Quantification of compound events evolution under climate change through multivariate bias correction methods

Grégoire JACQUEMIN (Mines Paris-PSL), Mathieu VRAC (LSCE), Denis ALLARD (INRAE), Xavier FREULON (Mines Paris-PSL) & Valentin ARJAILLES (ENSTA)

Journée scientifique de la chaire Géolearning, 08/04/2026

GEOLEARNING
CHAIRE /// Data Science for the Environment

ANDRA
Agence nationale pour la gestion des déchets radioactifs

BNP PARIBAS

CCR

SCOR
FONDATION POUR LA SCIENCE

MINES PARIS | PSL

INRAE

LSCE

Outline

1 Context

2 Modeling compound events

3 Data and methods

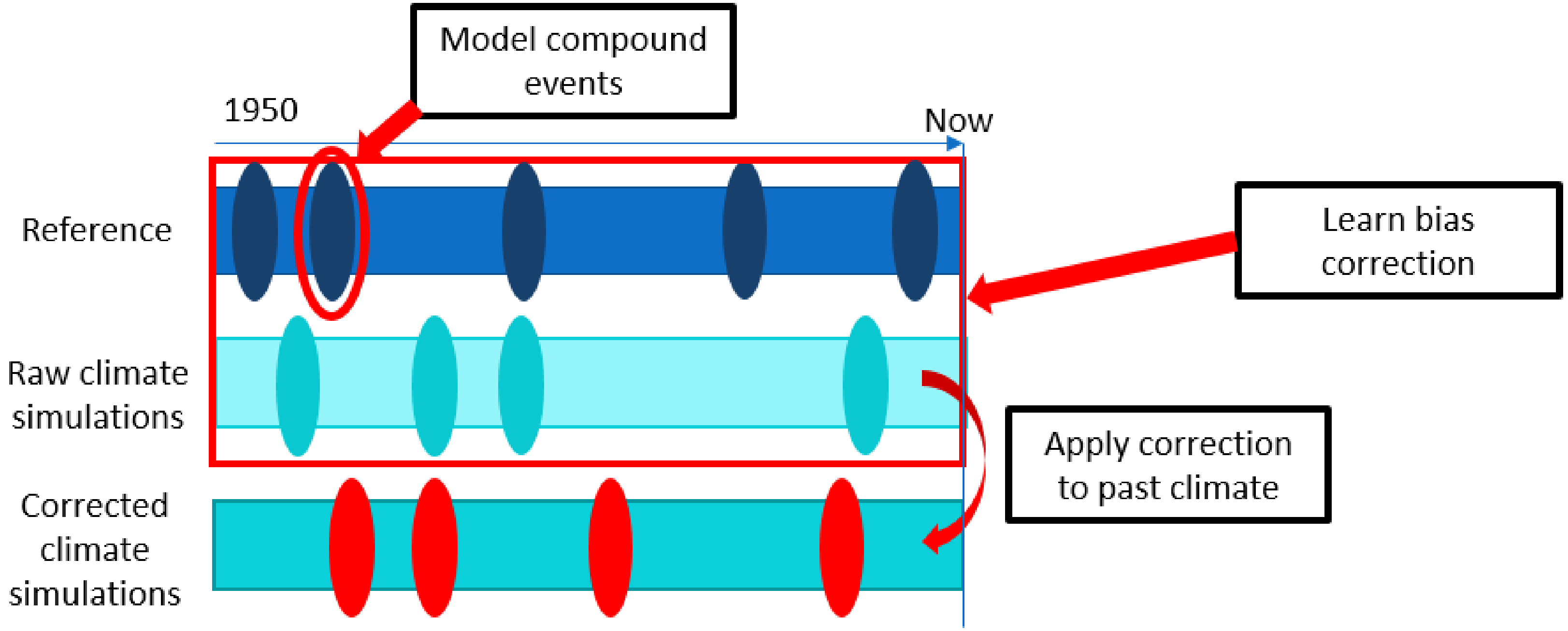
4 Bias correction of compound events

5 Extension to Europe

6 Conclusion and perspectives

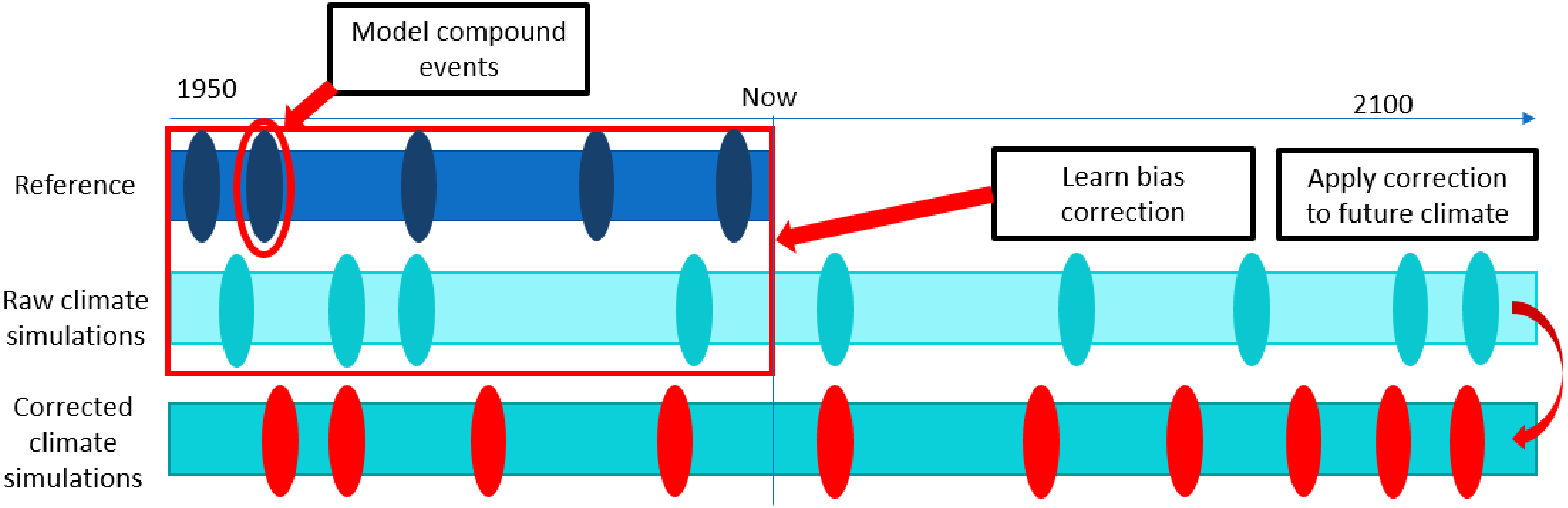
1. Context

Quantifying the evolution of compound events



1 Do bias correction methods improve the realism of compound events?

Quantifying the evolution of compound events

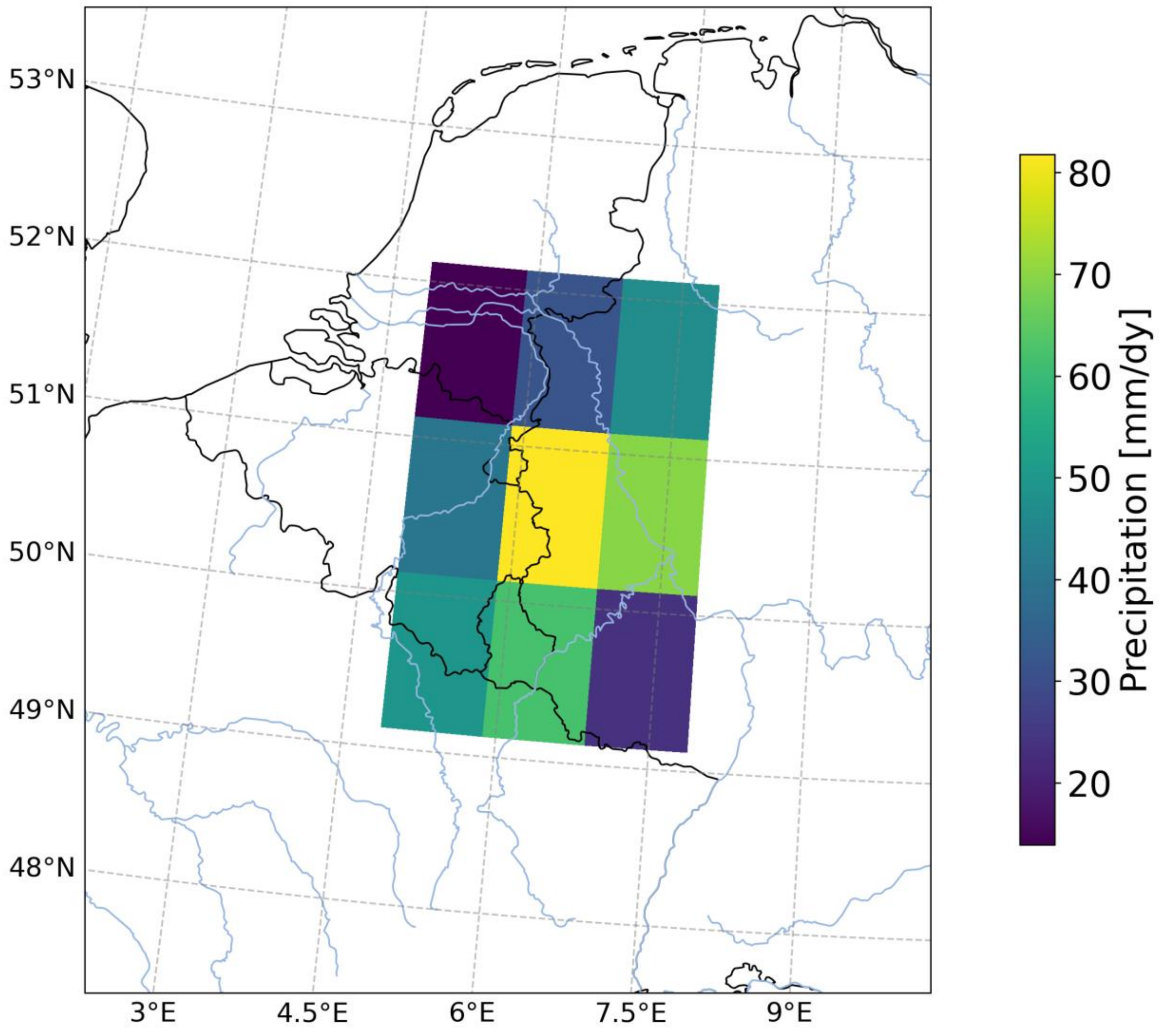


1 Do bias correction methods improve the realism of compound events?

2 Do bias correction methods preserve climate change signal seen in raw climate simulations?

The Ahr event (Belgium/Germany 2021)
Preconditioned compound event

- Extremely heavy precipitation after moderate precipitation led to a massive flood of the Ahr river in July 2021 (van Oldenborgh et al., 2016).
- The daily precipitation (PR) and the API are used to model the event.
- $API_j = \sum_{i=1}^{N} PR_{j-i} * k^{i-1}$
- The API (with k = 0.9 and N = 30) is used as a proxy for soil moisture.
- Daily precipitation are averaged over the shown area for June, July and August between 1992 and 2021 (on ERA5 1°x1° grid).



2. Modeling compound events

Extended Generalized Pareto Distribution (EGPD)

The Generalized Pareto Distribution (GPD) is used in a peaks-over-threshold context to model the exceedances of a variable above a defined threshold.

The Extended Generalized Pareto Distribution (EGPD) (Naveau et al., 2016) allows a complete modeling of the distribution.

Cumulative distribution function of the EGPD

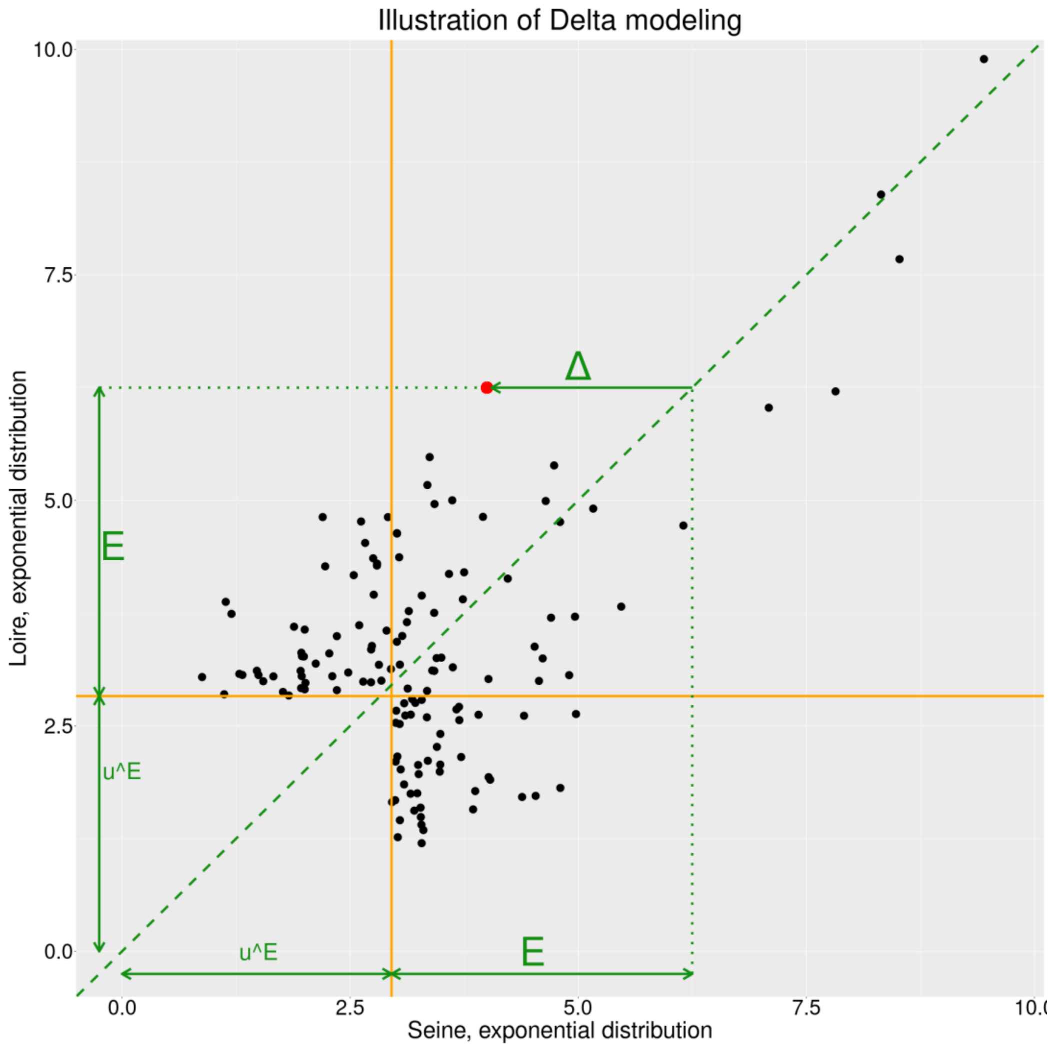
$f\big(G(x|\xi,\sigma)\big) = pG(x|\xi,\sigma)^{\kappa} + (1-p)G(x|\xi,\sigma)^{\delta}$, with $\kappa, \delta > 0$.

G is the GPD cumulative distribution function ($\xi > -0.5$).

- ▶ No threshold choice is required.
- ▶ The tail of the distribution behaves like a GPD.
- ▶ The bulk of the distribution is also modeled.
- ▶ There exists other forms for f .

Delta decomposition for bivariate GPD
The bi-GPD approach in Jacquemin et al. (2026)

- With the EGPD, margins are transformed to exponential.
- Two thresholds are defined (here the 95th quantile).
- Points below both thresholds are removed.
- For each point, a couple (E, Δ) is constructed.
- The bivariate exceedance probability of any point above both threshold can be numerically computed (**Our new method**).



Key takeaways from the article
Jacquemin et al, 2026

- This approach, called the bi-GPD approach, is formally defined and constructed.
- It is compared over simulated data to a more classical approach, the GPD-copula approach.
- In the context of asymptotic dependence, the bi-GPD approach is more accurate and less biased than the GPD-copula approach.
- Open access to the article on the Environmetrics website.

3. Data and methods

Data and materials

- 10 Global Climate Models (GCM).
- All the considered runs follow the ssp5-8.5 scenario.
- 4 Bias Correction (BC) methods are compared: no correction, CDF-t, dOTC and R2D2.
- BC methods are compared with two statistics: the bivariate return period **RP** and the coefficient of extremal dependence χ, computed with the GPD-copula approach (see Jacquemin et al, 2026).

Model Name	Run	Resolution
BCC-CSM2-MR	r1i1p1f1	~100 km
CanESM5	r10i1p1f1	~500 km
CNRM-CM6-1	r1i1p1f2	~250 km
CNRM-CM6-1-HR	r1i1p1f2	~100 km
CNRM-ESM2-1	r1i1p1f2	~250 km
INM-CM4-8	r1i1p1f1	~100 km
INM-CM5-0	r1i1p1f1	~100 km
IPSL-CM6A-LR	r14i1p1f1	~250 km
MIROC6	r1i1p1f1	~250 km
MRI-ESM2-0	r1i1p1f1	~100 km

Bias correction algorithms
Univariate

CDF-t (cumulative distribution function transform) (Michelangeli et al., 2009)

1. The transformation between the cdf (cumulative distribution function) of ERA5 and the cdf of the simulation over the same period is learn.
2. The transformation between the cdf of the simulation over the past and the cdf of the simulation over the future is learn.
3. The composition of these two transformation permits to construct the cdf of the corrected data over the future period.

Bivariate

R2D2 (rank resampling) (Vrac and Thao, 2020)

1. Apply univariate bias correction (CDF-t) on each variable separately.
2. Select one variable to be the pivot (the reference for the rank analogy).
3. Associate, in the rank space, points from the simulated data (calibration) to the reference data with respect to the pivot.
4. Replace the simulated values (projection) by the ones corresponding to the rank of the analogues.

dOTC (optimal transport) (Robin et al., 2019)

1. Multivariate optimal transport between the reference data and the model data of the reference period.
2. Multivariate optimal transport between the model data of the reference period and the projection period.
3. The two projection plans are combined to correct the projected data of the model.

Two extremal statistics

Two statistics are used for the comparison: the bivariate return period RP and the coefficient of extremal dependence χ.

▶ The return period RP is the expected waiting time between two exceedances above a "return level" x_{RP}.

▶ The univariate return period is: $\mathrm{RP} = \frac{1}{N\,\theta\, P[X > x_{RP}]}$.

▶ N is the number of data points per year and θ is the extremal index, which express the strength of the temporal dependence in extreme quantiles.

▶ A "similar" expression is available for compound events (Jacquemin et al., 2026, 10.1002/env.70063).

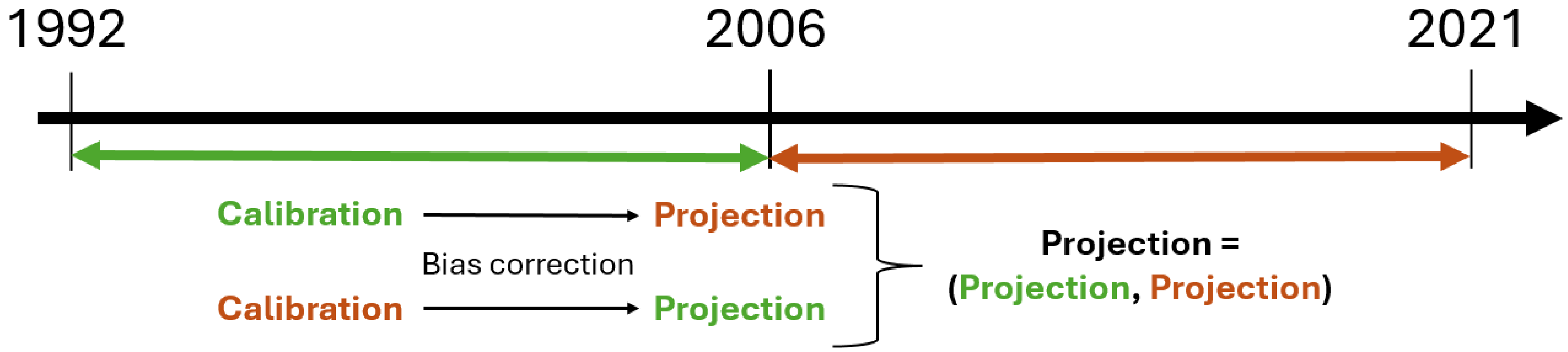
▶ The coefficient of extremal dependence χ is defined by: $\chi = \lim_{x\to\infty} P[X_1 > x \mid X_2 > x]$.

▶ 0 ≤ χ ≤ 1, with χ = 0 showing asymptotic independence and χ = 1 showing asymptotic perfect dependence.

4. Bias correction of compound events

Do bias correction methods improve the realism of compound events?
Sequential 2-fold experiment

- The first objective is to evaluate whether multivariate bias correction methods are reliable to correct extremal statistics.
- To do so, the sequential 2-fold experiment is applied over the reference period (1992-2021) for the 10 climate simulations.



Do bias correction methods improve the realism of compound events?

Results for the Ahr event

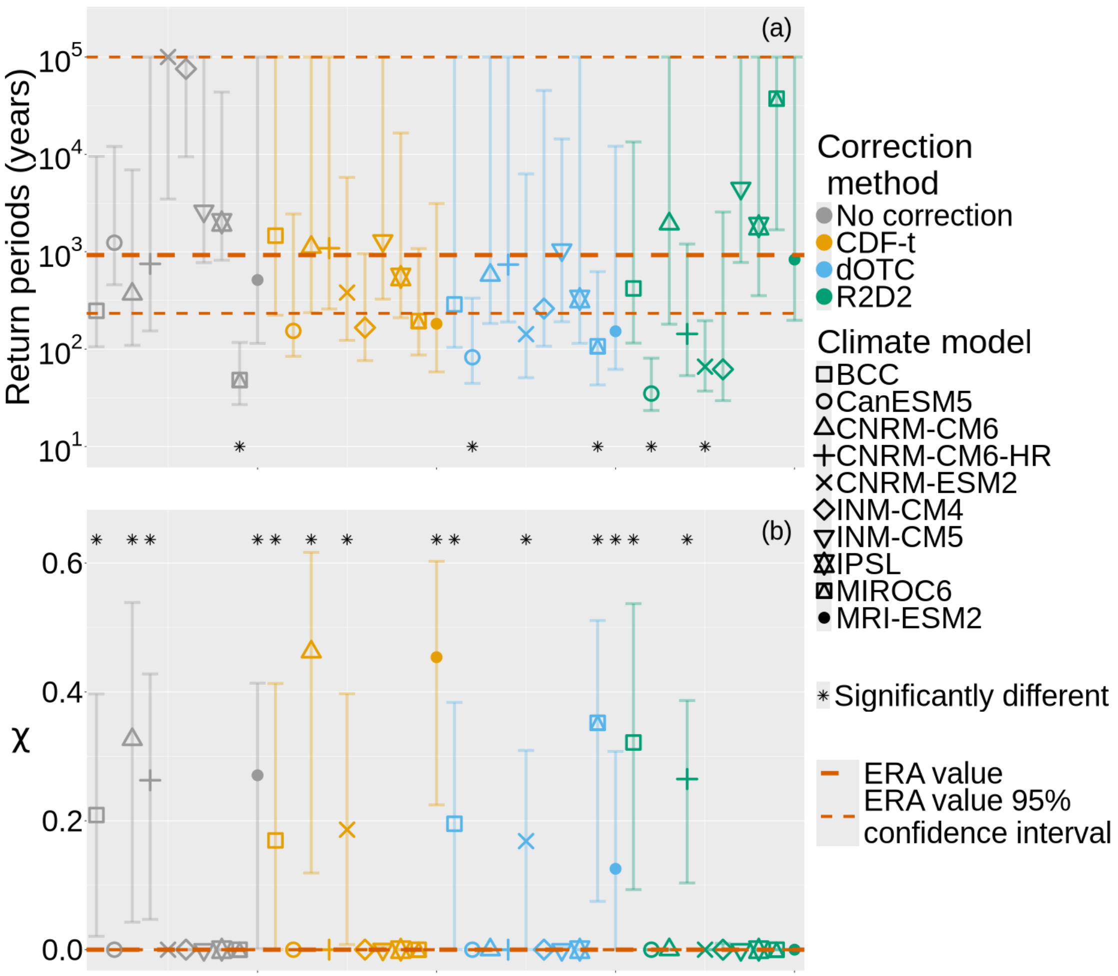
▶ Confidence intervals are constructed through bootstrap.

▶ A statistical test is performed to ensure that a value is significantly different from the ERA5 value (red dotted line).

▶ Uncorrected climate simulations are mostly correct (for **RP** and χ).

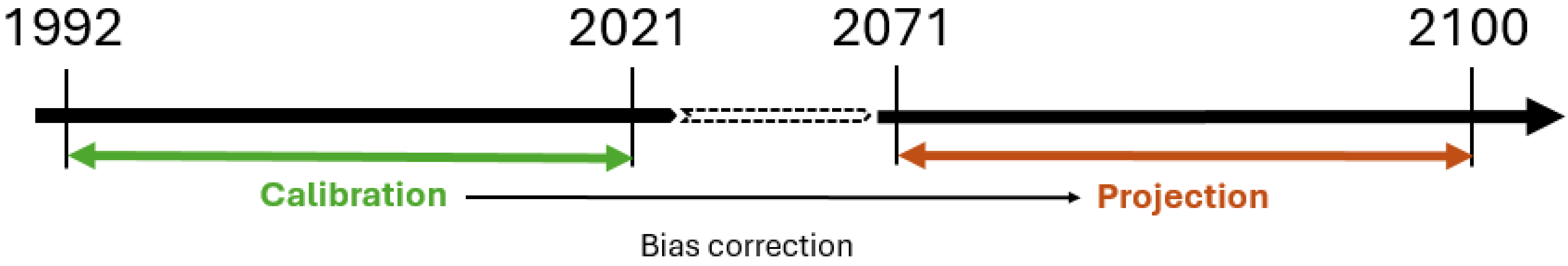
▶ CDF-t and dOTC bring most of the return periods closer to the ERA5 value, but without improvements of the χ.

▶ R2D2 struggles more than dOTC in terms of **RP** but is more efficient for χ.



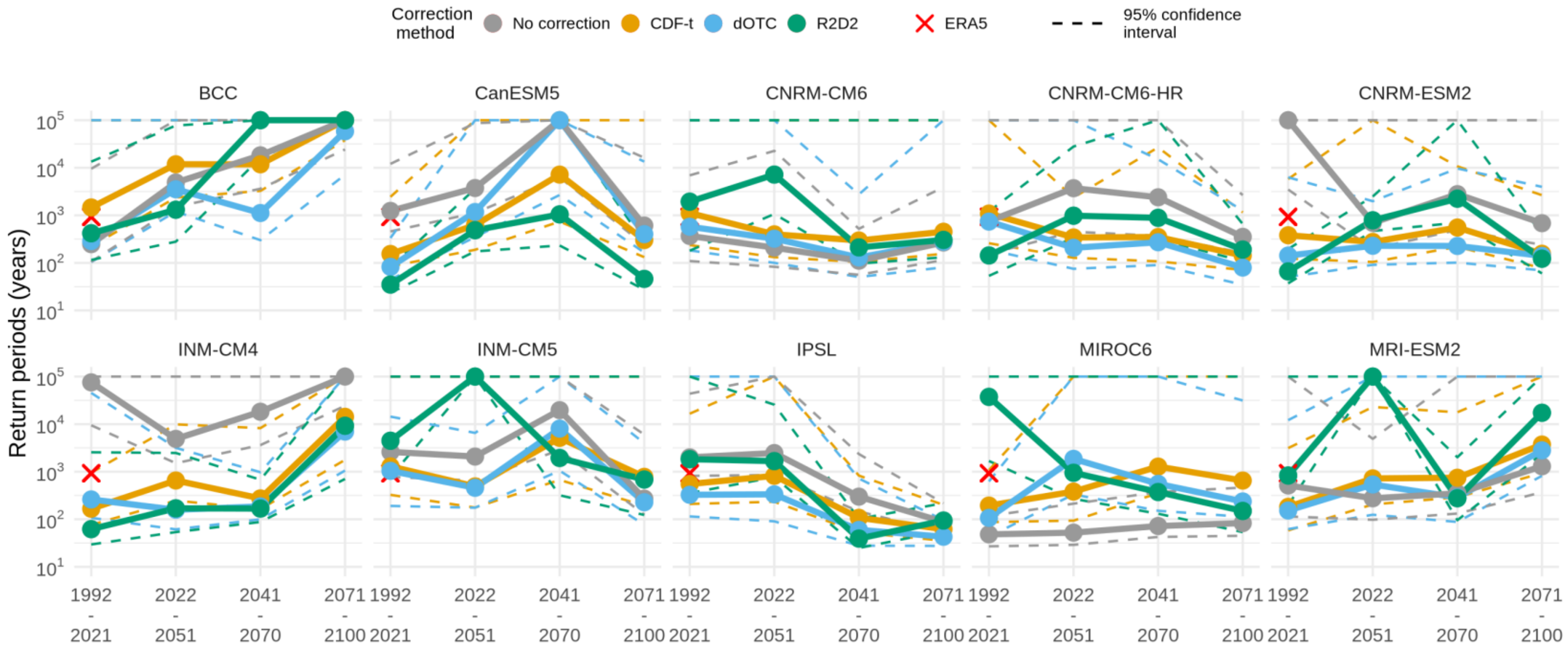
Do bias correction methods preserve climate change signal in raw climate simulations?
Experiment for future periods

- The second objective is to evaluate whether multivariate bias correction methods preserve the climate change signal of extremal statistics.
- To do so, the experiment for future periods is applied over the three projection period for the 10 climate simulations.



Do bias correction methods preserve climate change signal in raw climate simulations?

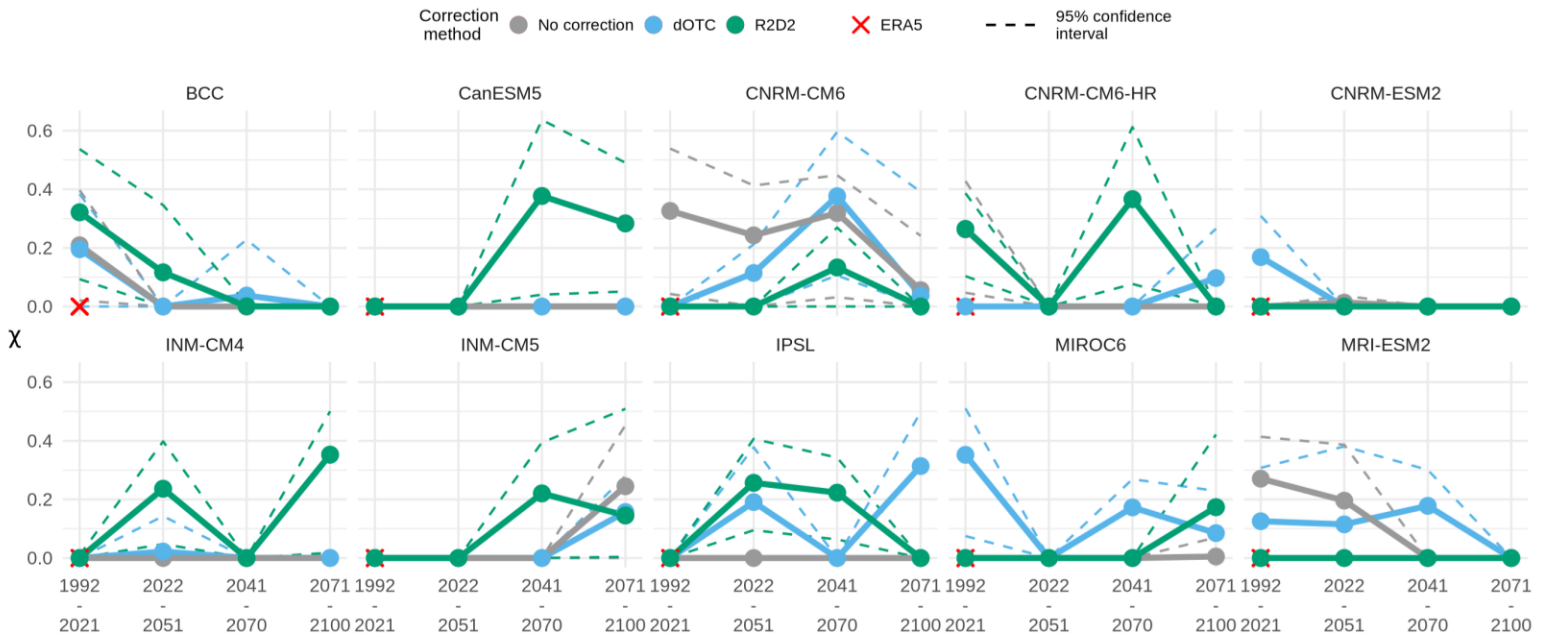
Return periods **RP**



- Great diversity of behaviors among the climate simulations.
- In around 2/3 of the cases, the sign of the evolution is preserved by the bias correction method.
- 4 simulations show an increase in **RP**, 3 show a decrease and 3 show no significative evolution.

Do bias correction methods preserve climate change signal in raw climate simulations?

Coefficient of extremal dependence χ



- CDF-t is not represented as it does not modify the dependence structure.
- BC methods may introduce some unexpected variations.
- No climate simulations show a significative evolution in terms of dependence structure.

5. Extension to Europe

Extension to Europe
Methodology for the extension

The objective is to compute **RP** and χ on every grid cell over Europe for an event “similar” to the Ahr event.

▶ On each grid cell, the variables PR and API are constructed similar to the Ahr event.

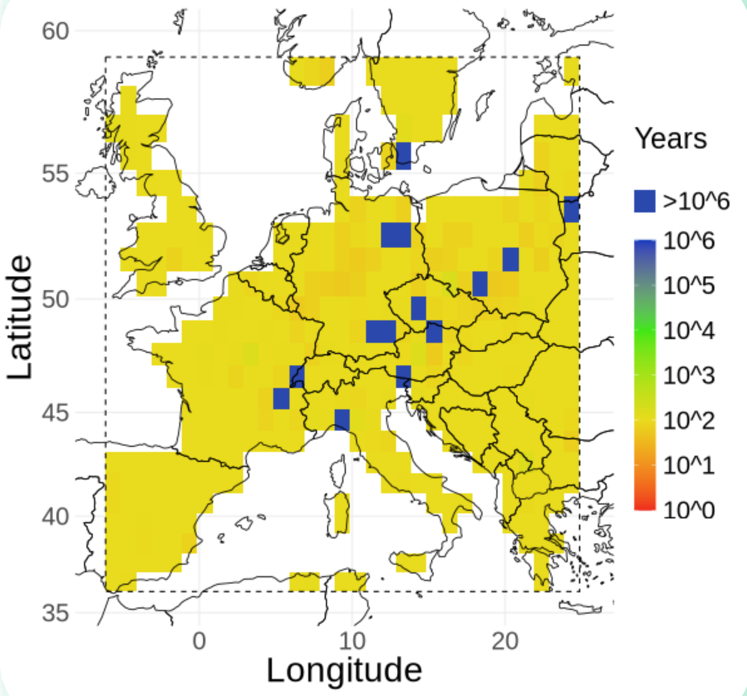
▶ The study is done with the IPSL climate simulation, and the CDF-t bias correction method for the projection period 2071-2100.

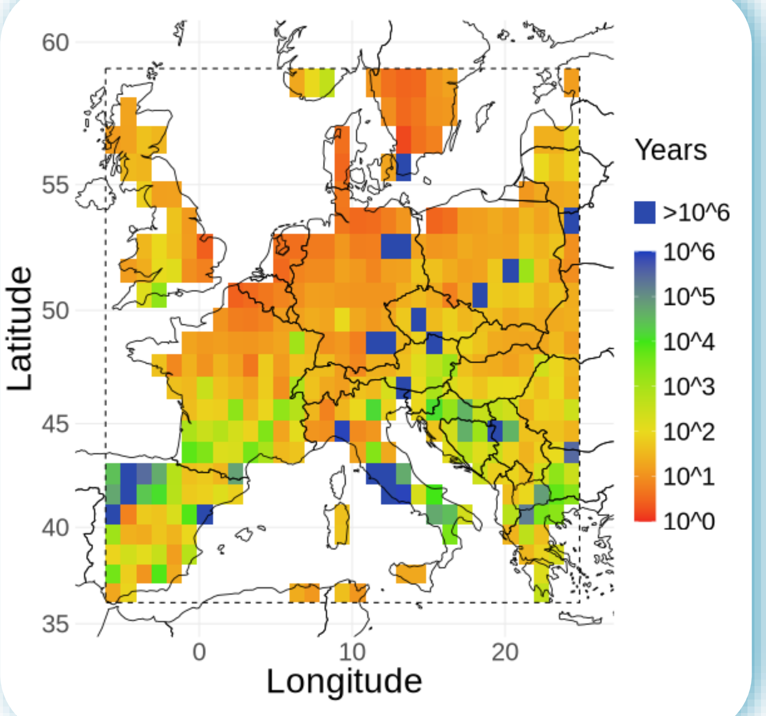
▶ Different choices can be made for the extension of the return levels. Here, return levels are deduced from a 100 years return period for ERA5 data.

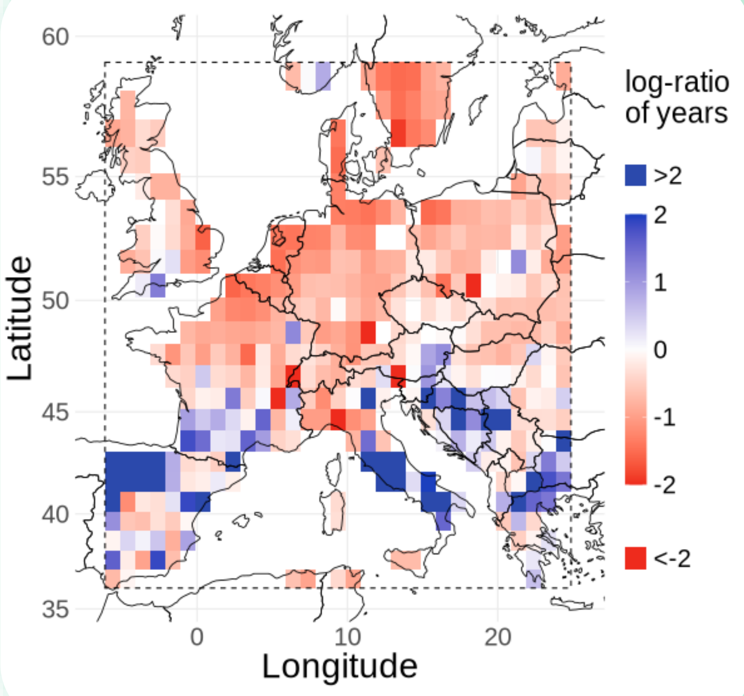
▶ For the reference period (1992-2021, ERA5 data), the return period is 100 years for each cell. For the projection period (2071-2100, IPSL data corrected with CDF-t), the return period corresponds to a100-year event of today, “similar to the Ahr event”.

▶ The log-ratio of the two return periods is computed: $Evolution = \log_{10} \frac{\boldsymbol{RP}_{Future}}{\boldsymbol{RP}_{Reference}}$, with $\boldsymbol{RP}_{Reference} = 100$ here.

Extension to Europe
Evolution of the return period **RP**







Return periods

ERA5 data, 1992-2021

Return periods

IPSL data corrected,

2071-2100

Evolution

Red indicates an increase in probability of occurrence, and blue a decrease.

Extension to Europe
Evolution of the spatial dependence

▶ This is an extension of the Seine/Loire event, with grid cells replacing watersheds.

▶ The API of each grid cell is considered.

▶ A specific grid cell is chosen, usually a major European city under flood threats (Paris, Prague, etc.)

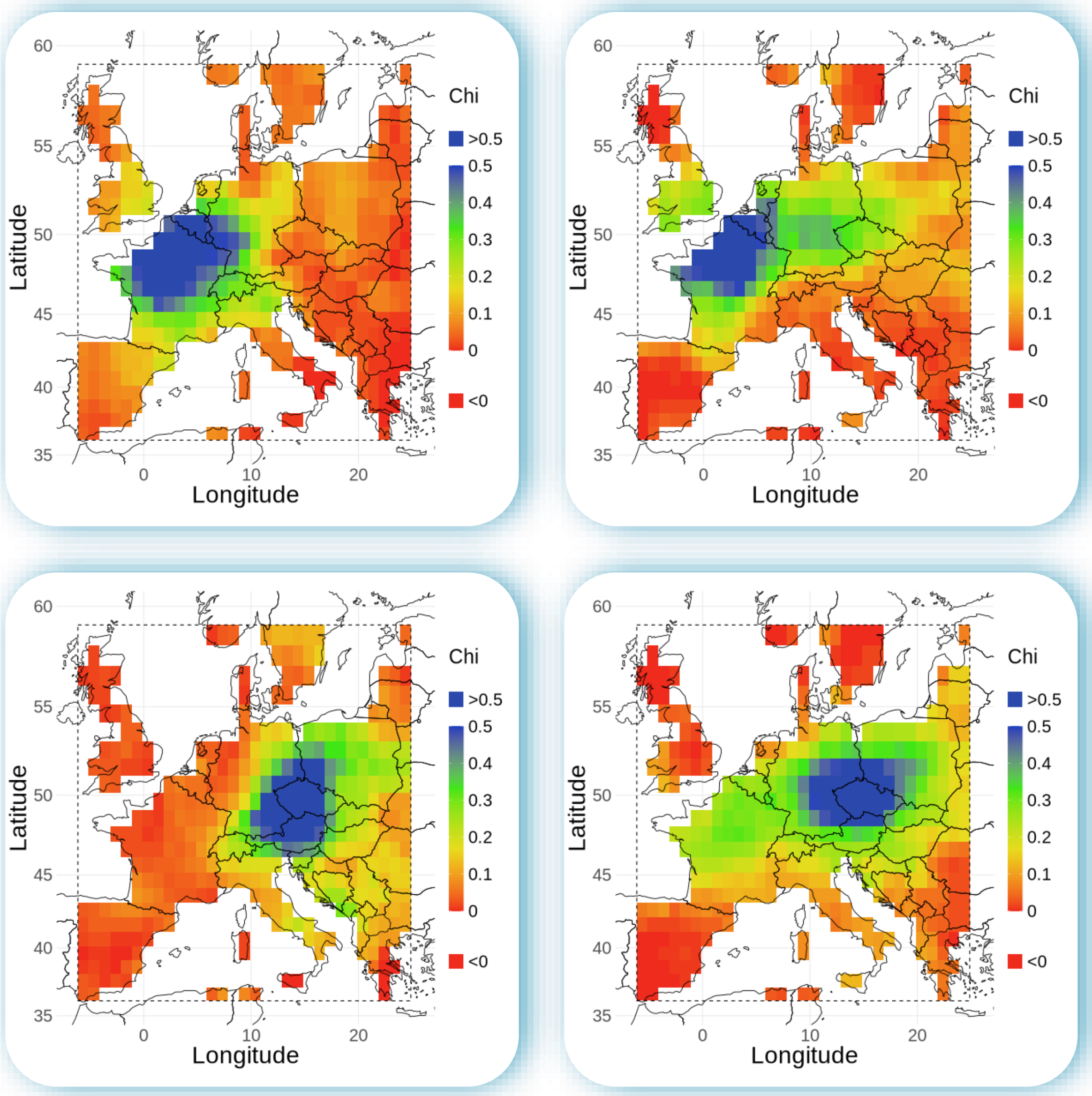
▶ The χ is computed between the API of the grid cell of the city, and the API of all the other grid cells.

▶ This is done for ERA5 data over the reference period (1992-2021) and for corrected IPSL data over the projection period (2071-2100).

1992-2021 **2071-2100**

Paris

Prague



6. Conclusion and perspectives

Conclusion and perspectives

Conclusions

- Necessity to use bias correction when projecting the evolution of compound events.
- Multivariate BC methods are required when the dependence is not correctly represented in the climate simulations. Otherwise, CDF-t is sufficient.
- Climate simulations show a wide variety of evolutions.
- At the European level, for the IPSL model, the “Ahr event” may become more frequent in northern Europe and less frequent in southern Europe.
- The dependence between extreme events may spatially evolve across Europe due to the climate change.

Perspectives

- Confirm return period evolutions with other climate simulations.
- Confirm the evolution of the spatial dependence with other climate simulation and multivariate bias correction methods.
- Similar study about hot and dry compound events has started and will be continued.

Thank You

Thanks to the **Geolearning chair** and its partners https://chaire-geolearning.org/

Thanks to TRACCS-EXTENDING and its partners: https://pepr-traccs.fr/projet/pc4-extending/