

Climate Extreme Attribution and Extreme Value Theory with a focus on RECORDS

joint work with Paula Gonzalez, Soulivanh Thao, Julien Worms (UVSQ)



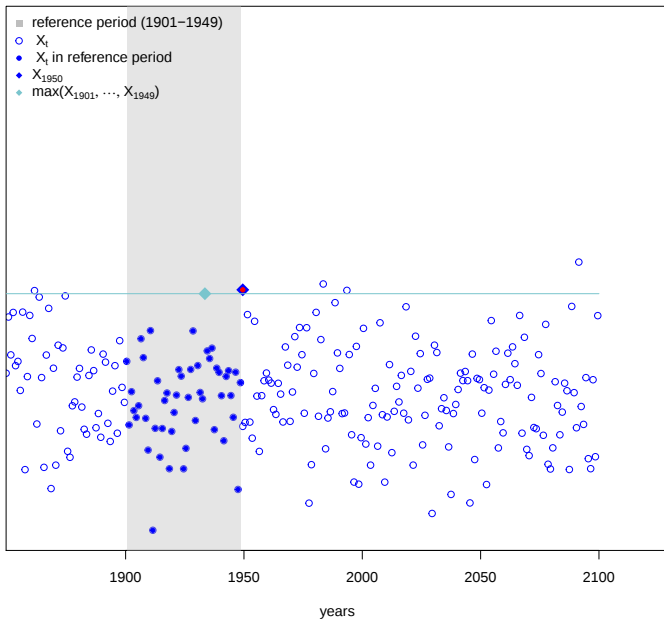
LSCE

LABORATOIRE DES SCIENCES DU CLIMAT
& DE L'ENVIRONNEMENT

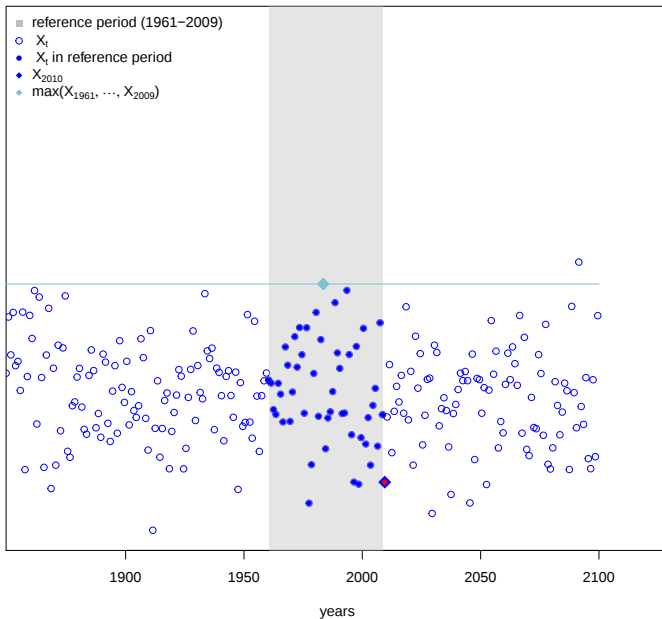
philippe.naveau@lsce.ipsl.fr

A record : the latest observation over a reference period is the largest

50-year record

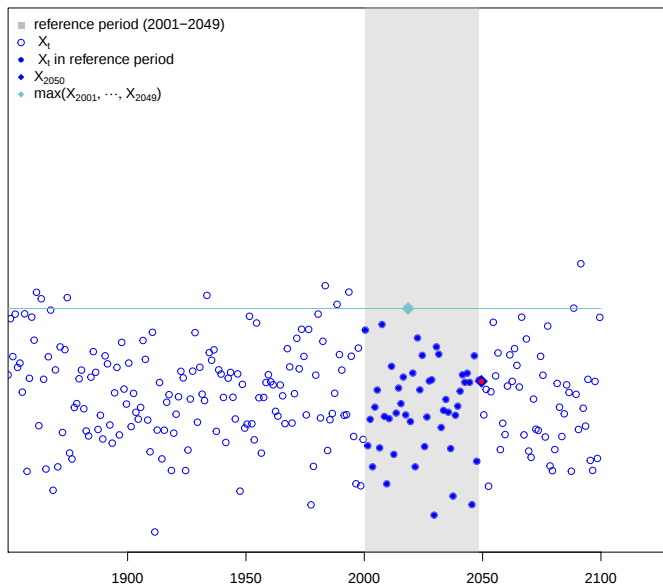


50-year record



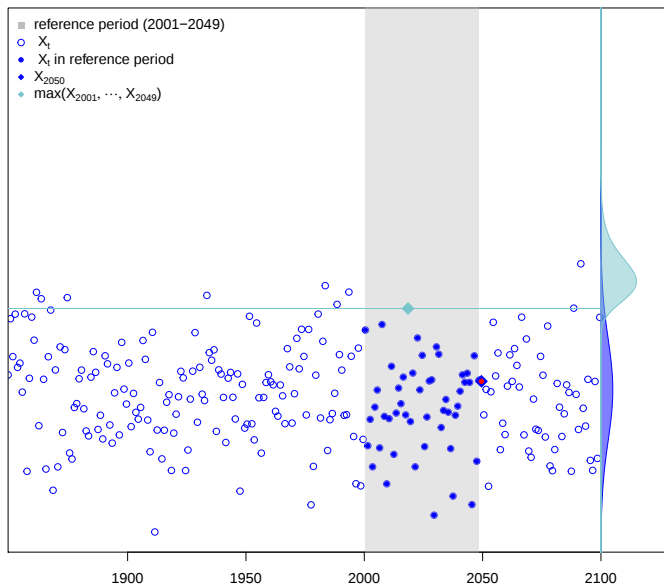
Records

50-year record



Records : $\mathbb{P}(X_t > \max(X_{t-1}, X_{t-2}, \dots, X_{t-r+1}))$

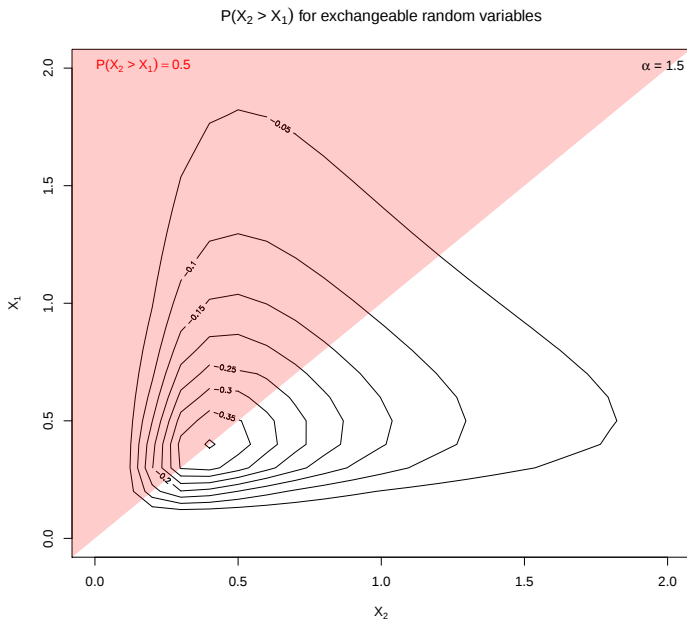
50-year record



$$\mathbb{P}(X_t > \max(X_{t-1}, X_{t-2}, \dots, X_{t-r+1})) = ?$$

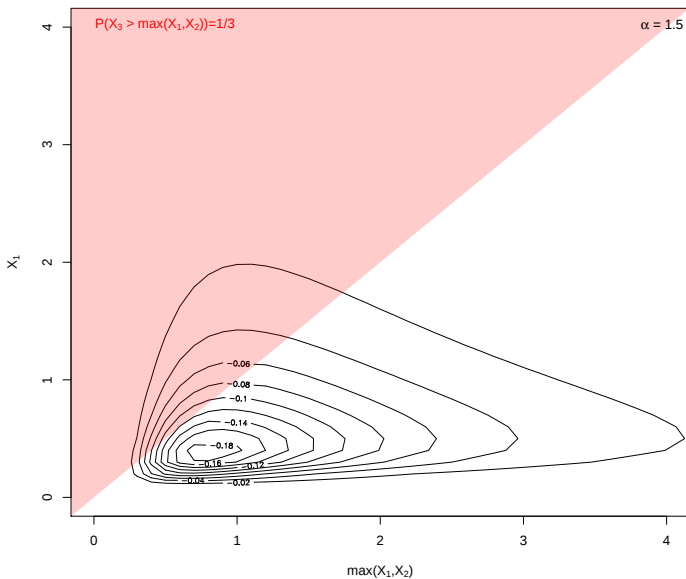
$$\mathbb{P}(X_2 > X_1) = ?$$

Example with max-stable logistic model



Example with max-stable logistic model

$P(X_3 > \max(X_1, X_2))$ for exchangeable random variables



Exchangeable random sequences

$$\mathbb{P}(X_t > \max(X_{t-1}, X_{t-2}, \dots, X_{t-r+1})) = \frac{1}{r}$$

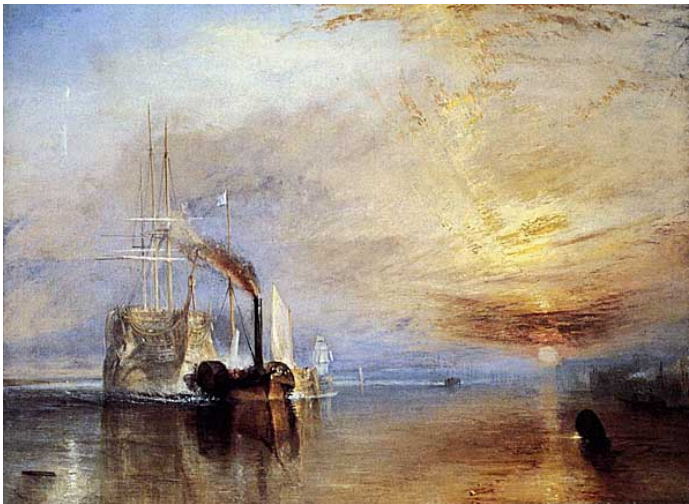
Exchangeable random sequences

$$\mathbb{P}(X_t > \max(X_{t-1}, X_{t-2}, \dots, X_{t-r+1})) = \frac{1}{r}$$

- No need to view data to compute this probability of record : *conceptual yardstick*
- This equality is *relative* and does not depend on the marginal type : *bypassing bias-correction*
- if $u = \max(X_{t-1}, X_{t-2}, \dots, X_{t-r+1})$ then $\mathbb{P}(X_t > u) = \frac{1}{r}$: *detection and attribution*
- Exchangeability is weaker than independence : *GCM ensemble dependence*

Detection and Attribution

Antropogenic forcings



Turner, The Fighting Temeraire - towed to her Last Berth to be broken up :
1838-39

Tambora 1815 (illustrations by G. & W.R. Harlin)



- ⇒ **Plutarch** noticed that the **eruption of Etna** in 44 B.C. attenuated the sunlight and caused crops to shrivel up in ancient Rome.
- ⇒ **Benjamin Franklin** suggested that the **Laki eruption** in Iceland in 1783 was related to the abnormally cold winter of 1783-1784.
- ⇒ **Mary Shelley, Frankenstein** (1818)
- ⇒ 1883 eruption of Krakatoa (Indonesian)
- ⇒ 1991 eruption of Mount Pinatubo

Detection & Attribution

Attribution

Evaluating the relative contributions of multiple causal factors¹ to a change or event with an assignment of statistical confidence.

Consequences

Need to assess whether the observed changes are

- consistent with the expected responses to external forcings (PS)
- inconsistent with alternative explanations (PN)

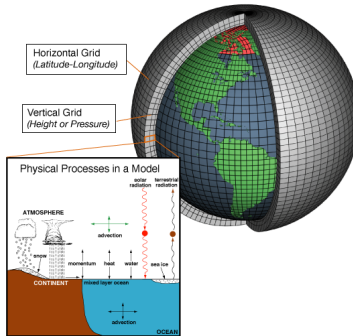
see, e.g., Naveau, Hannart, A Ribes (2020), Statistical methods for extreme event attribution in climate science, Annual Reviews of Statistics and its Application

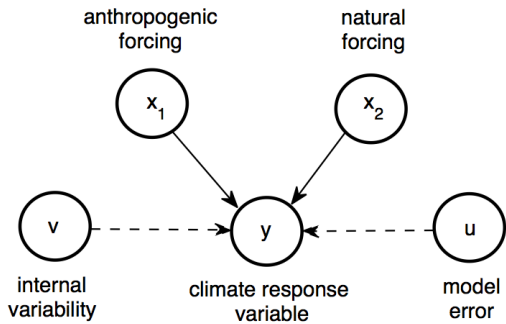
1. casual factors usually refer to external influences, which may be anthropogenic (GHGs, aerosols, ozone precursors, land use) and/or natural (volcanic eruptions, solar cycle modulations)

Main issue : only one earth

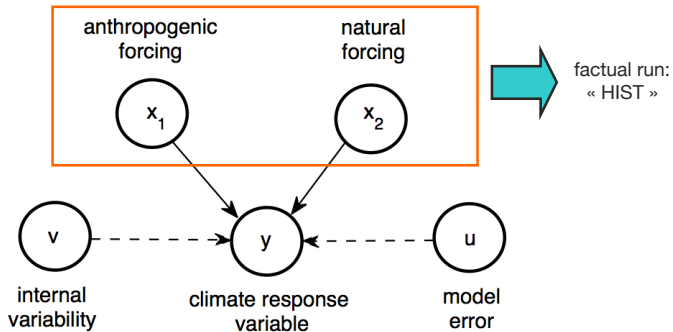
Climatology

- Global numerical physically based climate model (CMIP database)

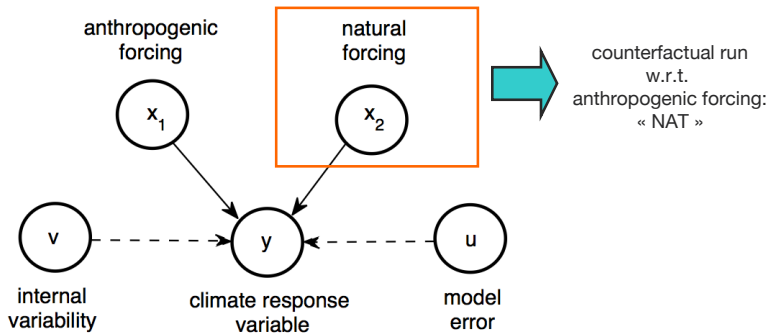




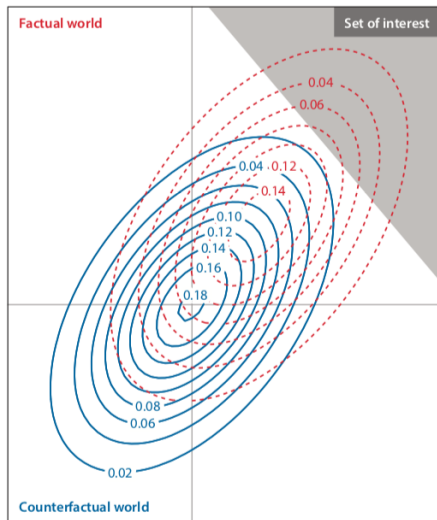
Factual world



Counterfactual world



Counterfactual and factual world in a probabilistic framework



Notations

- Counterfactual world

Sample **X** with $G(x) = \mathbb{P}(X \leq x)$

- Factual world

Sample **Z** with $F(z) = \mathbb{P}(Z \leq z)$

Notations

- Counterfactual world

Sample **X** with $G(x) = \mathbb{P}(X \leq x)$

- Factual world

Sample **Z** with $F(z) = \mathbb{P}(Z \leq z)$

Main quantities of interest in EEA where u is a high observed value

- Counterfactual world $p_0(u) = \mathbb{P}(X > u)$

- Factual world $p_1(u) = \mathbb{P}(Z > u)$

- Relative ratio

$$RR(u) = \frac{p_1(u)}{p_0(u)}$$

- Fraction of Attributable Risk (FAR)

$$FAR(u) = 1 - \frac{p_0(u)}{p_1(u)}$$

- Quantile changes

$$\Delta_t = I_t^{(F)} - I_t^{(G)} = F^{-1}(q_u) - G^{-1}(q_u) \text{ with } q \text{ quantile w.r.t. } u$$

In theory there is no difference between theory and practice. In practice there is²

2. see Benjamin Brewster from the Yale Literary Magazine of February 1882

Choices to be made in Extreme Event Attribution (EEA)

- Definition of the variable/index of interest
- Spatial and temporal resolution of the problem
- Statistical techniques to analyze the chosen data

Choices to be made in Extreme Event Attribution (EEA)

- Definition of the variable/index of interest
- Spatial and temporal resolution of the problem
- Statistical techniques to analyze the chosen data

Difference sources of error and/or uncertainty in D&A

- Natural climate internal variability
- Natural forcing variations
- Model uncertainty from approximating the true climate system with numerical experiments
- Observational uncertainties due to instrumental errors, homogenization pbs and mismatches in data sources
- Sampling uncertainty in space and time
- Statistical modeling error by assuming a specific statistical model, e.g., assuming a generalized extreme value distribution for block maxima
- Inferential uncertainties
- A rare event in the factual world can be almost impossible in the counterfactual world

Different approaches to infer $p_0(u) = \mathbb{P}(X > u)$ and $p_1(u) = \mathbb{P}(Z > u)$

Conditional (F. Otto, P. Stott and others)

$$\mathbb{P}(X > u | \text{SST}_u) \text{ and } \mathbb{P}(Z > u | \text{SST}_u)$$

- + Add local physical information
- Computationally intensive (numerous regional models runs) and no reproducible for a different event type

Unconditional (F. Zwiers, A. Ribes and others)

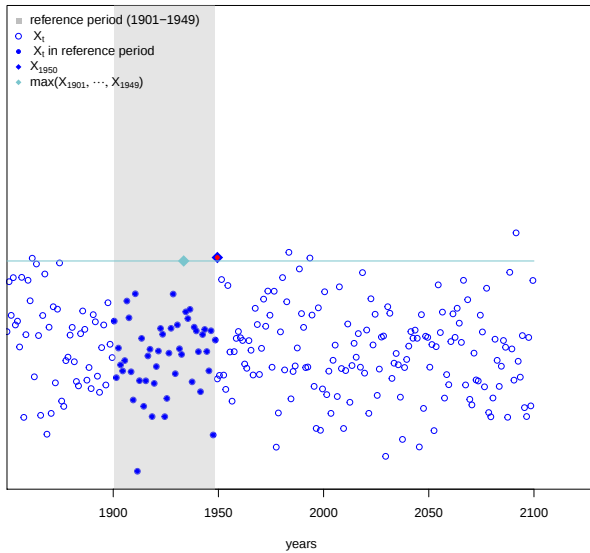
- + Computationally simple and reproducible for a different event type
- Not adapted for localized events

Story lines (T. Sheppard and others)

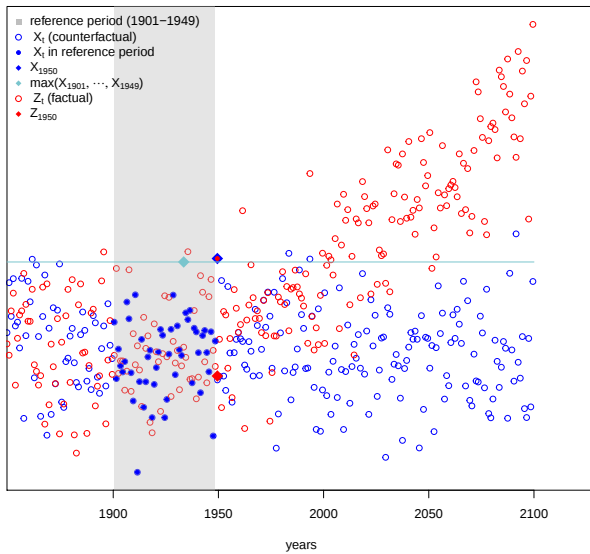
- + Meteorological interpretation of probable causes
- Not adapted to infer probabilities and uncertainties

Records in the context of Detection and Attribution

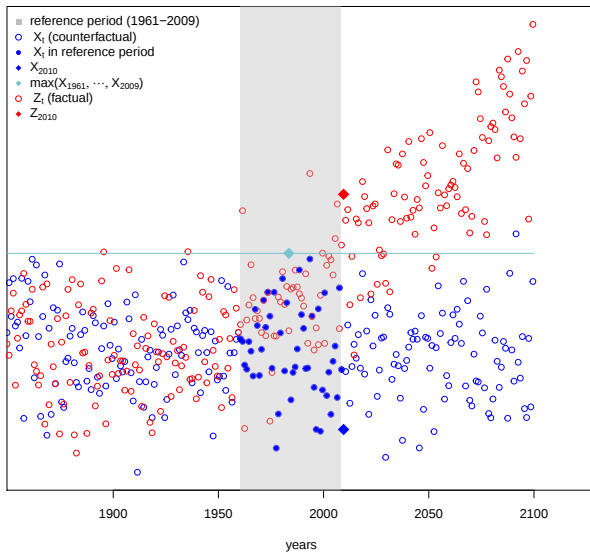
50-year record



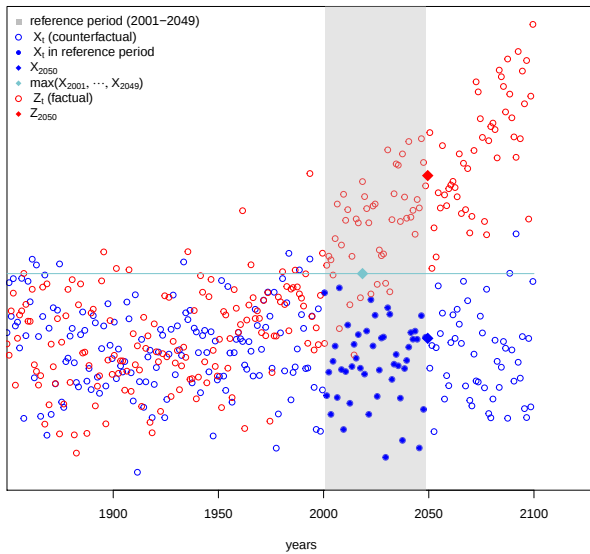
50-year record



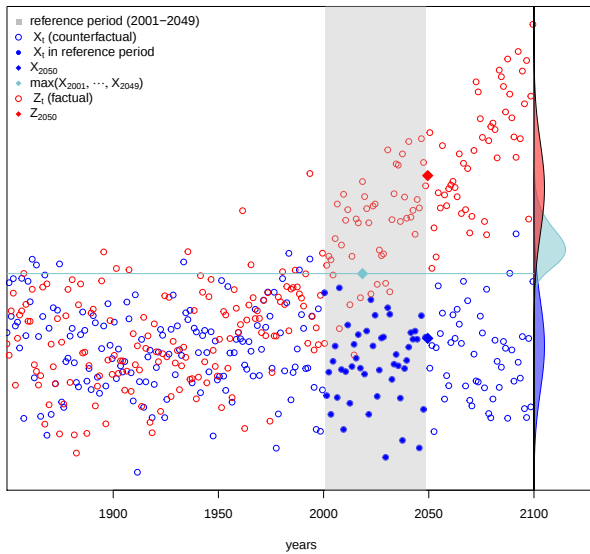
50-year record



50-year record



50-year record



Change of lenses in the causality framework

Instead of studying exceedance probabilities like

$$\mathbb{P}(X > u) \text{ and } \mathbb{P}(Z > u),$$

we work with record probabilities^a like

$$p_0(r) = \mathbb{P}(X > \max(X_1, \dots, X_{r-1})) \text{ \& } p_1(r) = \mathbb{P}(Z > \max(X_1, \dots, X_{r-1}))$$

a. the threshold u has been replaced by $\max(X_1, \dots, X_{r-1})$

Two probabilities of records

Counterfactual world

$$p_{0,r} = \mathbb{P}(X_r > \max(X_1, \dots, X_{r-1})) = \frac{1}{r}$$

Factual world

$$p_{1,r} = \mathbb{P}(Z_r > \max(X_1, \dots, X_{r-1}))$$

A new FAR, PN

$$PN(r) = far(r) = 1 - \frac{p_{0,r}}{p_{1,r}}$$

Two probabilities of records

Counterfactual world

$$p_{0,r} = \mathbb{P}(X_r > \max(X_1, \dots, X_{r-1})) = \frac{1}{r}$$

Factual world

$$p_{1,r} = \mathbb{P}(Z_r > \max(X_1, \dots, X_{r-1}))$$

A new FAR, PN

$$PN(r) = far(r) = 1 - \frac{p_{0,r}}{p_{1,r}}$$

A key variable $W = -\log G(Z)$

$$p_{1,r} = \mathbb{P}(Z_r > \max(X_1, \dots, X_{r-1})) = E(\exp((r-1)W))$$

Distribution of $W = -\log G(Z)$ when G describes annual maxima and Z too ?

Extreme Value Theory

$$\lim_{n \rightarrow \infty} \Pr(M_n \leq a_n x + b_n) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}$$



Gumbel (1891-1966)



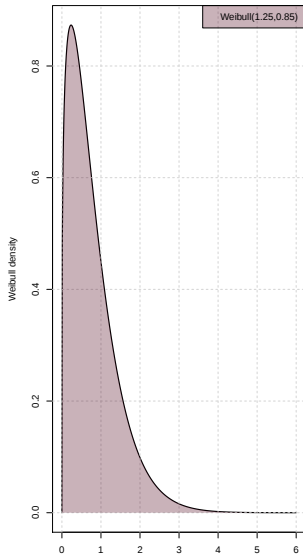
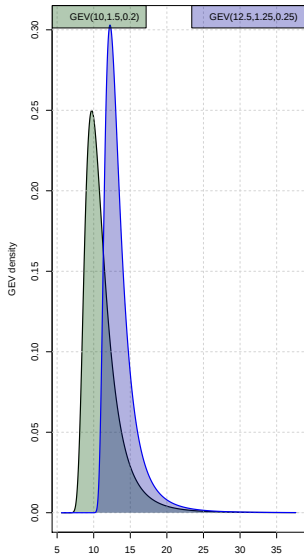
Weibull (1887-1979)



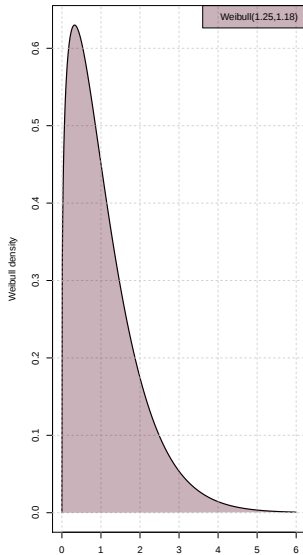
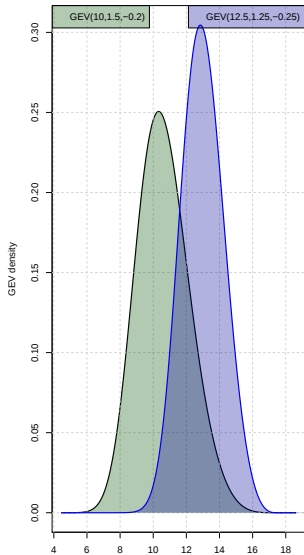
Fréchet (1878-1973)

see, e.g. *Statistics of Extremes*, Davison and Huser *Annual Review of Statistics and Its Application* 2015 2 :1, 203-235

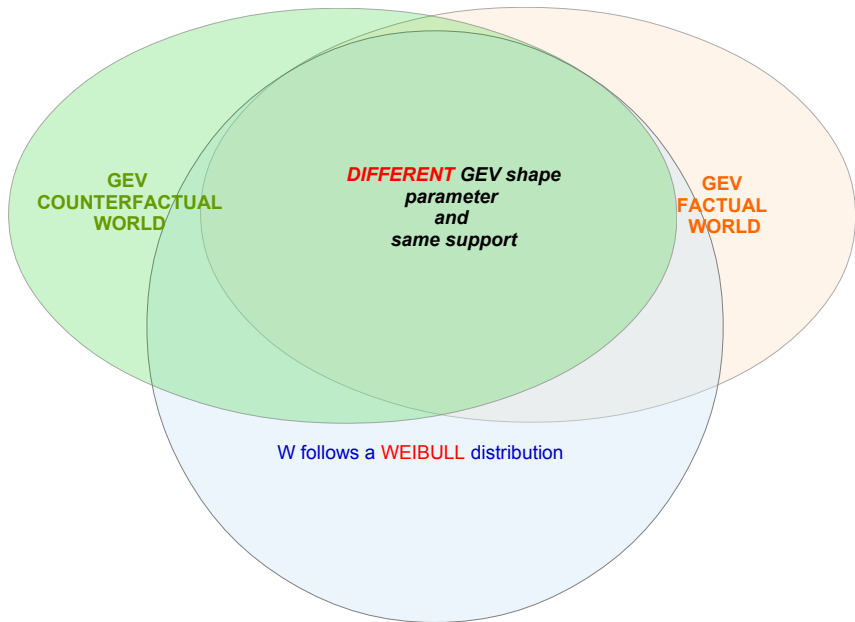
$W = -\log G(Z)$ with different GEV shape parameters



$W = -\log G(Z)$ with different GEV shape parameters



$W = -\log G(Z)$ with different GEV shape parameters



$W = -\log G(Z)$ with different GEV shape parameters³

If $X = \text{GEV}(\mu_X, \sigma_X, \xi_X)$ and $Z = \text{GEV}(\mu_Z, \sigma_Z, \xi_Z)$,
then

$W \sim \text{Weibull}(k, \lambda)$ with $k = \xi_X/\xi_Z$ and $\lambda = (k \times \sigma_Z/\sigma_X)^{-1/\xi_X}$,

and

$$p_{1,r} = \frac{k\Gamma(k)}{\lambda^k(r-1)^k} - \frac{k\Gamma(2k)}{\lambda^{2k}(r-1)^{2k}} + O(r^{-3k}),$$

and consequently,

$$\lim_{r \rightarrow \infty} \text{far}(r) = \begin{cases} 1 & \text{if } k < 1, \\ 1 - \lambda & \text{if } k = 1, \\ -\infty & \text{if } k > 1. \end{cases}$$

3. Note : $k = \xi_X/\xi_Z$ and $\lambda = (k \times \sigma_Z/\sigma_X)^{-1/\xi_X}$, see Worms and Naveau (2022), Record events attribution in climate studies, HAL link

Estimation algorithm of $p_{1,r}(t) = \mathbb{P}(Z_t > \max(X_{t-1}, \dots, X_{t-r+1})) = E(\exp((r-1)W_t))$

Under Weibull assumption : $p_{1,r}(t) = \int_0^1 \exp(-(r-1)\lambda_t(-\log x)^{1/k_t}) dx$

Estimation algorithm of $p_{1,r}(t) = \mathbb{P}(Z_t > \max(X_{t-1}, \dots, X_{t-r+1})) = E(\exp((r-1)W_t))$

Under Weibull assumption : $p_{1,r}(t) = \int_0^1 \exp(-(r-1)\lambda_t(-\log x)^{1/\hat{k}_t}) dx$

Estimation algorithm :

Step 1⁶ : $\forall t$

$$\widehat{p}_{1,2}(t) = \sum_{j=1}^n \frac{K_h(t-t_k)}{\sum_{k=1}^n K_h(t-t_k)} \mathbb{G}(Z_{t_j}) \quad \widehat{p}_{1,3}(t) = \sum_{j=1}^n \frac{K_h(t-t_k)}{\sum_{k=1}^n K_h(t-t_k)} \mathbb{G}^2(Z_{t_j})$$

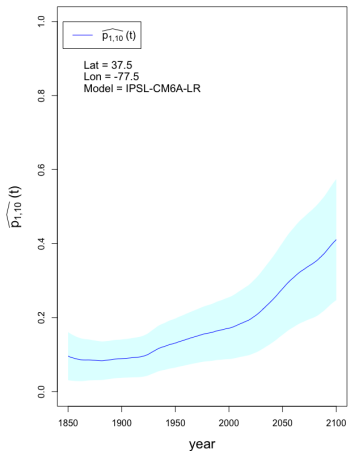
Step 2 : for a chosen t

$$\begin{cases} \widehat{p}_{1,2}(t) = \int_0^1 \exp(-\hat{\lambda}_t(-\log x)^{1/\hat{k}_t}) dx \\ \widehat{p}_{1,3}(t) = \int_0^1 \exp(-2\hat{\lambda}_t(-\log x)^{1/\hat{k}_t}) dx. \end{cases} \rightsquigarrow (\hat{k}_t, \hat{\lambda}_t)$$

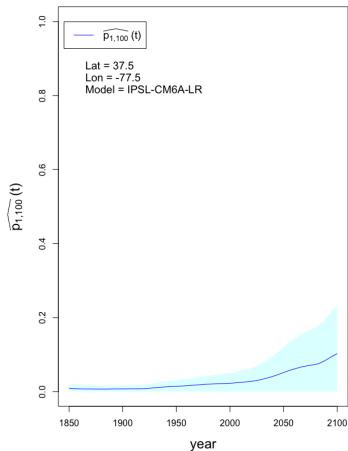
Step 3 : $\widehat{p}_{1,r}(t) \leftarrow \int_0^1 \exp(-(r-1)\hat{\lambda}_t(-\log x)^{1/\hat{k}_t}) dx$

Yearly maxima of daily precipitation at Richmond grid-point (rcp8.5)

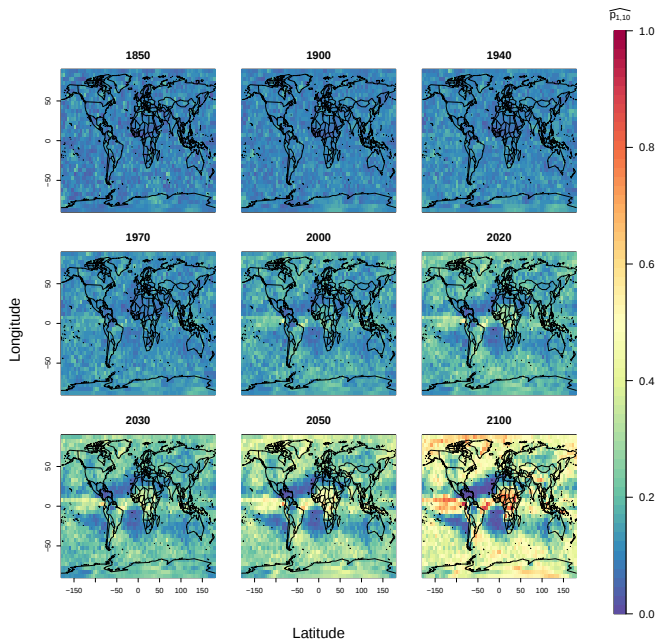
Decadal record probability



Centennial record probability



Decadal records of yearly maxima of daily precipitation

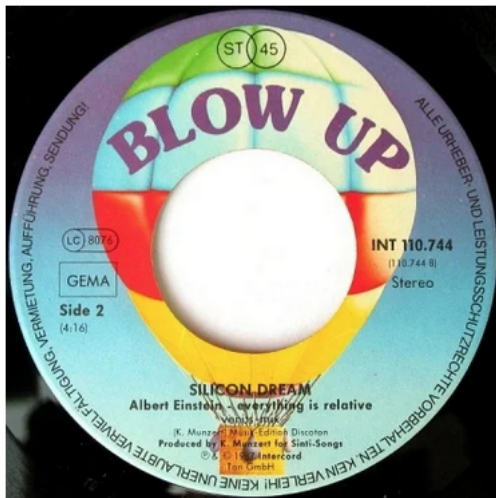


What's about multi-model error ?

Global climate models available for yearly maxima of daily precipitation

CMIP6 climate models		
Institute	Name	Runs
CCCma	CanESM5	r10i1p1f1
CNRM-CERFACS	CNRM-CM6-1	r1i1p1f2
IPSL	IPSL-CM6A-LR	r1i1p1f1
MRI	MRI-ESM2-0	r1i1p1f1
CMIP5 climate models		
Institute	Name	Runs
CCCma	CanESM2	r1i1p1
CNRM-CERFACS	CNRM-CM5	r1i1p1
CSIRO-BOM	ACCESS1-3	r1i1p1
CSIRO-QCCCE	CSIRO-Mk3-6-0	r1i1p1
IPSL	IPSL-CM5A-LR	r1i1p1
IPSL	IPSL-CM5A-MR	r1i1p1
MIROC	MIROC-ESM	r1i1p1
MIROC	MIROC-ESM-CHEM	r1i1p1
MRI	MRI-CGCM3	r1i1p1
NCAR	CCSM4	r1i1p1
NCC	NorESM1-M	r1i1p1
NSF-DOE-NCAR	CESM1-CAM5	r1i1p1

Back to our main message : everything is relative



Multi-model errors

Idea : $\mathbb{P}(X_t > Z_t) = \frac{1}{2}$ if X_t and Z_t exchangeable

Two types of IDEAL world

- Counterfactual

Sample **X** with $G(x) = \mathbb{P}(X \leq x)$

- Factual

Sample **Z** with $F(z) = \mathbb{P}(Z \leq z)$

Two types of IDEAL world

- Counterfactual

Sample **X** with $G(x) = \mathbb{P}(X \leq x)$

- Factual

Sample **Z** with $F(z) = \mathbb{P}(Z \leq z)$

We never observe **X** for the counterfactual climate and only partially **Z**

Two types of IDEAL world

- Counterfactual

Sample $\mathbf{X}$ with $G(x) = \mathbb{P}(X \leq x)$

- Factual

Sample $\mathbf{Z}$ with $F(z) = \mathbb{P}(Z \leq z)$

Two types of IMPERFECT numerical climate runs

- Counterfactual world from model m

Sample $\mathbf{X}^{(m)}$ with $G^{(m)}(x) = \mathbb{P}(X^{(m)} \leq x)$

- Factual world from model m

Sample $\mathbf{Z}^{(m)}$ with $F^{(m)}(z) = \mathbb{P}(Z^{(m)} \leq z)$

Linking factual and counterfactual truth with their numerical avatars

and

$$X \stackrel{d}{=} \left(G^{\leftarrow} \circ G^{(m)} \right) \left(X^{(m)} \right)$$

$$Z \stackrel{d}{=} \left(F^{\leftarrow} \circ F^{(m)} \right) \left(Z^{(m)} \right)$$

Linking factual and counterfactual truth with their numerical avatars

and

$$X \stackrel{d}{=} \left(G^{\leftarrow} \circ G^{(m)} \right) \left(X^{(m)} \right)$$

$$Z \stackrel{d}{=} \left(F^{\leftarrow} \circ F^{(m)} \right) \left(Z^{(m)} \right)$$

A fundamental condition in D&A

$$(A) : F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$$

A fundamental condition in D&A

$$(A) : F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$$

Climatological interpretation

The discrepancy between numerical model m and the true world stays the same in the factual and counterfactual worlds.

A fundamental condition in D&A

$$(A) : F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$$

Mathematical consequence of (A)

Under (A), it is possible to easily make **relative** comparisons of probabilities

Under exchangeability between X and X'

$$p_0 = \frac{1}{2} \text{ and } p_1 = \mathbb{P}(Z > X)$$

Under exchangeability between X and X'

$$p_0 = \frac{1}{2} \text{ and } p_1 = \mathbb{P}(Z > X)$$

If Model m satisfies condition (A), then

$$p_1 = p_1^{(m)} \text{ with } p_1^{(m)} = \mathbb{P}(Z^{(m)} > X^{(m)})$$

Under exchangeability between X and X'

$$p_0 = \frac{1}{2} \text{ and } p_1 = \mathbb{P}(Z > X)$$

If Model m satisfies condition (A), then

$$p_1 = p_1^{(m)} \text{ with } p_1^{(m)} = \mathbb{P}(Z^{(m)} > X^{(m)})$$

Under (A), there is no need to correct Model m . Huge gain !!

Inference of $\mathbb{P}(Z^{(m)} > X^{(m)})$

CLIMATE MODEL RUNS

COUNTERFACTUAL

1850-2014

2015-2100

FACTUAL

1850-2014

2015-2100

Temporal indexing to deal with transient runs

$$p_t^{(m)} = P\left(Z_t^{(m)} > X_t^{(m)}\right)$$

Temporal indexing to deal with transient runs

$$\hat{p}_t^{(m)} = P\left(Z_t^{(m)} > X_t^{(m)}\right)$$

Kernel Nadaraya-Watson regression

$$\hat{p}_t^{(m)} = \frac{1}{\sum K(t - t_j)} \sum_{j=1}^J K(t - t_j) \mathbb{G}_I^{(m)}\left(Z_{t_j}^{(m)}\right),$$

where K weighting Kernel and $\mathbb{G}^{(m)}$ empirical cdf from $X_{1:t}^{(m)}$

Temporal indexing to deal with transient runs

$$\hat{p}_t^{(m)} = P\left(Z_t^{(m)} > X_t^{(m)}\right)$$

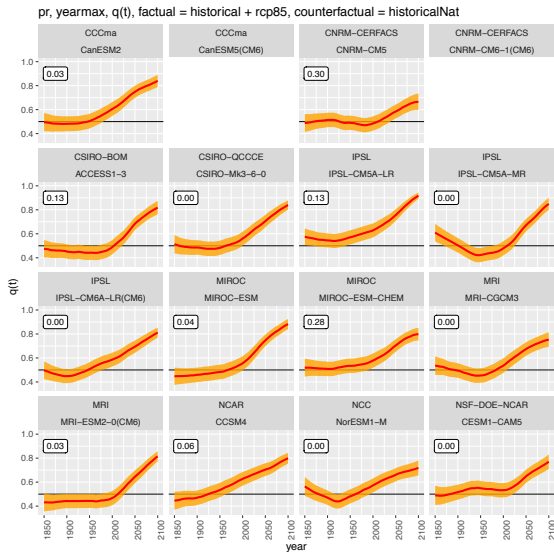
Kernel Nadaraya-Watson regression

$$\hat{p}_t^{(m)} = \frac{1}{\sum K(t - t_j)} \sum_{j=1}^J K(t - t_j) \mathbb{G}_I^{(m)}\left(Z_{t_j}^{(m)}\right),$$

where K weighting Kernel and $\mathbb{G}^{(m)}$ empirical cdf from $X_{1:t}^{(m)}$

Asymptotic confidence intervals build by coupling Kernel results with empirical process results, see e.g. Mason and Van Zwet (1987), provide asymptotic confidence intervals when $\hat{G}^{(m)}$ is replaced by $\mathbb{G}^{(m)}$

$\hat{p}_t^{(m)}$ for yearly maxima of rainfall at Oxford



How to judge and combine all the $\hat{q}_t^{(m)}$?

TRUE WORLDS

COUNTERFACTUAL

1850-1900

1901-2014

2015-2100

FACTUAL

1850-1900

1901-2014

2015-2100

TRUE WORLDS

	COUNTERFACTUAL	
1850-1900	1901-2014	2015-2100
	FACTUAL	
1850-1900	1901-2014	2015-2100
COUNTERFACTUAL IS EQUAL TO FACTUAL		

Notations reminder

$$X \stackrel{d}{=} \left(G^{\leftarrow} \circ G^{(m)} \right) \left(X^{(m)} \right) \text{ and } Z \stackrel{d}{=} \left(F^{\leftarrow} \circ F^{(m)} \right) \left(Z^{(m)} \right)$$

$$(A) : F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$$

Notations reminder

$$X \stackrel{d}{=} \left(G^{\leftarrow} \circ G^{(m)} \right) \left(X^{(m)} \right) \text{ and } Z \stackrel{d}{=} \left(F^{\leftarrow} \circ F^{(m)} \right) \left(Z^{(m)} \right)$$

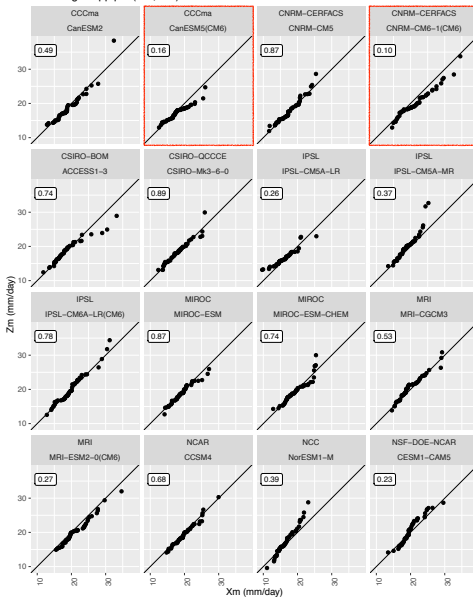
$$(A) : F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$$

During the pre-industrial period $\mathcal{T} = \{1850, \dots, 1900\}$

- if $F_t = G_t$ and $F_t^{(m)} = G_t^{(m)}$ for $t \in \mathcal{T}$, then (A) holds for $t \in \mathcal{T}$

Pre-industrial yearly maxima of rainfall at Oxford

Checking A: qq-plot(X_m , Z_m)



A convex combination

$$\hat{q}_t = \sum_{m=1}^M w_m \times \hat{p}_t^{(m)}, \text{ with } w_i \geq 0 \text{ and } w_1 + \dots + w_M = 1.$$

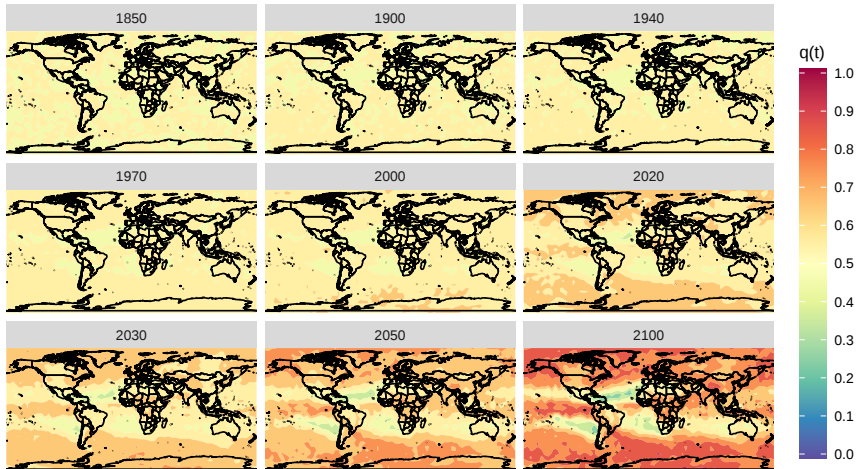
with weights minimizing with Kullback Leibler divergence w.r.t. to $q_{t \in \mathcal{T}} = c(.5, \dots, .5)$

Yearly maxima of rainfall at Oxford grid point



GCM with largest Anderson-Darling p-value Weighted convex combination

pr, yearmax, q(t) multimodel



Emergence times at yearly maxima of rainfall

A clear referential

$$p_t = P(Z_t > X_t) = \frac{1}{2}, \text{ for all year } t \in \mathcal{T}$$

Emergence times at yearly maxima of rainfall

A clear referential

$$p_t = P(Z_t > X_t) = \frac{1}{2}, \text{ for all year } t \in \mathcal{T}$$

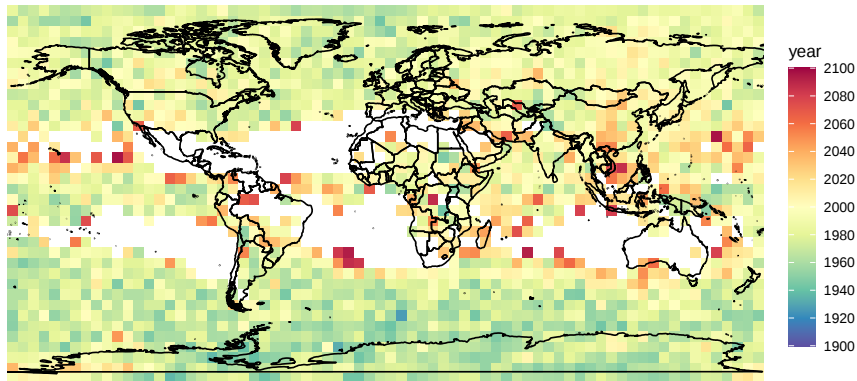
Emergence times

$$\tau_p = \operatorname{argmin}\{t \text{ for all } t' \geq t, \text{ we have } P(\hat{p}_{t'} > 0.5) > p\}.$$

All years after the emergence time τ_p have a q_t significantly higher than 0.5 (at level p)

Emergence times at yearly maxima of rainfall

pr, yearmax, multimodel time of emergence



Conclusions

- D&A deals with relative questions
- Simple and fast algorithms to combine climate model outputs. No need to implement classical “bias” correction approaches
- Emergence of anthropogenic forcing appears around the year 1990 for yearly maxima of precipitation in CMIP world
- Move from six GEV parameters to two Weibull parameters
- Record changes can be analyzed quickly

Conclusions

- D&A deals with relative questions
- Simple and fast algorithms to combine climate model outputs. No need to implement classical “bias” correction approaches
- Emergence of anthropogenic forcing appears around the year 1990 for yearly maxima of precipitation in CMIP world
- Move from six GEV parameters to two Weibull parameters
- Record changes can be analyzed quickly

Open questions

- How to introduce observations in the setup ?
- How to move from grid-point to regions with records ?

Collaborators list and related articles [2, 1, 3, 4, 5, 6, 7, 8, 9, 10]



J. Worms and P. Naveau.

Record events attribution in climate studies.
Environmetrics, 2022.



P. Naveau and S. Thao.

Multi-model errors and emergence times in climate attribution studies.
Journal of Climate, pages 4791–4804, 2022.



J. Rohmer, J. Louisor, G. Le Cozannet, P. Naveau, S. Thao, and X. Bertin.

Attribution of Extreme Wave Height Records along the North Atlantic Coasts using Hindcast Data : Feasibility and Limitations.
Journal of Coastal Research, 95(sp1) :1268 – 1272, 2020.



P. Naveau, A. Hannart, and A. Ribes.

Statistical methods for extreme event attribution in climate science.
Annual Review of Statistics and Its Application, 7(1) :89–110, 2020.



A. Kiriliouk and P. Naveau.

Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory.
Annals of Applied Statistics, 14(3) :1342–1358, 2020.



P. Naveau, A. Ribes, F. W. Zwiers, A. Hannart, A. Tuel, and P. Yiou.

Revising return periods for record events in a climate event attribution context.
Journal of Climate, 31 :3411–3422, 2018.



A. Hannart and P. Naveau.

Probabilities of causation of climate changes.
Journal of Climate, 31 :5507–5524, 2018.



P. Yiou, A. Jezequel, P. Naveau, F. E. L. Otto, R. Vautard, and M. Vrac.

A statistical framework for conditional extreme event attribution.
Advances in Statistical Climatology, Meteorology and Oceanography, 3 :17–31, 2017.



A. Ribes, F. W. Zwiers, J. M. Azaïs, and P. Naveau.

A new statistical approach to climate change detection and attribution.
Climate Dynamics, pages 29–56, 2017.



A. Hannart, J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil.

Counterfactual causality theory for the attribution of weather and climate-related events

Weights of yearly maxima of rainfall

