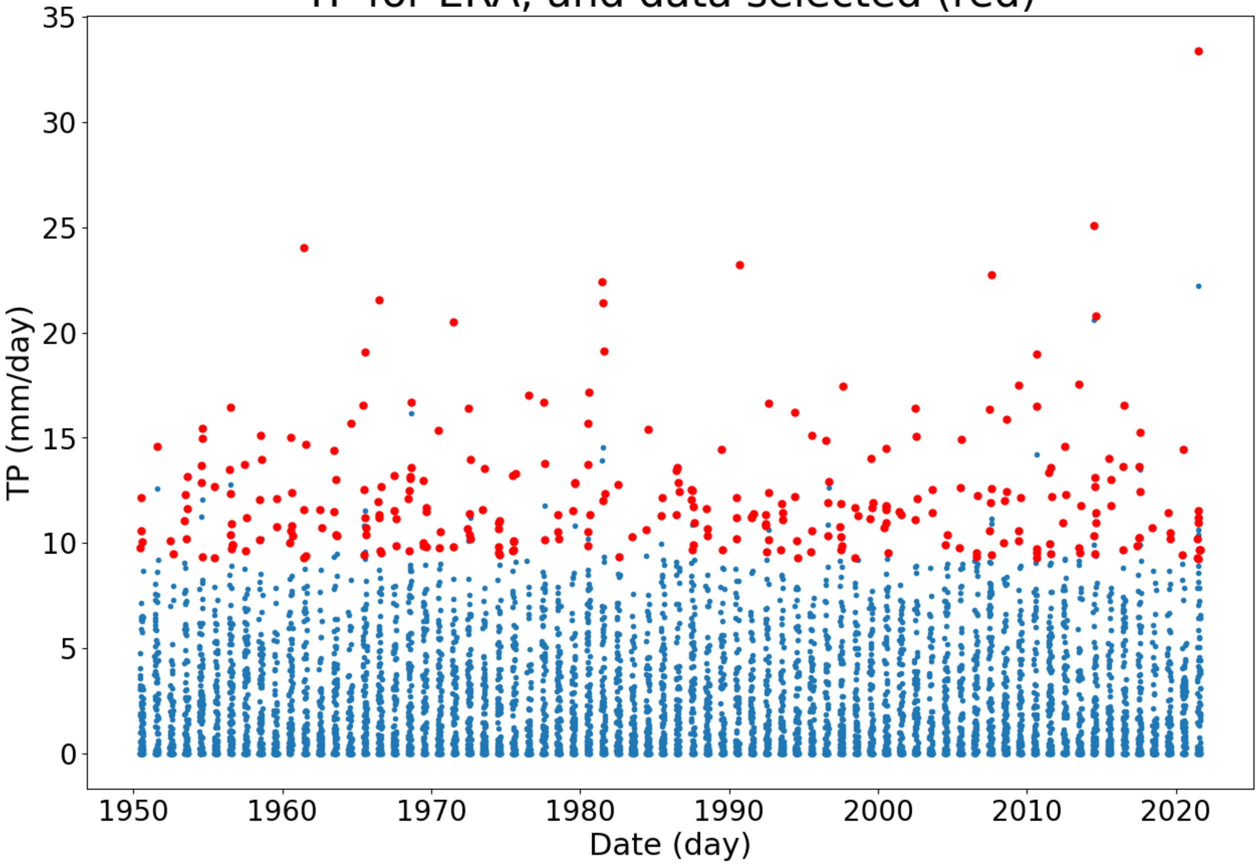
- For TP: select points above the 95th quantile, separated by at least 2 days
- For API: select points above the 95th quantile, and weakly correlated : $\rho\left(API_j, API_{j+\boldsymbol{h}}\right) < 0.1$
- We find $\boldsymbol{h} = $ **20 days**

Data selection for 2D

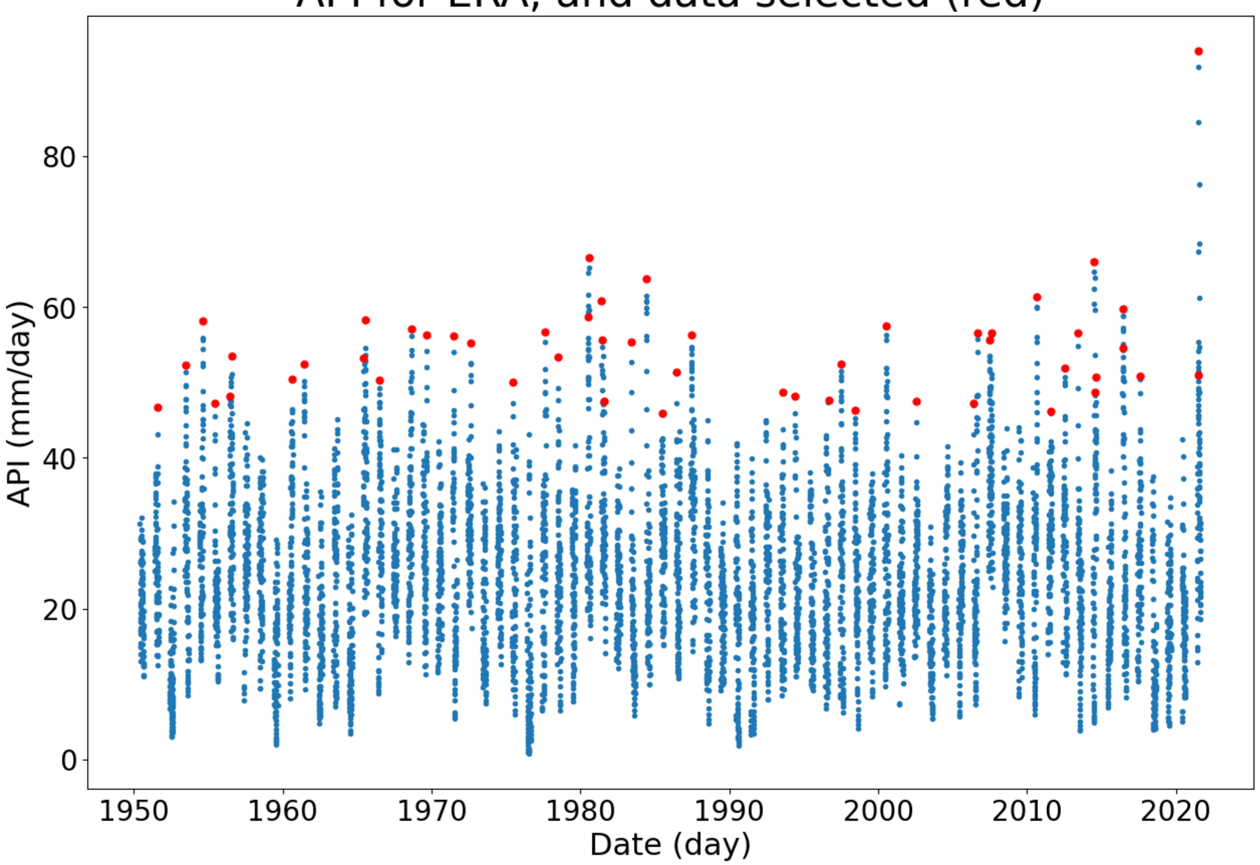
- Select (TP_i, API_j) with $TP_i > Q95_{TP}$, $API_j > Q95_{API}$ and $i - 5 \leq j \leq i$
- Then select couples separated by at least h days, according to the highest TP value

Univariate selection July event

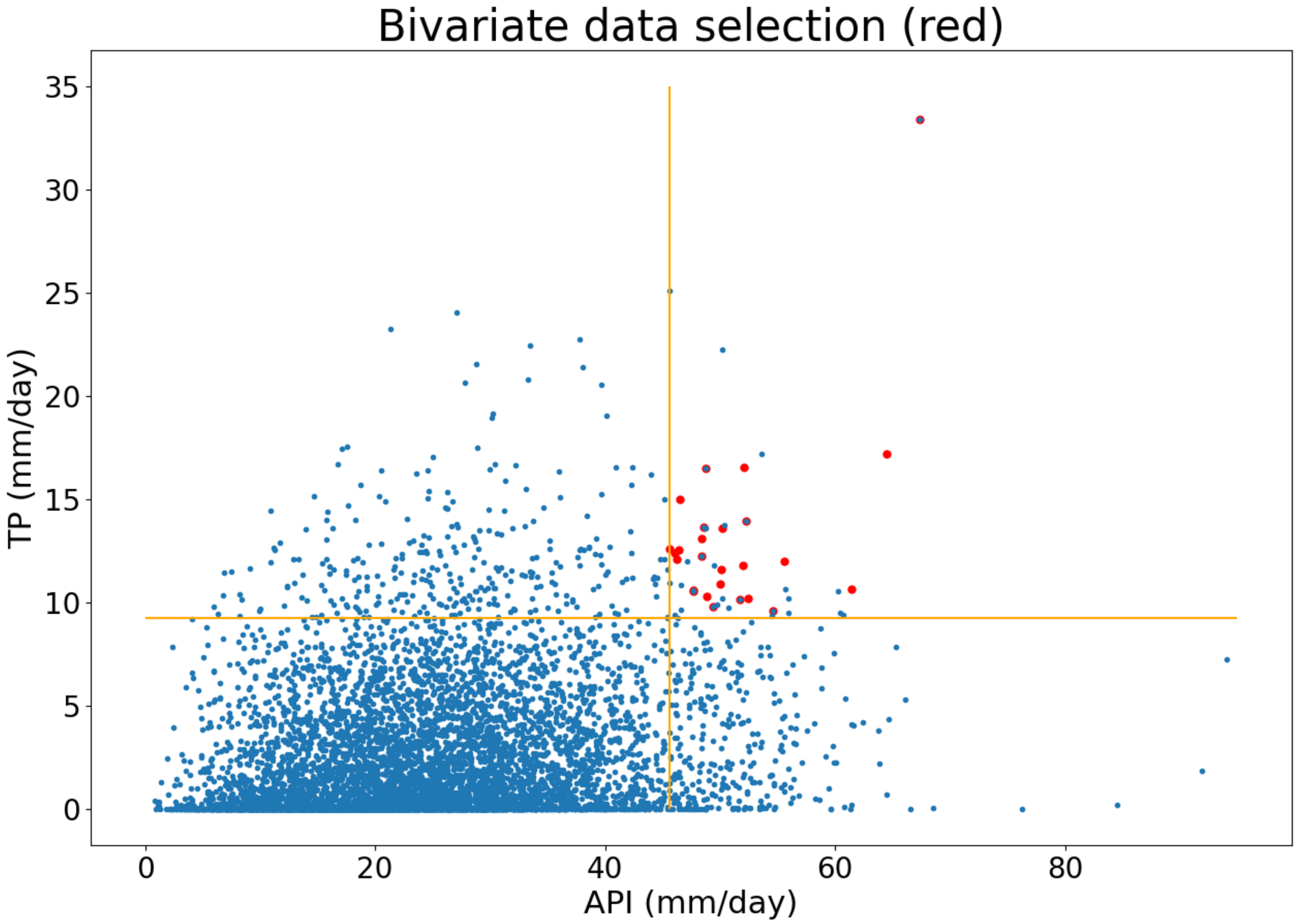
TP for ERA, and data selected (red)



API for ERA, and data selected (red)



Bivariate selection July event



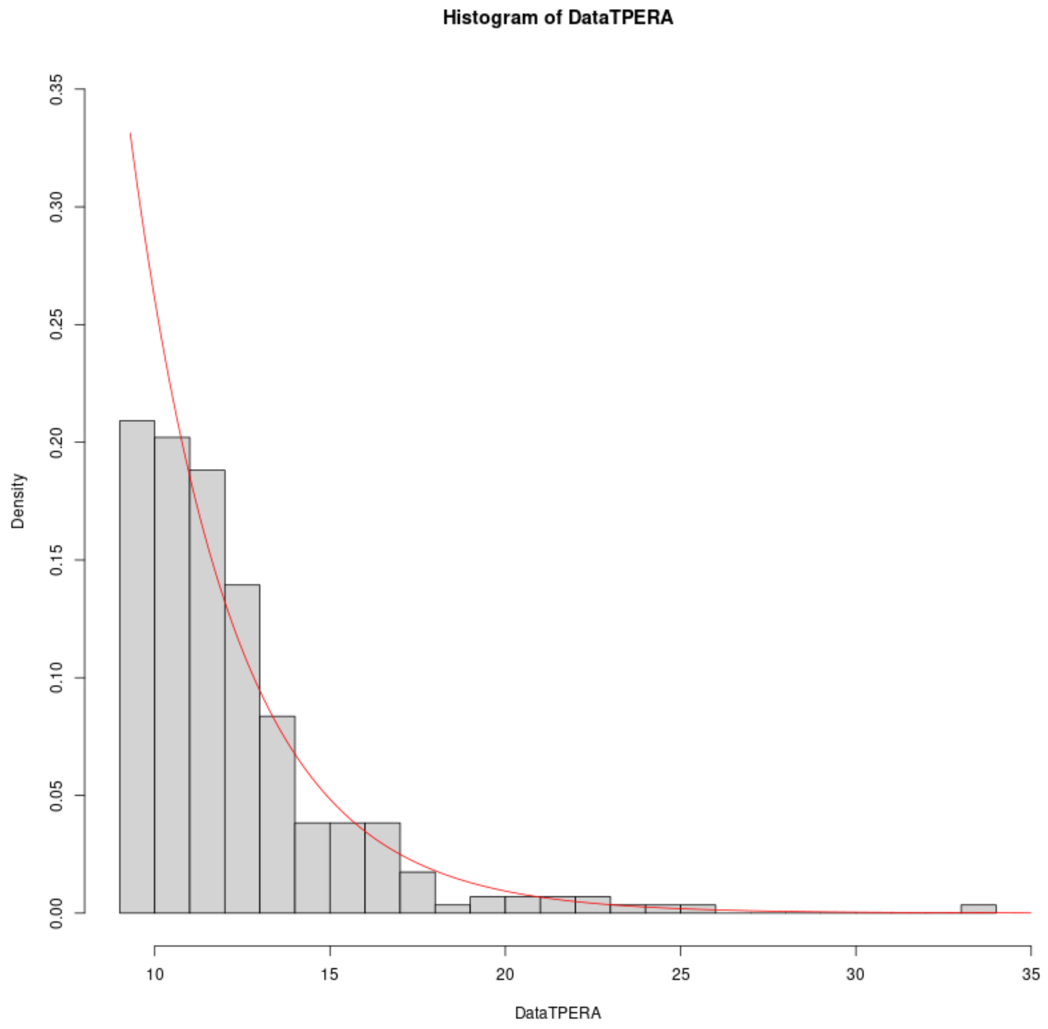
Generalized Pareto Distribution model

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

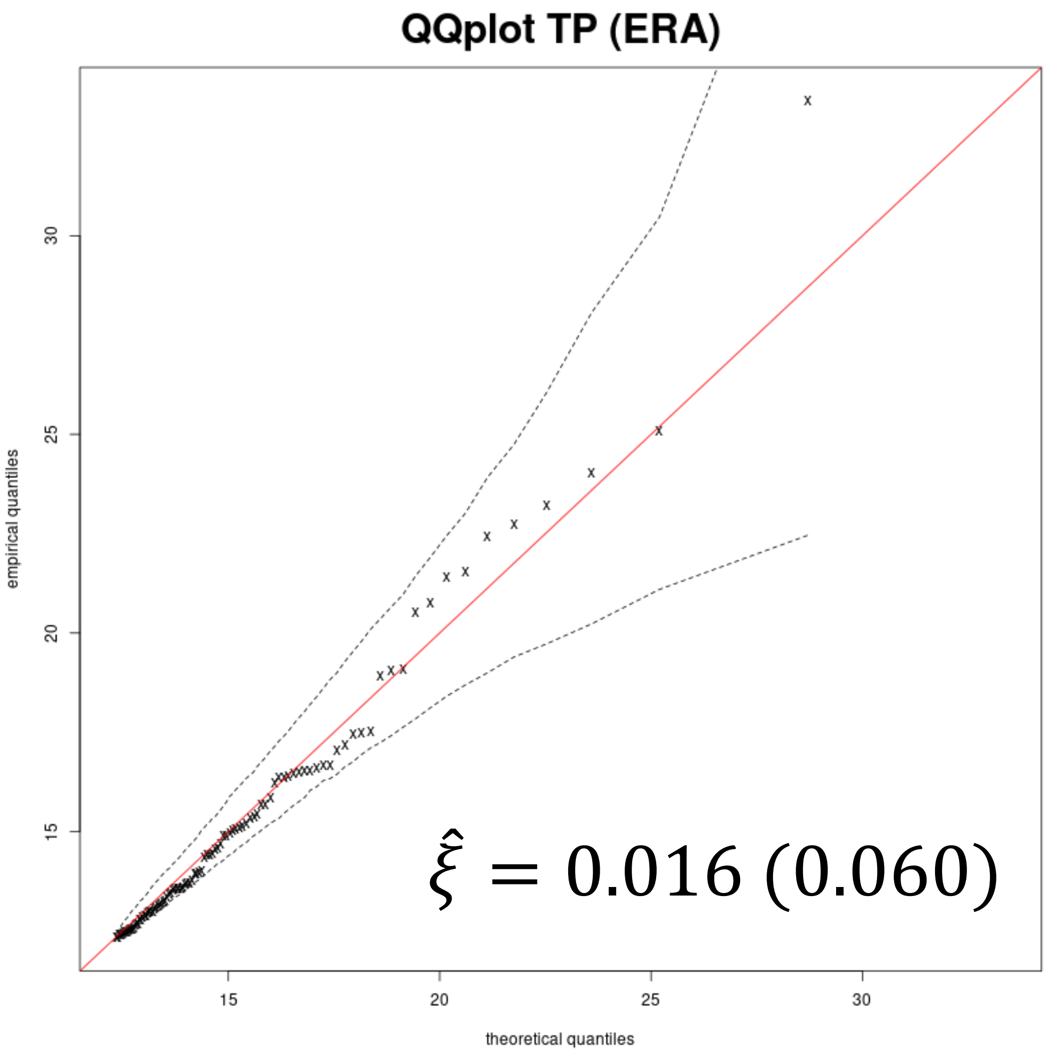
$$F(x) = 1 - (1 + \xi x)^{\frac{-1}{\xi}}$$

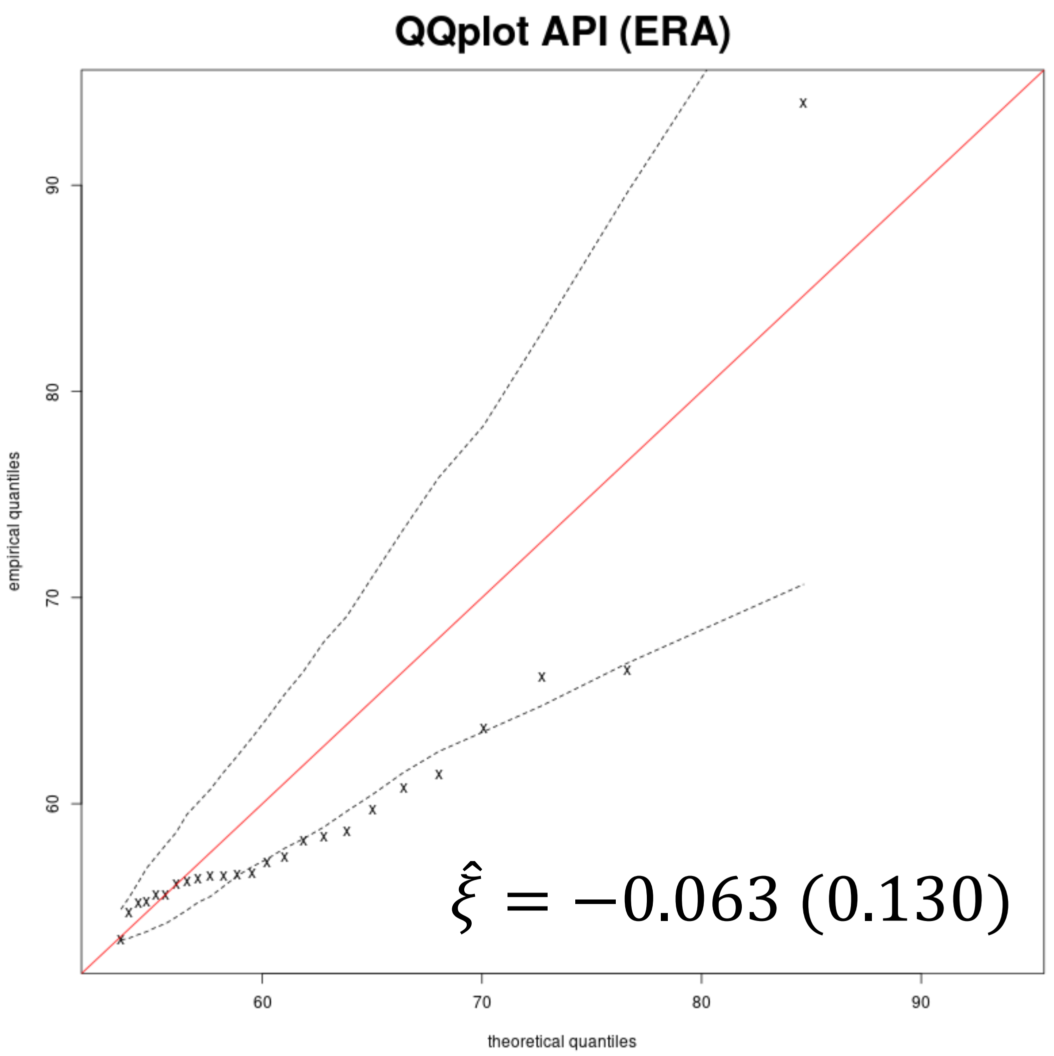
with $x \geq 0$ and $\xi \neq 0$

Parameters are estimated through maximum likelihood method



Quantile plots of GPD adjustment

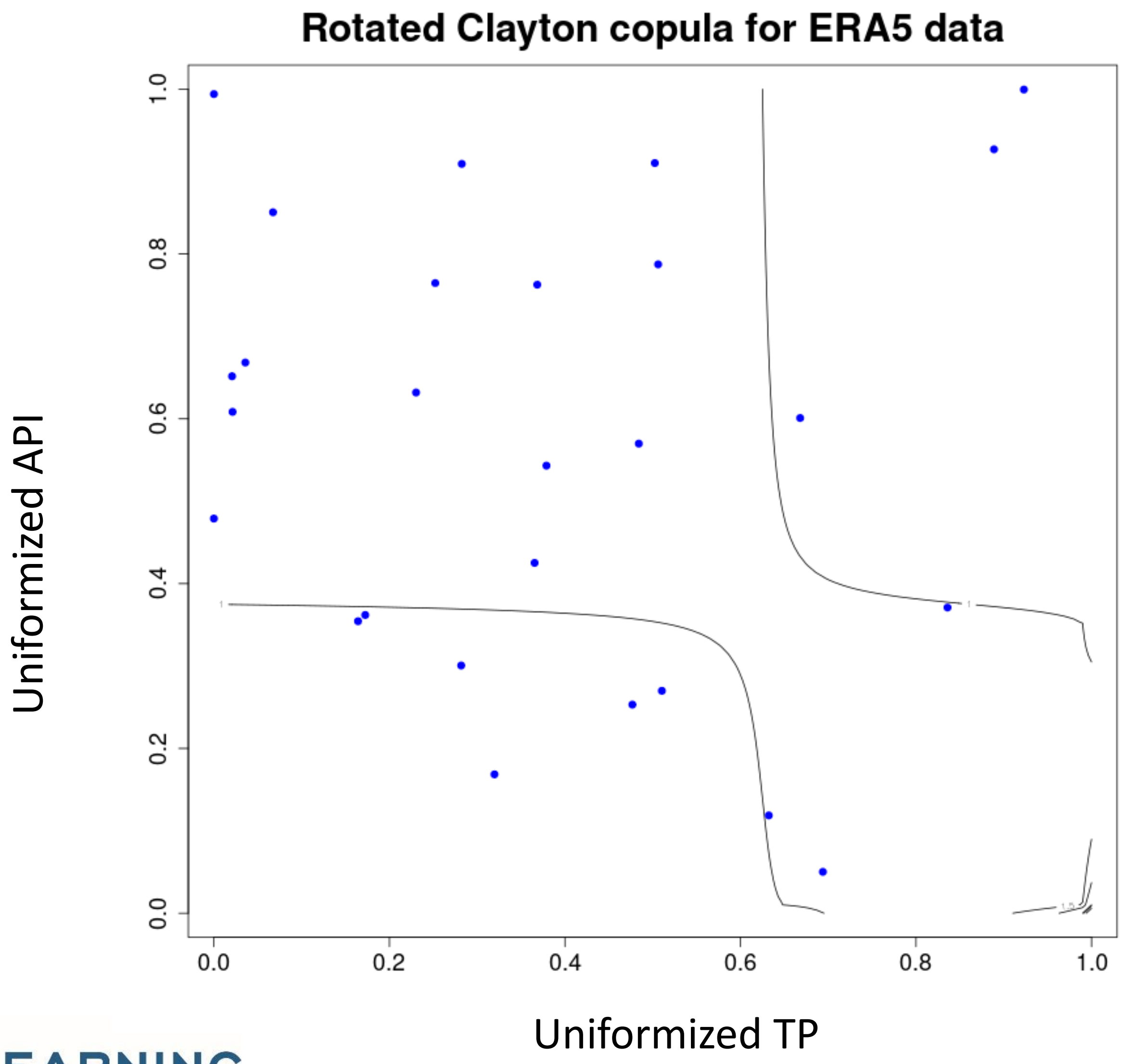




Copula model estimation

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 Bayesian Information Criteria (BIC)



Return periods

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

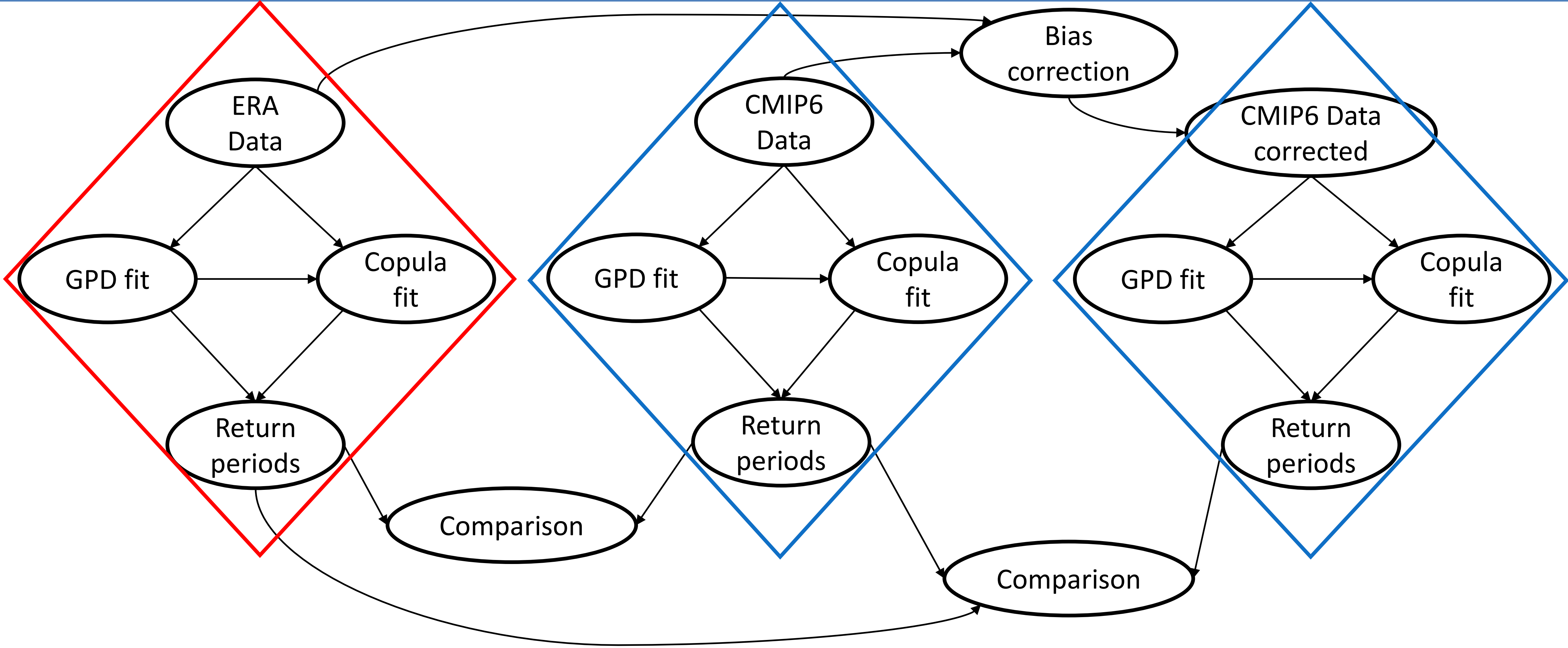
$$\boxed{T(x_{14.07}) = \frac{1/n}{1 - \boldsymbol{P}(X \leq x_{14.07})}}$$

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

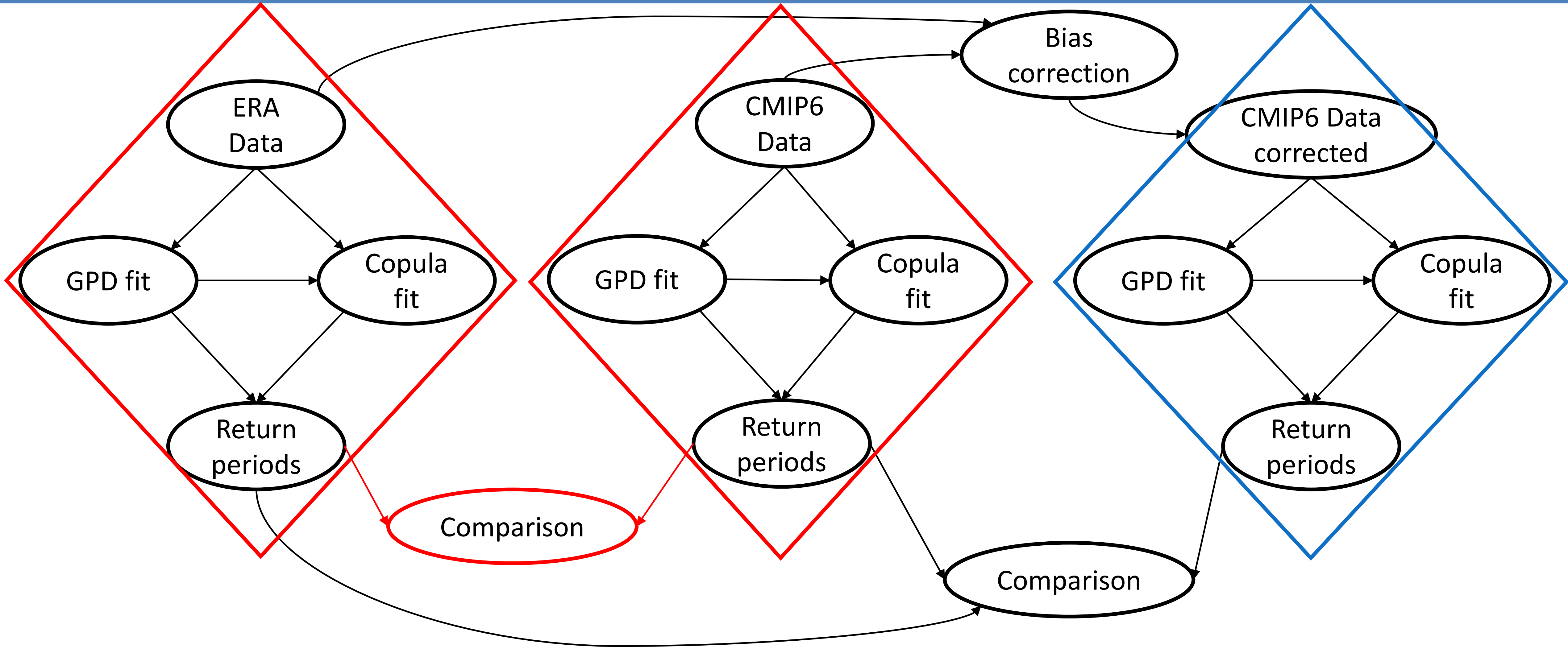
$$\boxed{T_B(TP_{14.07}, API_{14.07}) \approx \frac{1/n}{\frac{N_{u,v}}{N}[1 - U_{TP} - U_{API} + C(U_{TP}, U_{API})]}}$$

with $U_X = F(x_{14.07})$, C the copula, N the total number of points and $N_{u,v}$ the number of points above u and v

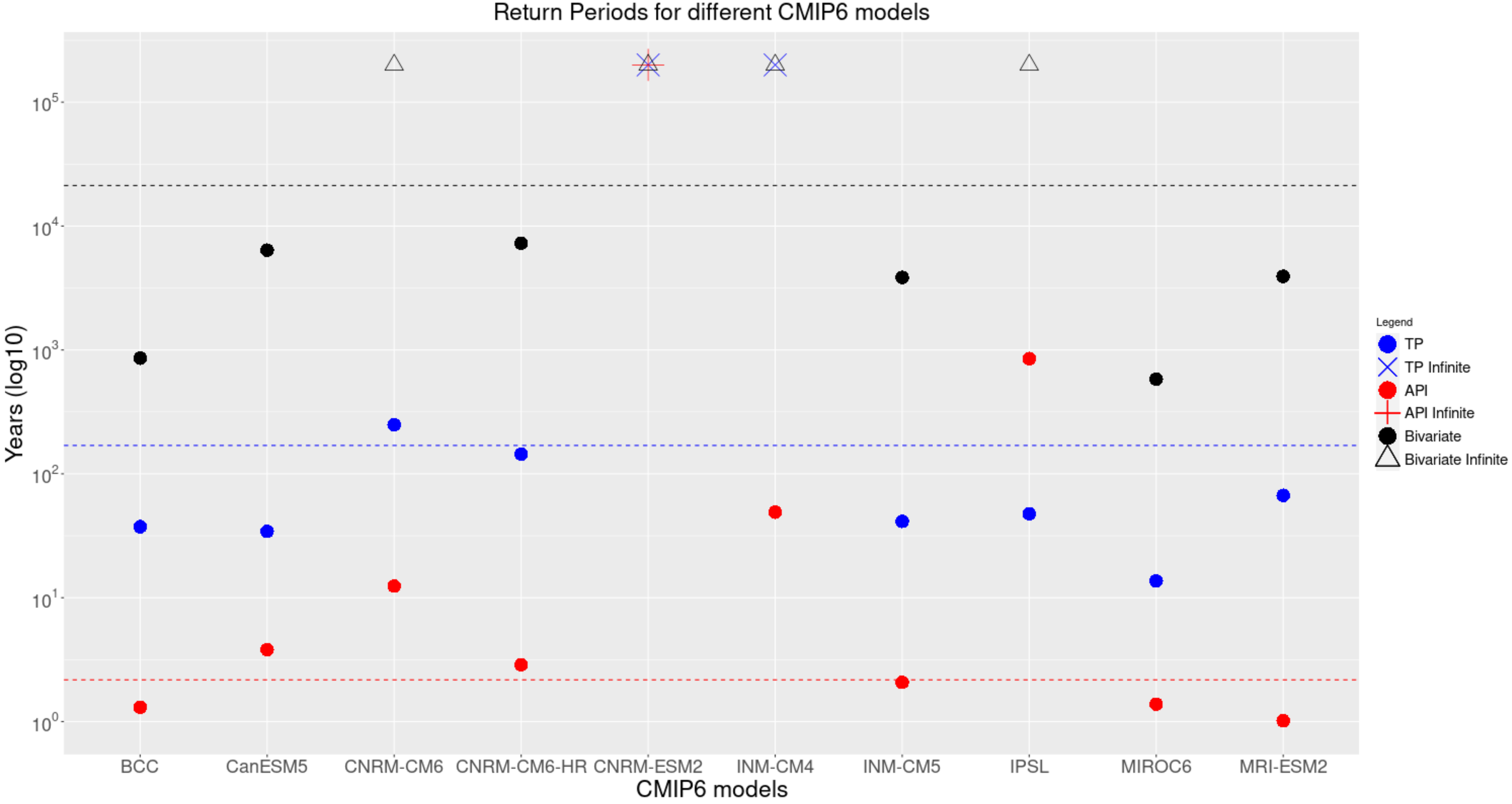
Method employed to assess events evolution



Method employed to assess events evolution



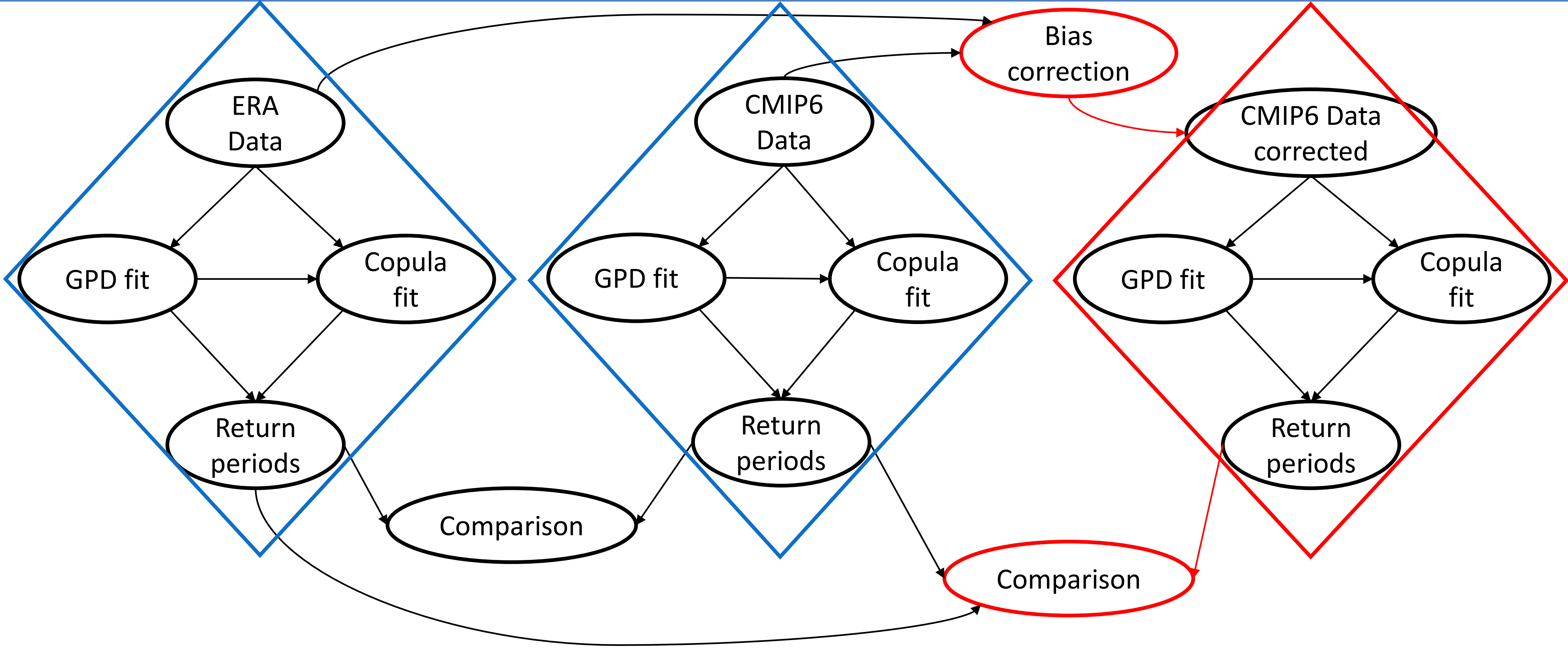
Results July event



- The dashed lines represent the ERA return periods
- These return periods are calculated on the 1992-2021 period
- The models' return periods seem off → need for bias correction

GEOLEARNING
CHAIRE /// Data Science for the Environment

Method employed to assess events evolution



Climate models

- All the considered runs follow the **ssp5-8.5 scenario** (worst scenario) *(IPCC report 6)*

→ We plan to consider more scenarios in the future

- A climatic period is usually considered to be 30 years. We have data between 1950 and 2100. We separated the data into 5 climatic periods:

1950-1979, 1992-2021, 2022-2051, 2041-2070, 2071-2100

with an historic period (1950-1979), a reference period (1992-2021) and 3 future periods

CDF-t

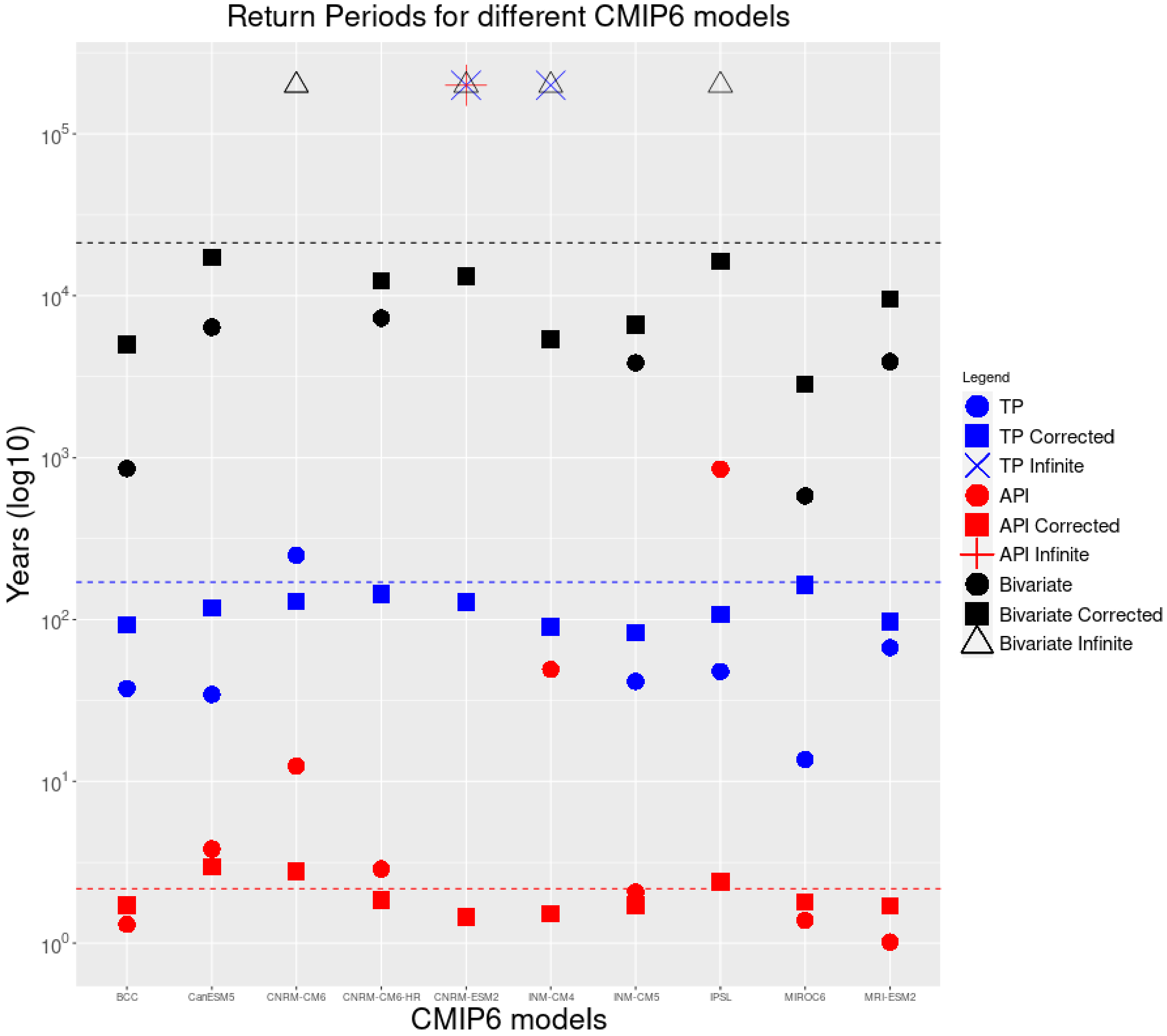
- We perform a univariate bias correction method: Cumulative Distribution Function-transform (CDF-t) *(Vrac et al., 2012)*

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

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

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

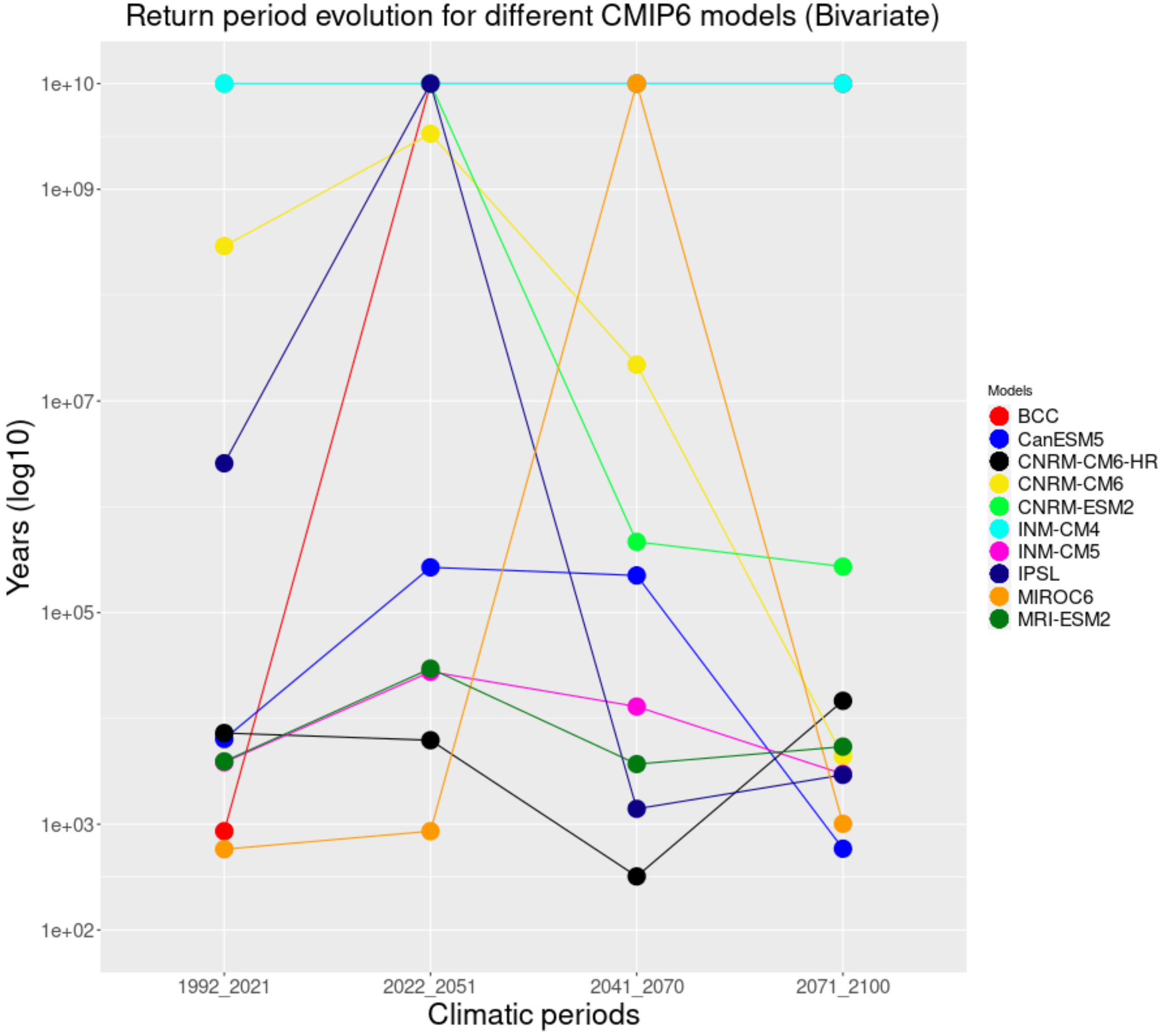
Results July event corrected

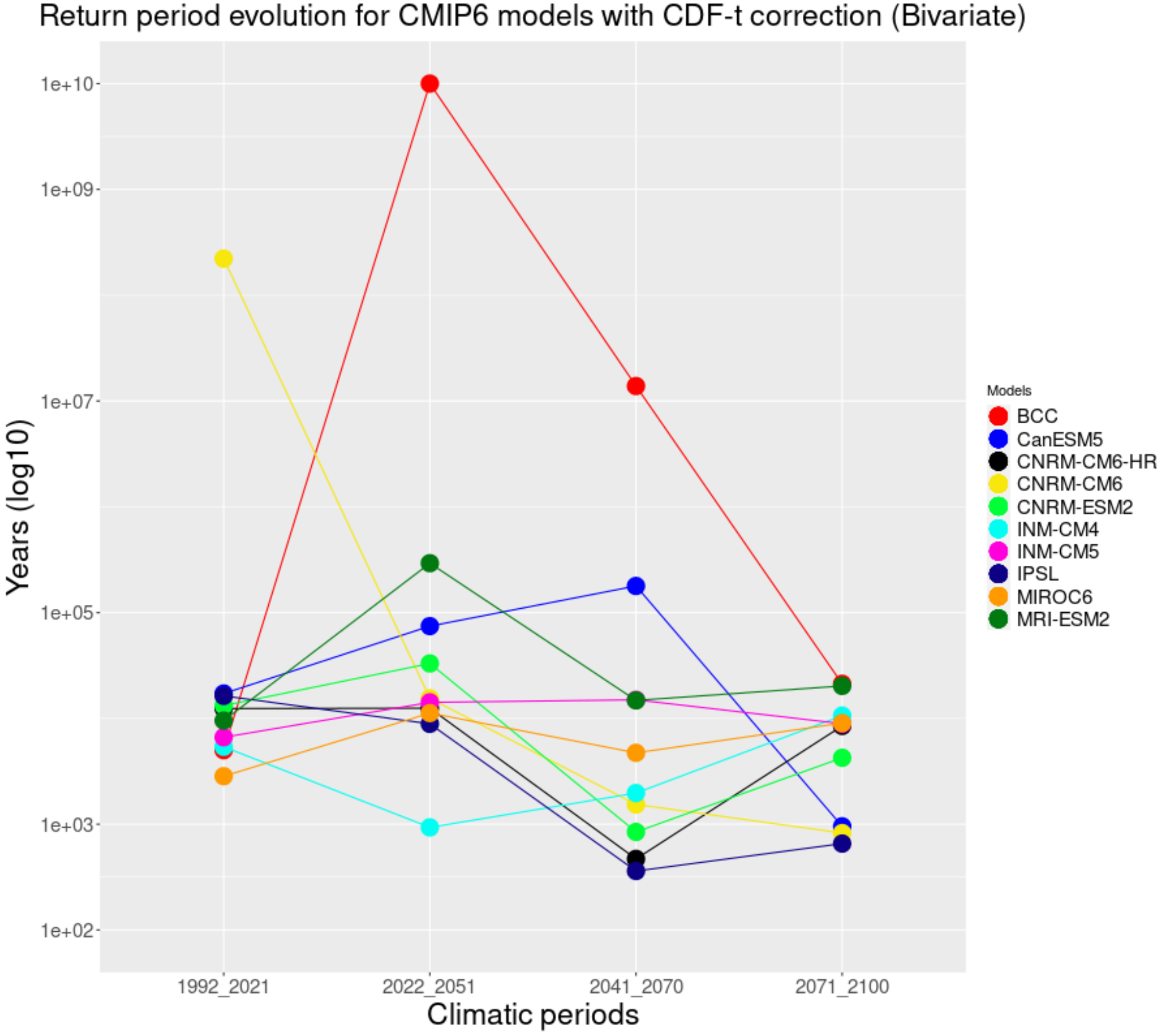


- The dashed lines represent the ERA return periods
- These return periods are calculated on the 1992-2021 period
- In univariate → CDF-t works
- In bivariate setting → values are closer, but still not fully satisfying
- Need for multivariate bias correction

.EARNING
a Science for the Environment

How does the return period evolve?





Seine/Loire event

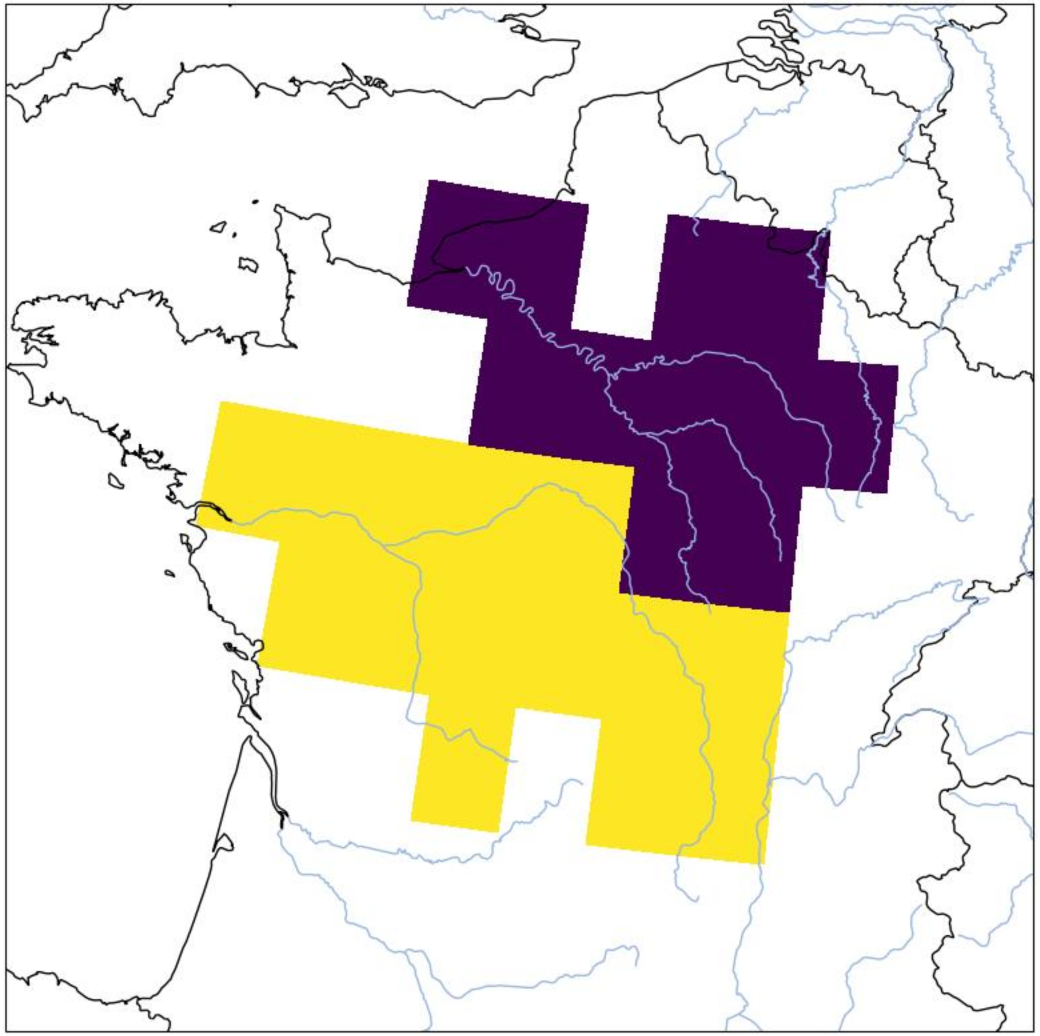
- Spatial daily precipitation averages over the Seine and the Loire watersheds for May and June

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

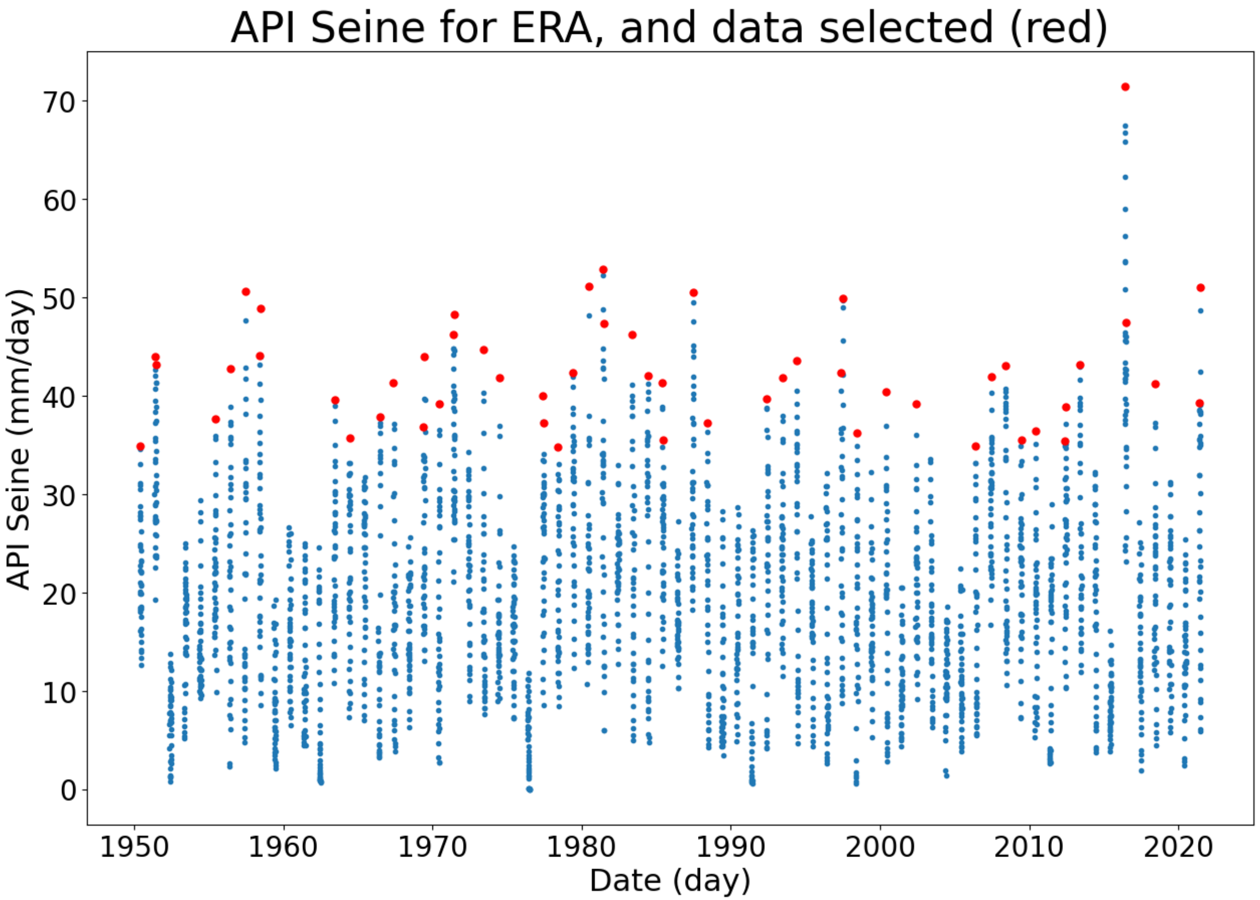
  with k = 0.9 and N = 20

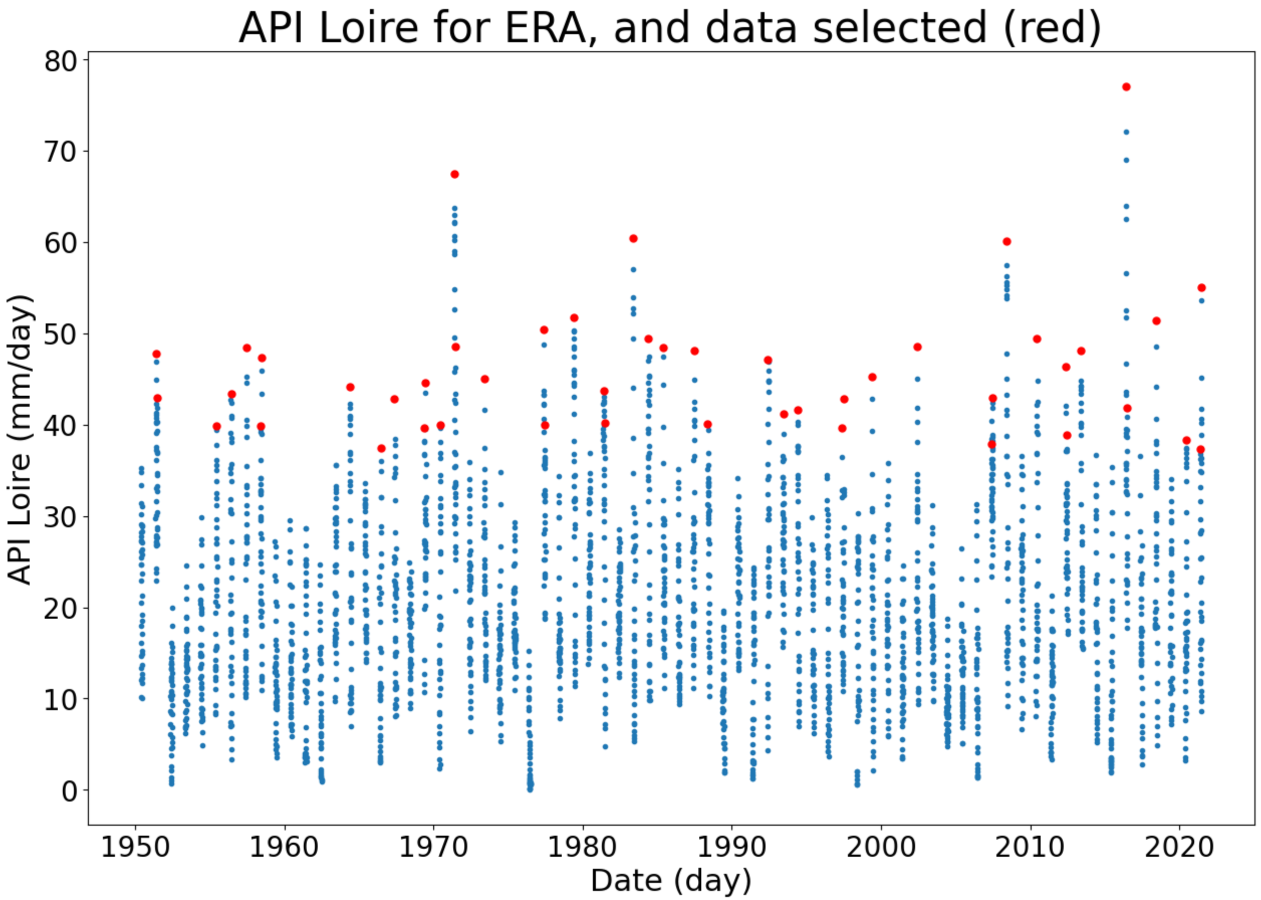
- Same methodology (Data selection, GPD model, copula…)

Seine and Loire selected areas

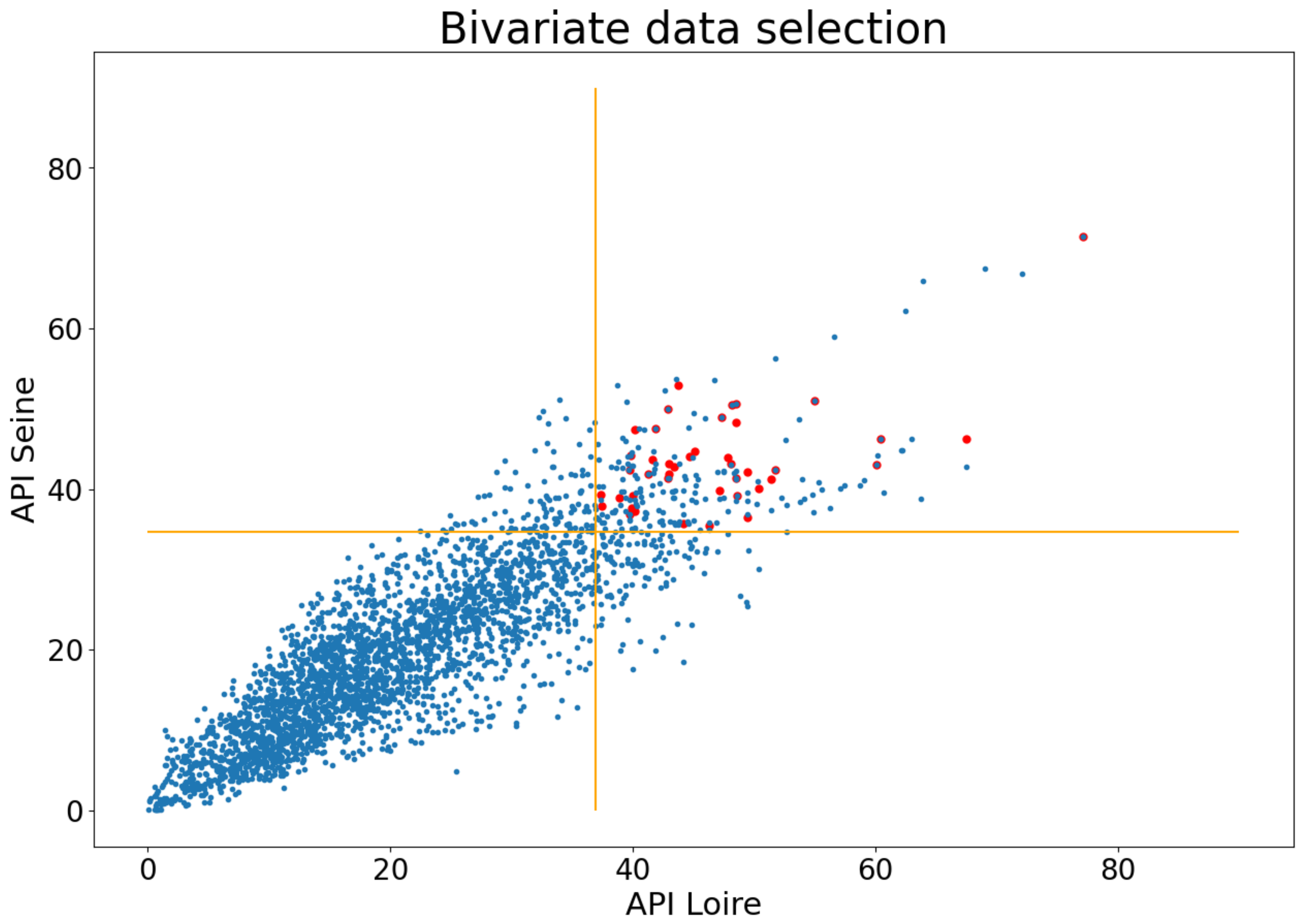


Univariate selection Seine/Loire event

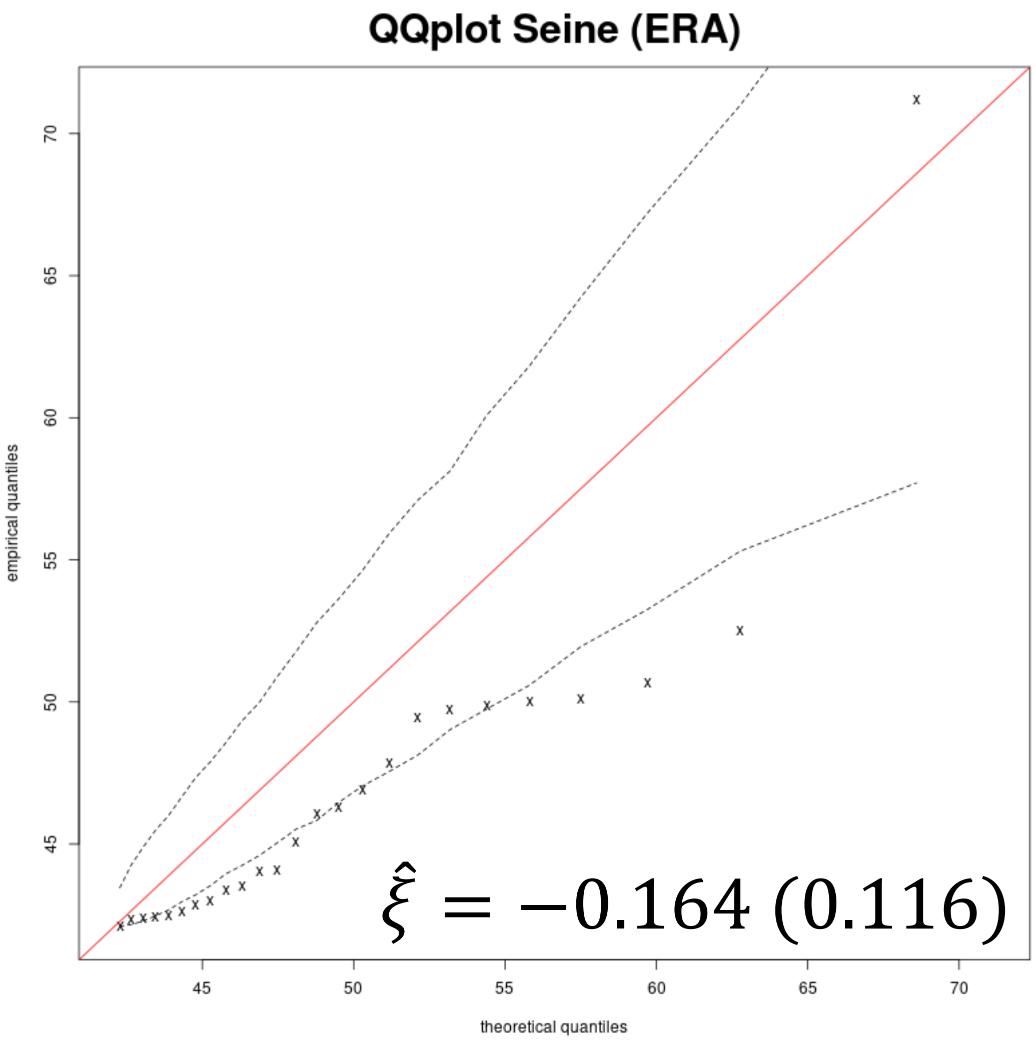


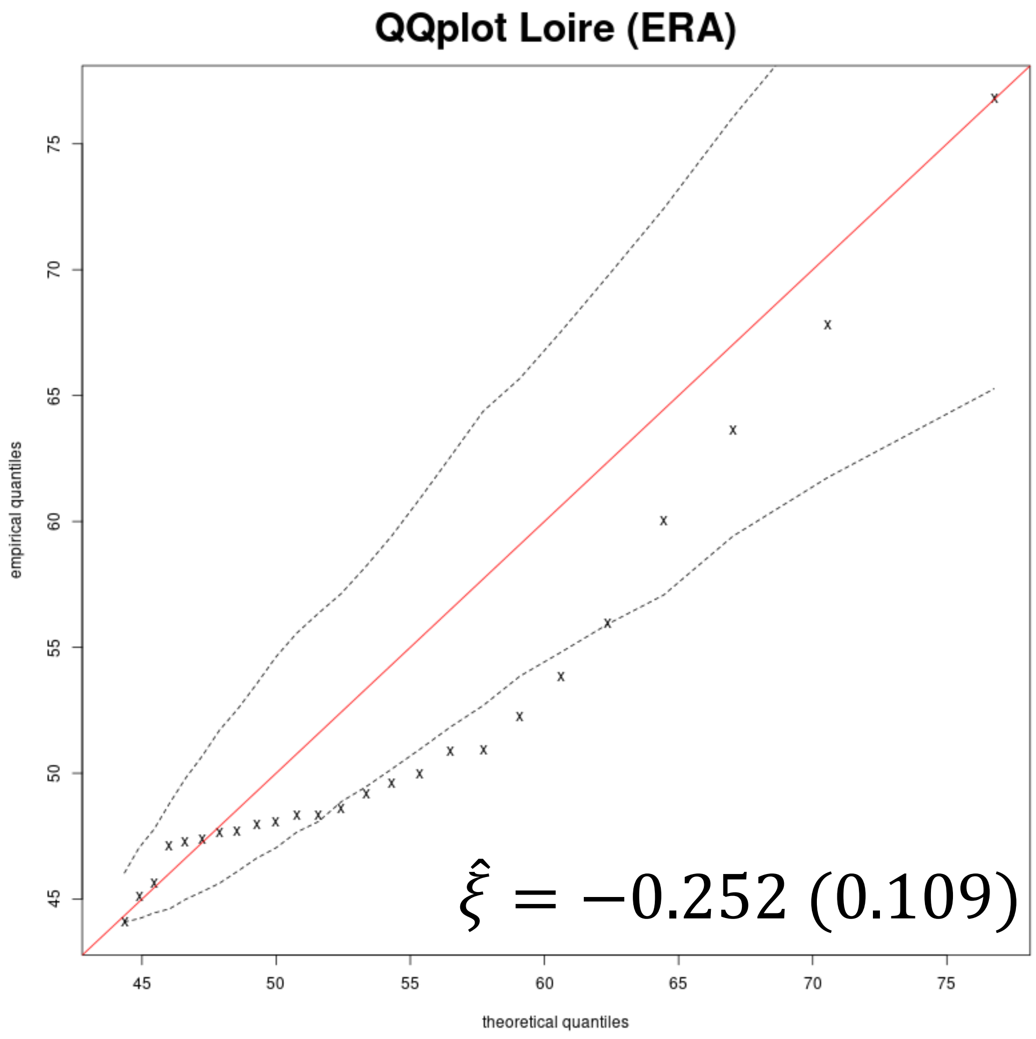


Bivariate selection Seine/Loire event

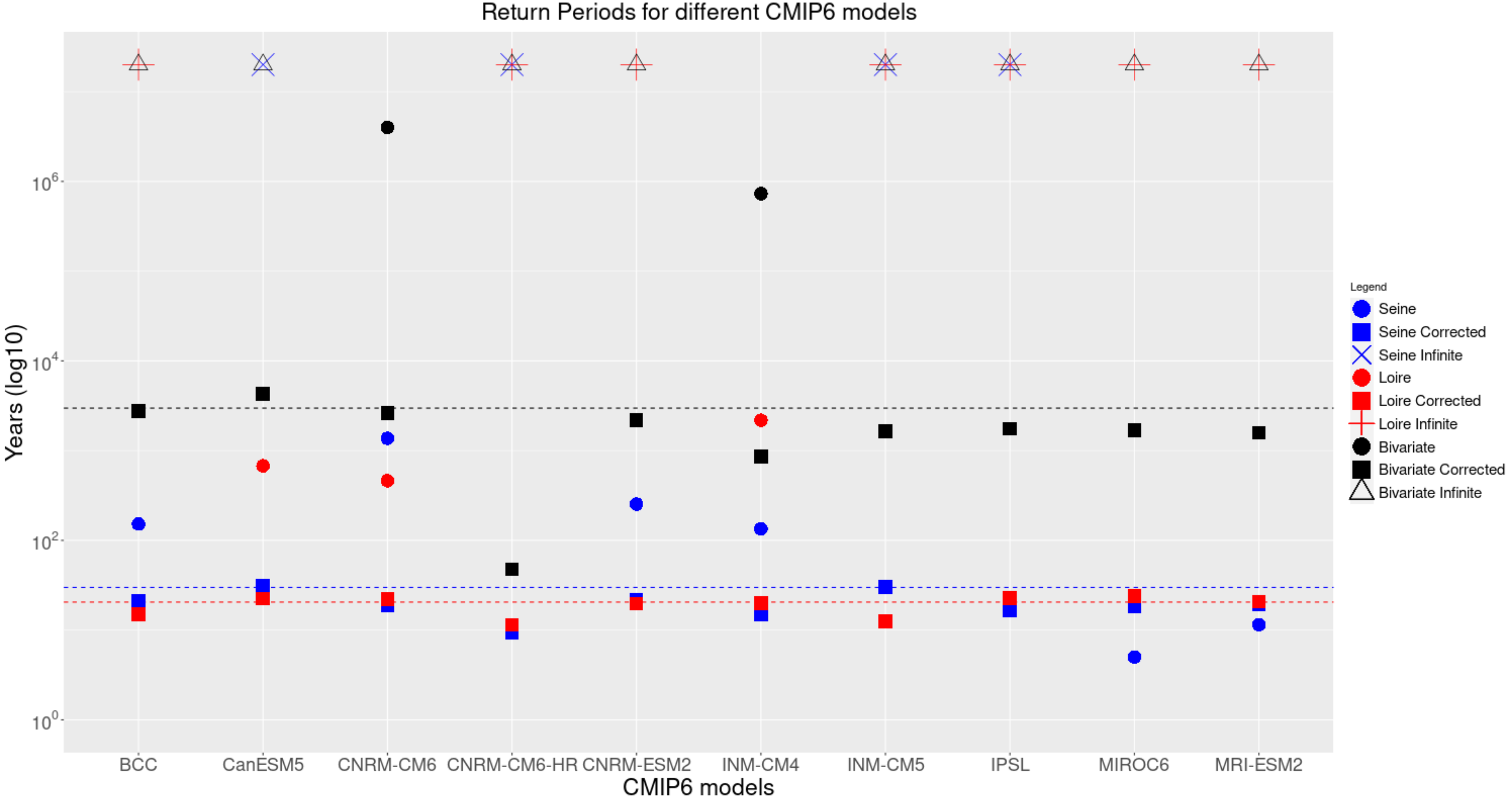


Quantile plots of GPD adjustment



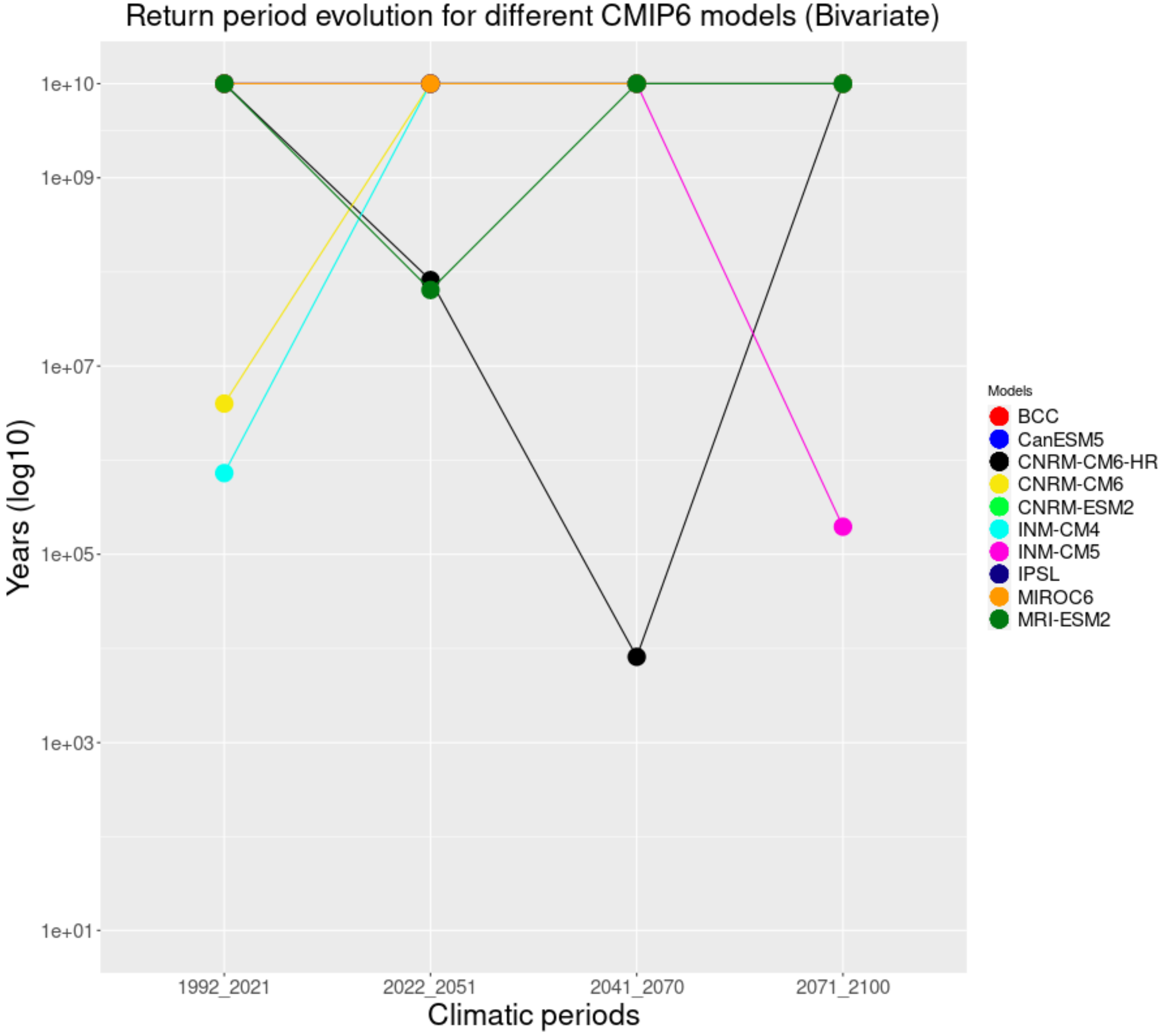


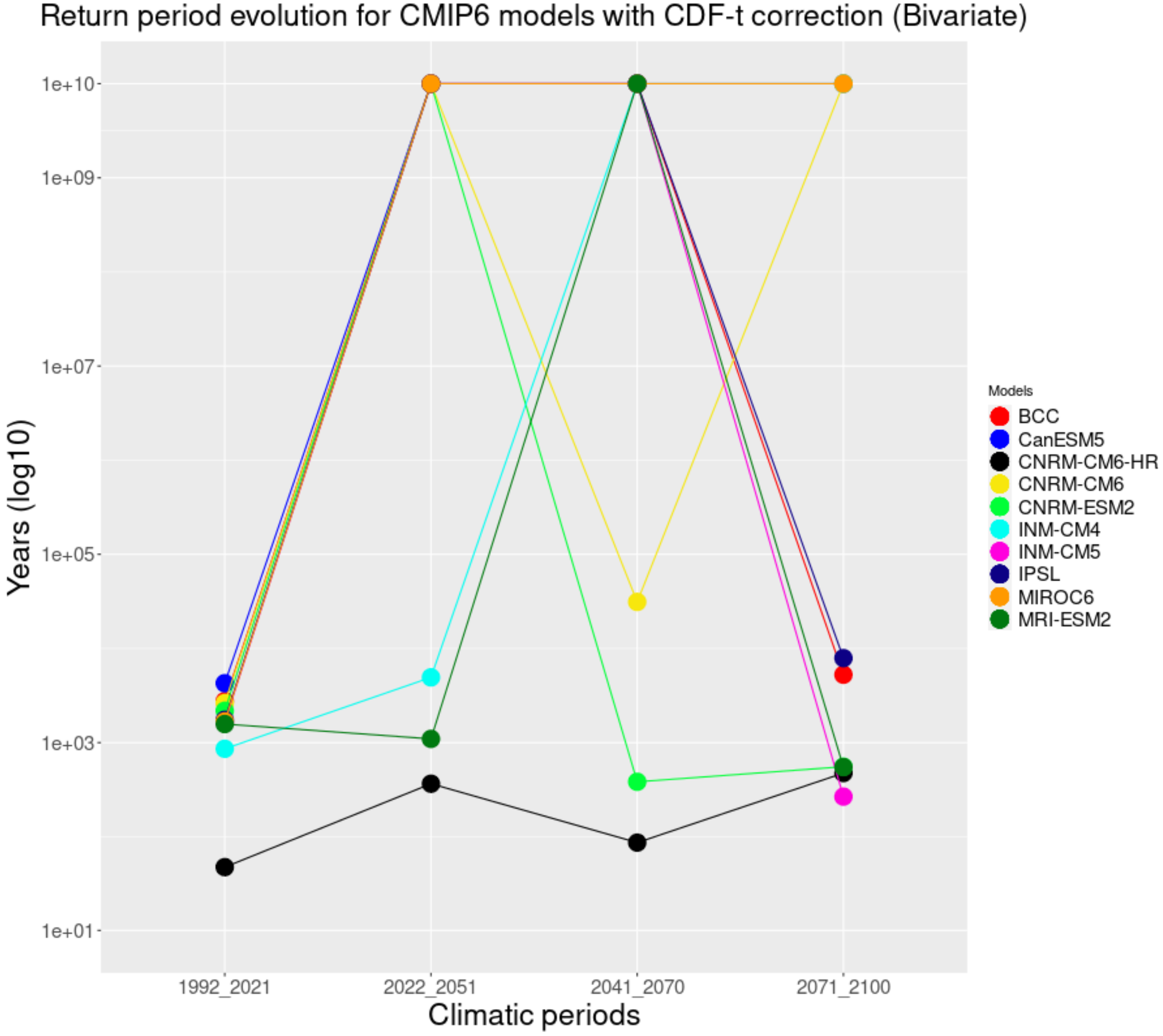
Return Periods and CDF-t correction (Seine/Loire event)



- The dashed lines represent the ERA return periods
- In bivariate setting → values are closer, but still not fully satisfying
- Need for multivariate bias correction

How does the return period evolve ?





GEOLEARNING
CHAIRE /// Data Science for the Environment

Conclusion

- We can model bivariate compound events, and calculate return periods

- With the use of a univariate bias correction method, we can get more coherent return periods for the reference period

- By correcting the projection period of the simulation, we can interpret the evolution of the climatic events

Perspectives

- Implement multivariate bias correction → correct both the marginals and the dependence structure. *(François et al., 2020)*
- Consider spatial extension over Europe
- Convective events → develop new methodological and theoretical aspects:
  - Select key variables (index, more than two variables …)
  - If more than two variables, need to extend the multivariate modeling→ Vine copula, Pareto processes …

References

- Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., ... & Wisotzky, C. (2022). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe. Part 1: Event description and analysis. Natural Hazards and Earth System Sciences Discussions, 2022, 1-44.
- van Oldenborgh, G. J., Philip, S., Aalbers, E., Vautard, R., Otto, F., Haustein, K., ... & Cullen, H. (2016). Rapid attribution of the May/June 2016 flood-inducing precipitation in France and Germany to climate change. Hydrology and Earth System Sciences Discussions, 2016, 1-23.
- Linsley Jr, R. K., Kohler, M. A., & Paulhus, J. L. (1975). Hydrology for engineers.
- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2006). Statistics of extremes: theory and applications. John Wiley & Sons.
- Bousquet, N., & Bernardara, P. (2021). Extreme Value Theory with Applications to Natural Hazards. Springer International Publishing.
- Weather and Climate Extreme Events in a Changing Climate (AR6, groupe I, chapitre 11)
- Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., & Somot, S. (2012). Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. Natural Hazards and Earth System Sciences, 12(9), 2769-2784.
- François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: which benefits for which losses?. Earth System Dynamics, 11(2), 537-562.
- Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., ... & Vignotto, E. (2020). A typology of compound weather and climate events. Nature reviews earth & environment, 1(7), 333-347.