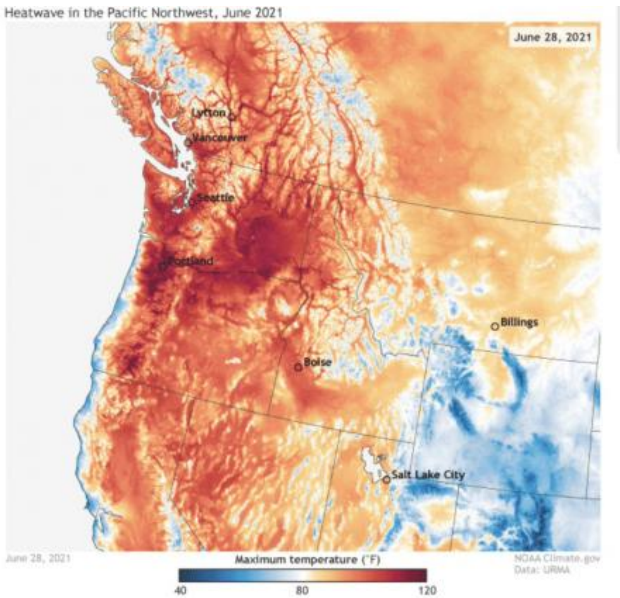
A few key elements to understand extreme event attribution studies.

Soulivanh Thao (LSCE),

Philippe Naveau (LSCE), Paula Gonzalez (LSCE), Julien Worms (UVSQ),

Aurelien Ribes (CNRM), Julien Cattiaux (CNRM)

World Weather Attribution: Pacific Northwest 2021 Heatwave



"extremely rare (a 1 in 1000 year event) in today's climate"

"virtually impossible without human-caused climate change."

"at least 150 times rarer without human-induced climate change."

"2°C hotter than it would have been [...] at the beginning of the industrial revolution (when global mean temperatures were 1.2°C cooler than today)."

Attribution of climate change

Source:
IPCC AR5, Chapter 10.

Attribution of climate change

Attribution

> *The process of evaluating the relative contributions of multiple* ***causal factors*** *to change or event with an assignment of* ***statistical*** *confidence.*

Source:
IPCC AR5, Chapter 10.

Attribution of climate change

Attribution

> *The process of evaluating the relative contributions of multiple **causal factors** to change or event with an assignment of **statistical** confidence.*

Consequences

Need to assess whether the observed changes are

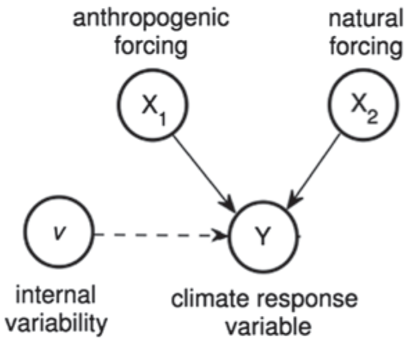
- consistent with the expected responses to external forcings
- inconsistent with alternative explanations

Source:
IPCC AR5, Chapter 10.

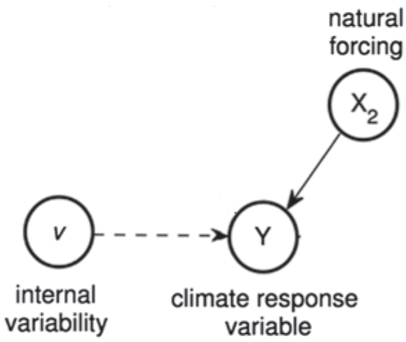
Attribution and causality

Assessing the causal effect of human activity through the notion of intervention and counterfactuals

factual world



counterfactual world



Source:
Hannart, A., J. Pearl, F.E. Otto, P. Naveau, and M. Ghil, 2016: Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events. *Bull. Amer. Meteor. Soc.*, 97, 99–110, https://doi.org/10.1175/BAMS-D-14-00034.1

Importance of climate models in D&A

Counterfactual

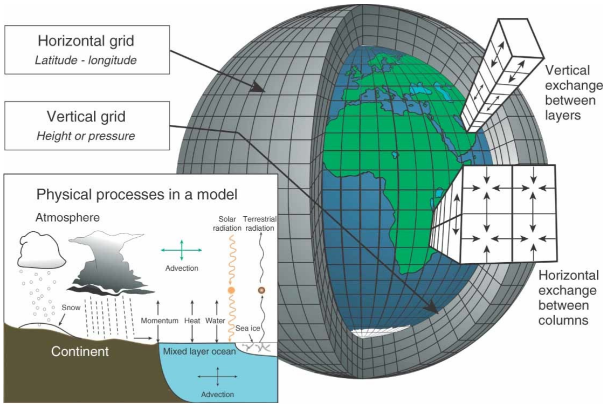
can never be observed in practice.

Intervention

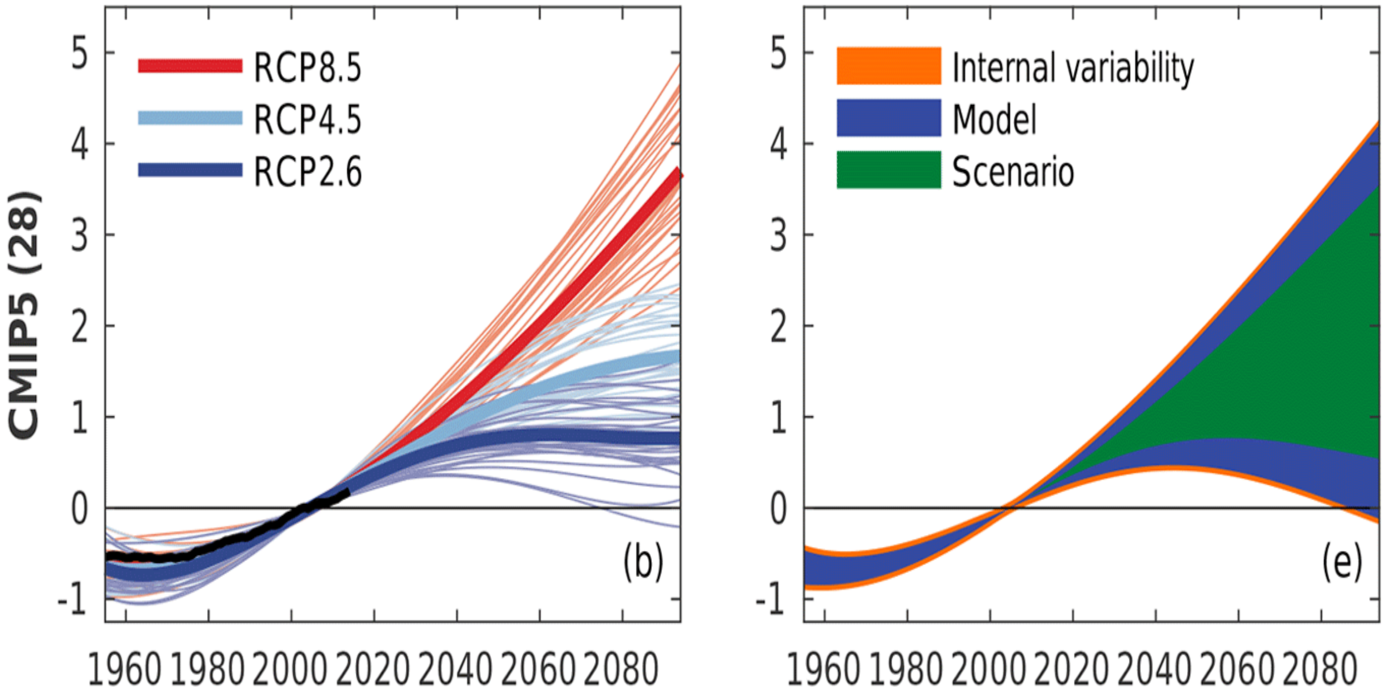
requires to have an experimental group and a control group, which is not possible with only one Earth.

Solution ?

Use climate models to perform controlled experiements.

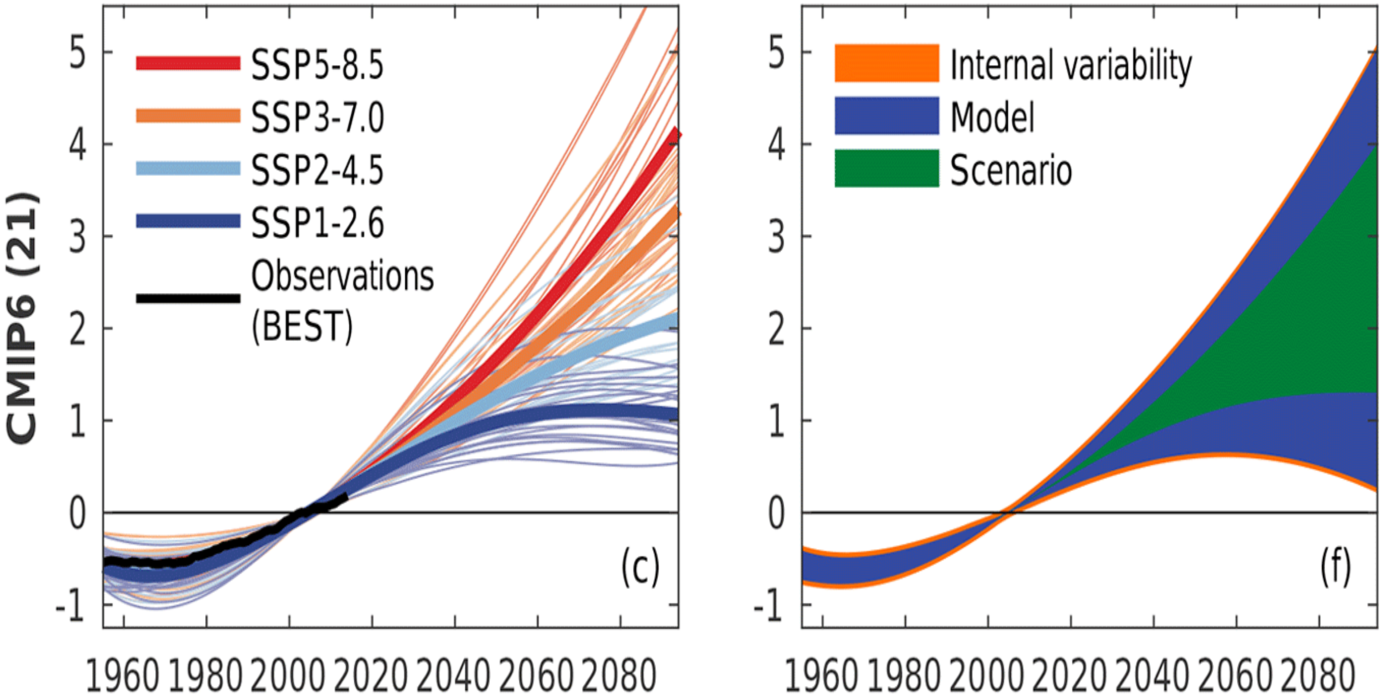


CMIP5 and temperature projections



Source:
Lehner et al. 2020. "Partitioning Climate Projection Uncertainty with Multiple Large Ensembles and CMIP5/6." Earth System Dynamics 11 (2): 491–508. https://doi.org/10.5194/esd-11-491-2020.

CMIP6 and temperature projections

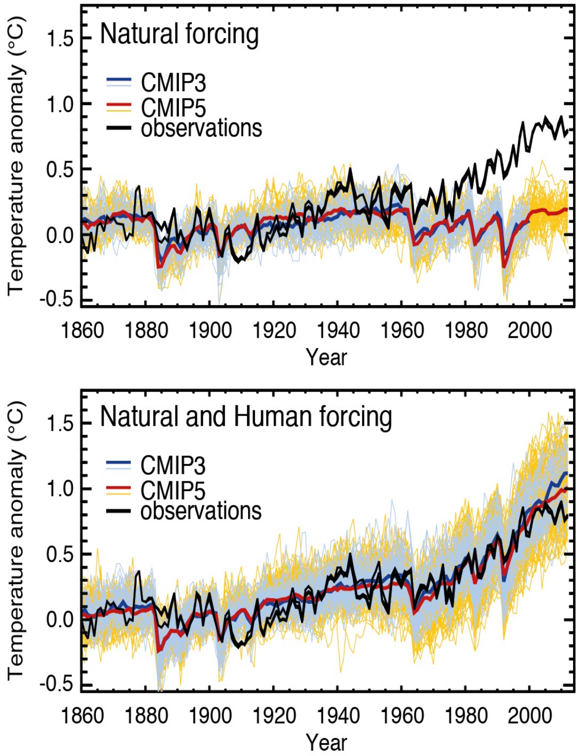


Source:
Lehner et al. 2020. “Partitioning Climate Projection Uncertainty with Multiple Large Ensembles and CMIP5/6.” Earth System Dynamics 11 (2): 491–508. https://doi.org/10.5194/esd-11-491-2020.

But we still need observations and statistical models

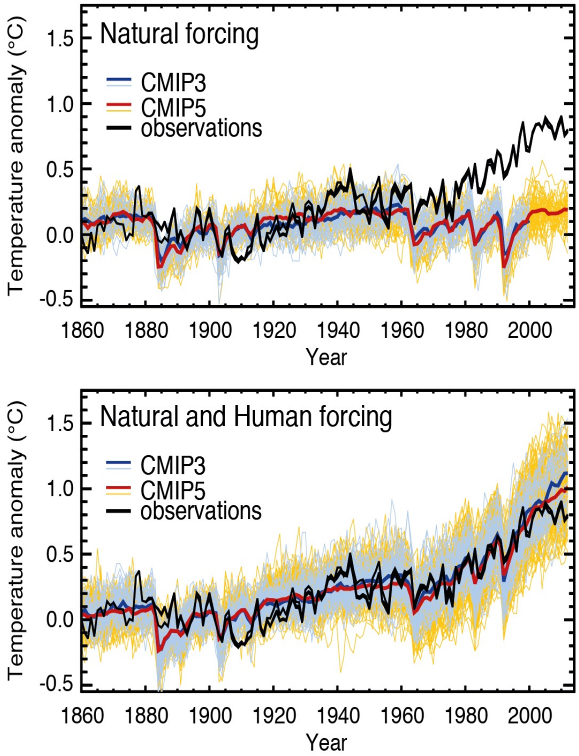
- To check wether climate models are consistents with observations
- To link climate models and observations
- To infer quantities of interest

Attribution of Global Mean Surface Temperature Trends



Source:
IPCC AR5, FAQ 10.1, Figure 1 |

Attribution of Global Mean Surface Temperature Trends



Statistical Model

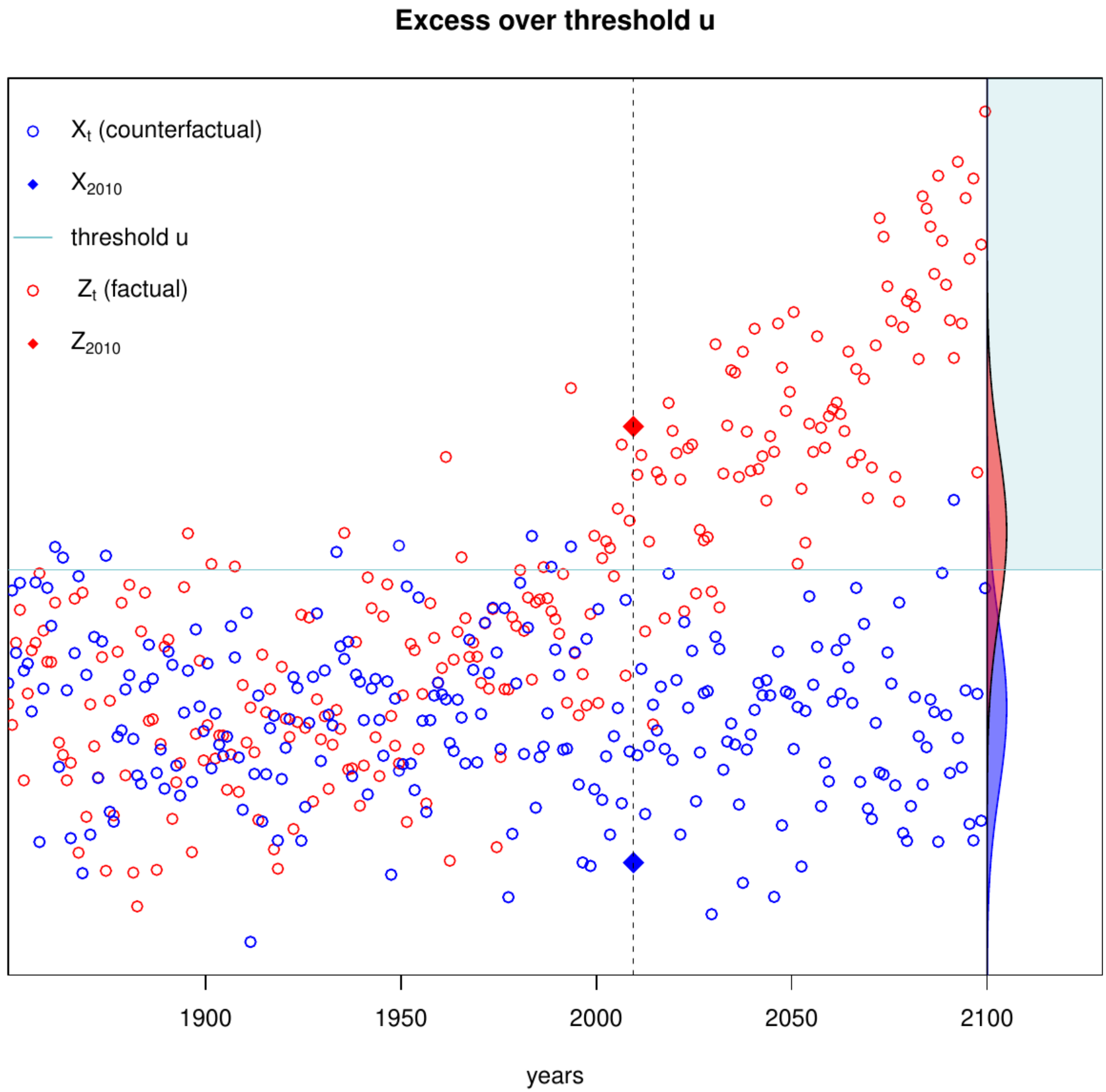
$$Z^* = \mu + \sum_{i=n}^{N} \beta_i R_i^*$$
$$Z = Z^* + \varepsilon_Z$$
$$R_i = R_i^* + \varepsilon_{R_i} \quad \text{for } i = 1, \dots, N.$$

Z is the climate state observed at a given time, which is therefore impacted by climate internal variability ε_Z.

R_i is a response to forcing i simulated by a climate model. It includes the contribution of the internal variability ε_{R_i} simulated by the climate model.

Source:
IPCC AR5, FAQ 10.1, Figure 1 |

Attribution of Extreme Events

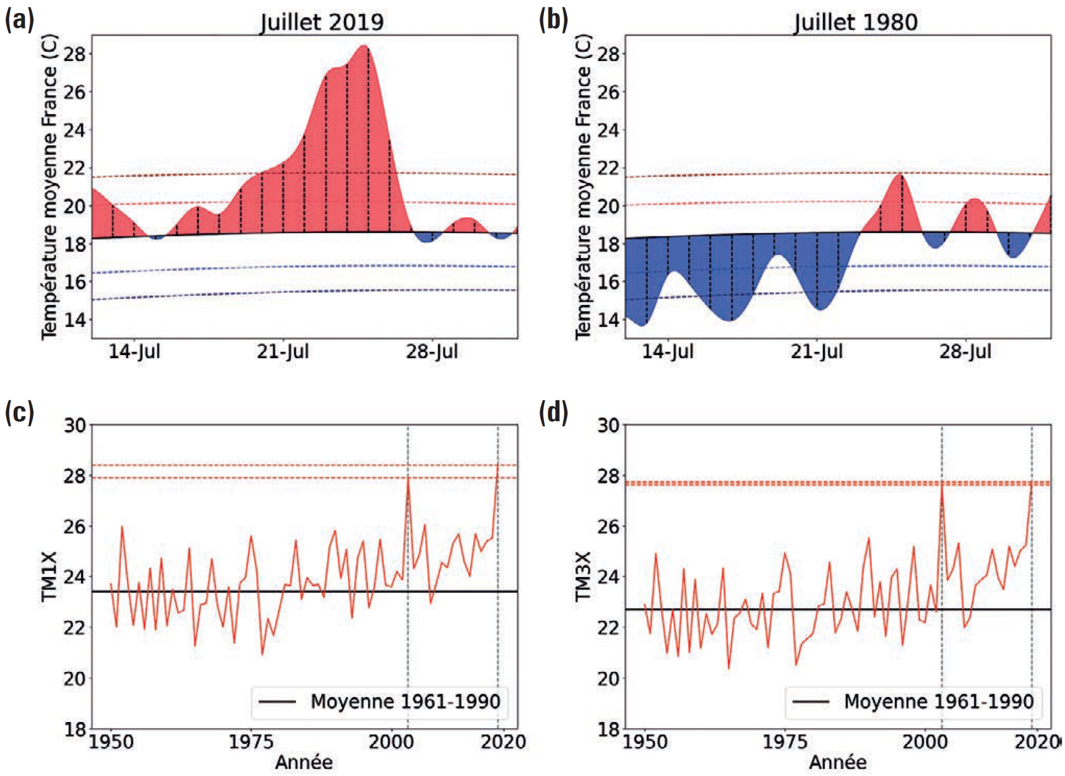


Defining the event (only univariate)

1. Define the climate variable of interest.
   e.g. temperature, precipitation, ...
2. Define the spatio-temporal aggregation.
   e.g. maximum over Europe of 10-day temperature averages
3. Define the conditioning.
   e.g. only during JJA, only for NAO+ circulation type, ...
4. Define the threshold
   e.g. based on recent event

More on this:
Cattiaux, J. and A. Ribes, 2018: Defining Single Extreme Weather Events in a Climate Perspective. *Bull. Amer. Meteor. Soc.*, 99, 1557–1568, https://doi.org/10.1175/BAMS-D-17-0281.1

Case study: 2019 July Heatwave in France



Event definition:

Annual maxima of 3-day average temperature anomalies in France, i.e. [42N - 51N] x [5W -10E] above +4.98 K.

Source:
Robin, Yoann ; Drouin, Agathe ; Soubeyroux, Jean-Michel ; Ribes, Aurélien ; Vautard, Robert. Comment attribuer une canicule au changement climatique ?. *La Météorologie*, 115, 28-36, 2021.

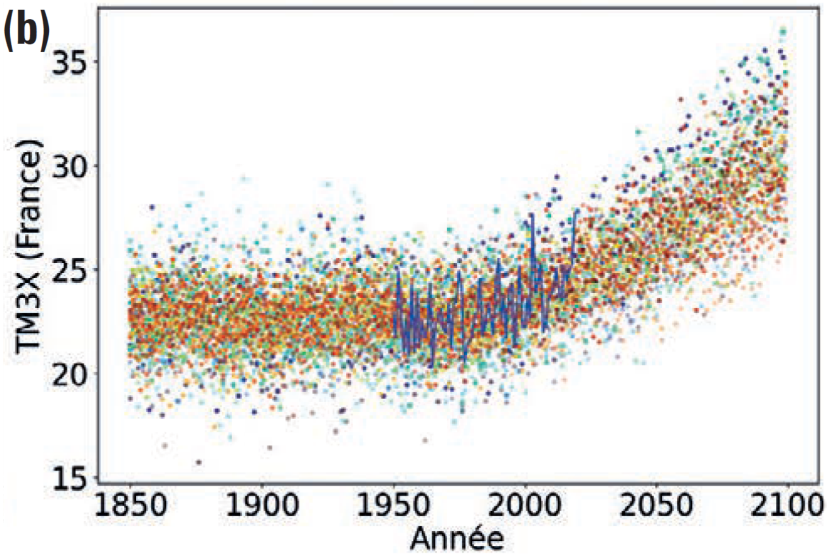
Data

Simulations from 26 models from the Coupled Model Intercomparison Project 5 (CMIP5).

Data are extracted for the experiences:

- **historical**, corresponding to the climate in the factual world (1850-2005)
- **rcp2.5, rcp8.5**, corresponding to climate projections for the future of the factual world(2006-2100)

Météo-France thermal index is used as observation. It corresponds to the average of observations from 30 ground stations, showing data available between 1947 and 2019



Generalized Extreme Value distribution

The model focuses on the statistical behavior of

$$M_n = max\{X_1, \ldots, X_n\},$$

where $X_l, \ldots, X_n$, is a sequence of independent random variables having a common distribution function F

Theorem
If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\mathbb{P}\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \text{ as } n \rightarrow \infty$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

Bayesian Model

Bayesian Model

Factual world

$$
\begin{aligned}
Z_t &\sim GEV(\mu_0 + \mu_1 C_t^F, exp(\sigma_0 + \sigma_1 C_t^F), \xi) \\
C_t^F &= C_t^{NAT} + C_t^{ANT} + \epsilon_{IV} \\
\epsilon_{IV} &\sim \mathcal{N}(0, \sigma_{IV}^2)
\end{aligned}
$$

Bayesian Model

Factual world

$$\begin{aligned} Z_t &\sim GEV(\mu_0 + \mu_1 C_t^F, exp(\sigma_0 + \sigma_1 C_t^F), \xi) \\ C_t^F &= C_t^{NAT} + C_t^{ANT} + \epsilon_{IV} \\ \epsilon_{IV} &\sim \mathcal{N}(0, \sigma_{IV}^2) \end{aligned}$$

Counterfactual world

$$\begin{aligned} X_t &\sim GEV(\mu_0 + \mu_1 C_t^{CF}, exp(\sigma_0 + \sigma_1 C_t^{CF}), \xi) \\ C_t^{CF} &= C_t^{NAT} + \epsilon_{IV} \\ \epsilon_{IV} &\sim \mathcal{N}(0, \sigma_{IV}^2) \end{aligned}$$

Bayesian Model

Factual world

$$\begin{aligned} Z_t &\sim GEV(\mu_0 + \mu_1 C_t^F, exp(\sigma_0 + \sigma_1 C_t^F), \xi) \\ C_t^F &= C_t^{NAT} + C_t^{ANT} + \epsilon_{IV} \\ \epsilon_{IV} &\sim \mathcal{N}(0, \sigma_{IV}^2) \end{aligned}$$

Counterfactual world

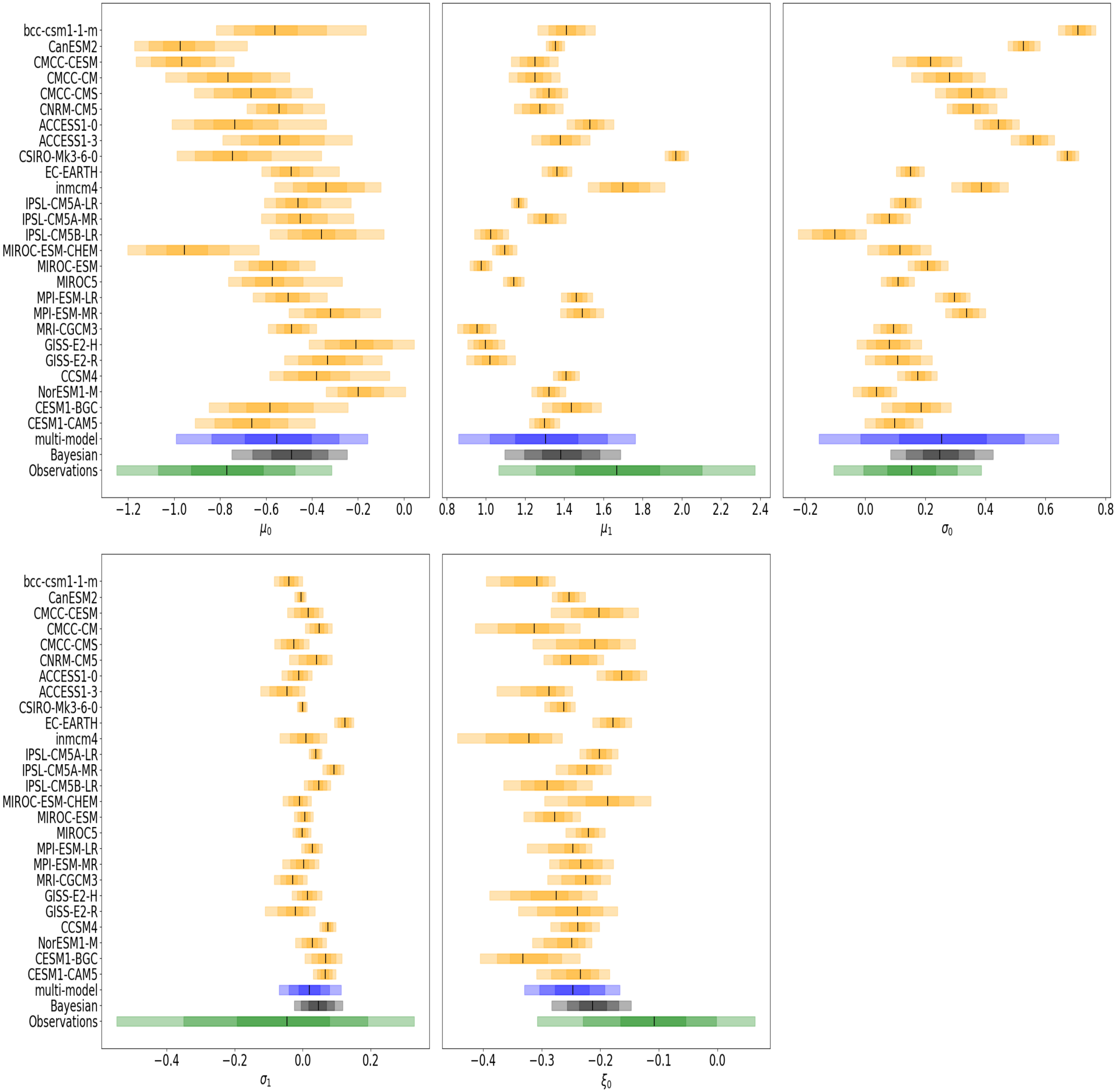
$$\begin{aligned} X_t &\sim GEV(\mu_0 + \mu_1 C_t^{CF}, exp(\sigma_0 + \sigma_1 C_t^{CF}), \xi) \\ C_t^{CF} &= C_t^{NAT} + \epsilon_{IV} \\ \epsilon_{IV} &\sim \mathcal{N}(0, \sigma_{IV}^2) \end{aligned}$$

Prior distribution

$$\begin{aligned} \theta &= [\mu_0, \mu_1, \sigma_0, \sigma_1, \xi, C^{NAT}, C^{ANT}] \\ &\sim \mathcal{N}(\mu_\theta, \sigma_\theta^2) \end{aligned}$$

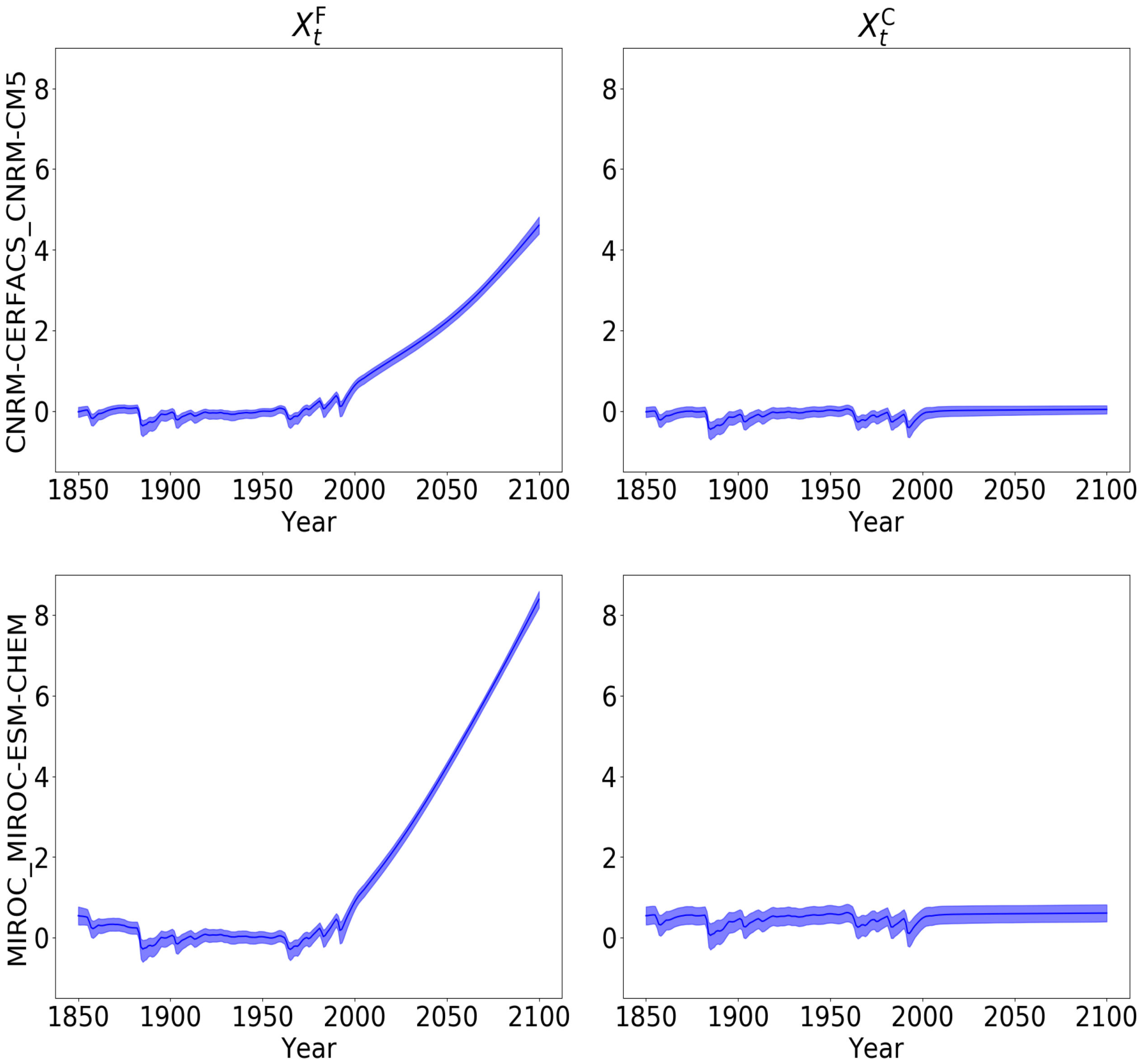
We use the distribution of the CMIP5 models as prior

GEV fits

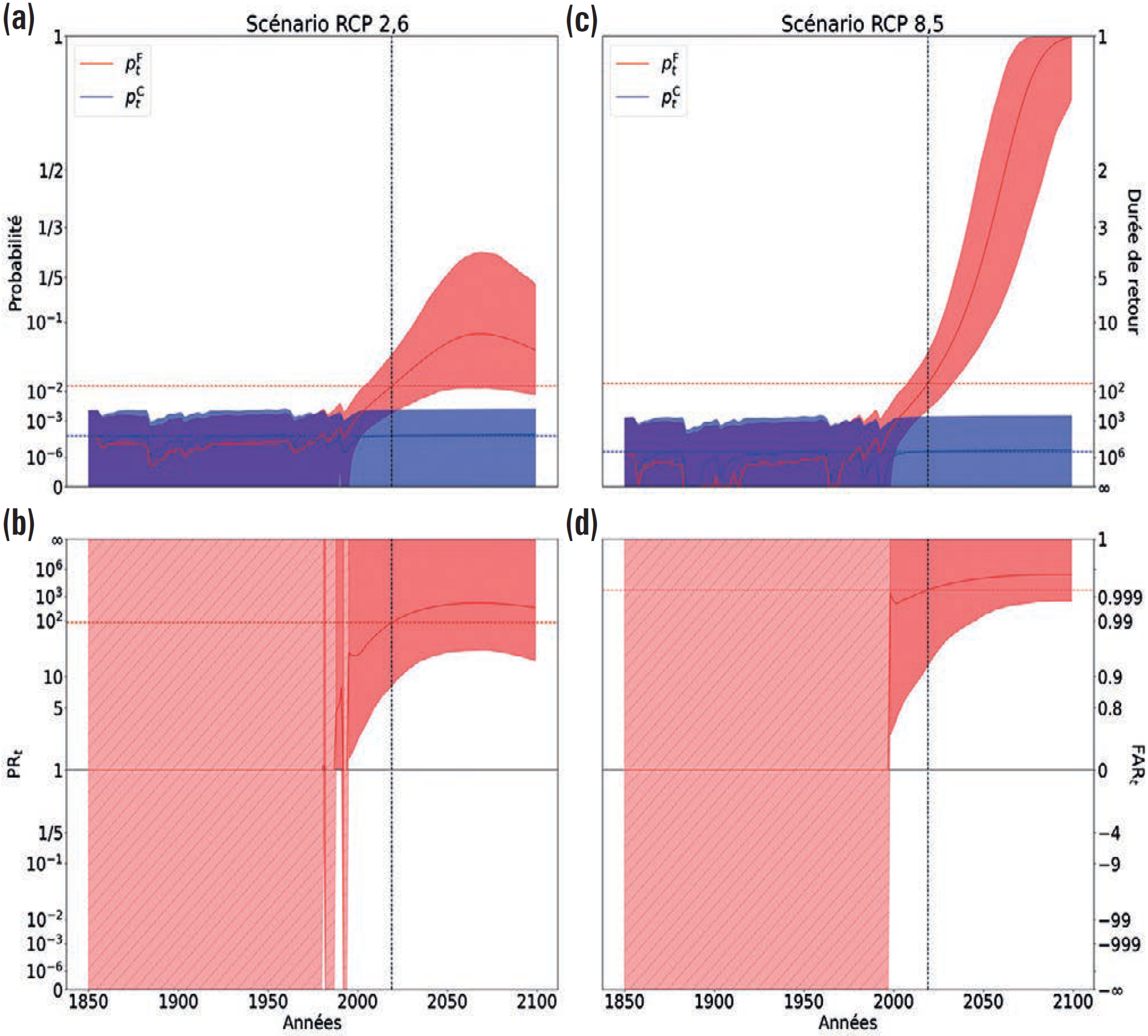


bcc-csm1-1-m
CanESM2
CMCC-CESM
CMCC-CM
CMCC-CMS
CNRM-CM5
ACCESS1-0
ACCESS1-3
CSIRO-Mk3-6-0
EC-EARTH
inmcm4
IPSL-CM5A-LR
IPSL-CM5A-MR
IPSL-CM5B-LR
MIROC-ESM-CHEM
MIROC-ESM
MIROC5
MPI-ESM-LR
MPI-ESM-MR
MRI-CGCM3
GISS-E2-H
GISS-E2-R
CCSM4
NorESM1-M
CESM1-BGC
CESM1-CAM5
multi-model
Bayesian
Observations
μ_0
μ_1
σ_0
σ_1
ξ_0

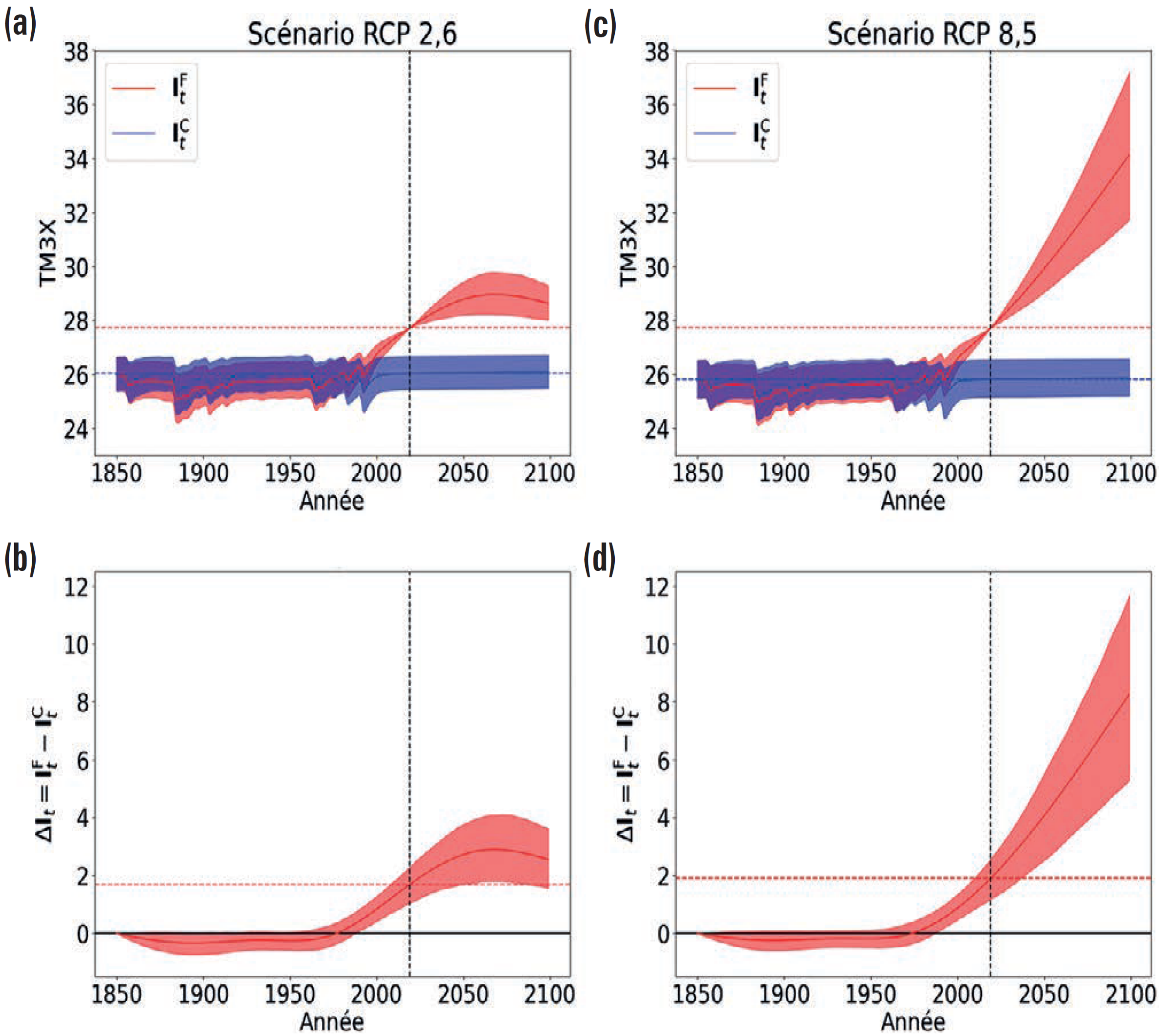
Global Mean Surface Temperature as covariate



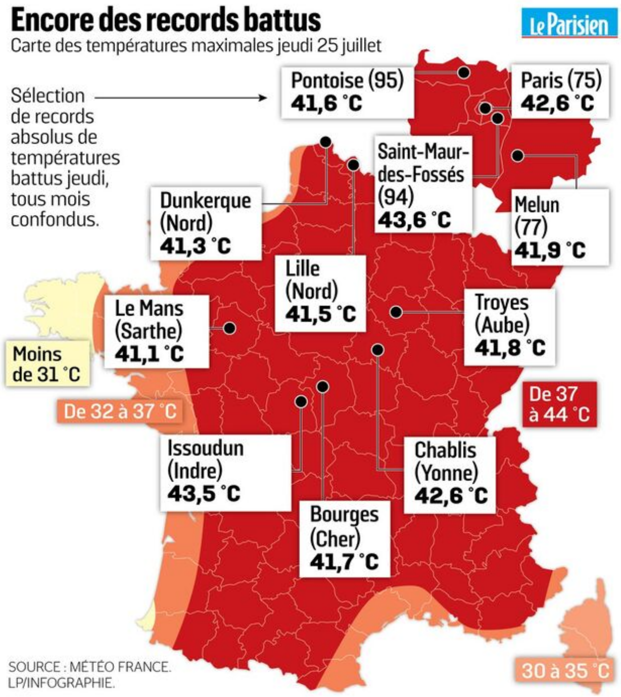
Posterior: change in probability



Posterior: change in intensity



Conclusions of the case study



"currently the risk of occurrence of this type of heat wave has been multiplied by at least 10"

"In 2040 it will be multiplied by at least 20".

"in the best case (rcp2.5) we return to the current situation in 2100"

Source:
Robin, Yoann ; Drouin, Agathe ; Soubeyroux, Jean-Michel ; Ribes, Aurélien ; Vautard, Robert. Comment attribuer une canicule au changement climatique ?. *La Météorologie*, 115, 28-36, 2021.

Let's take a few steps back

Event attribution is about comparing two distributions

- **Counterfactual**

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

- **Factual**

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

Hard to do with observations...

- We never observe $\mathbf{X}$ for the counterfactual climate
- Often lack of available data $\mathbf{Z}$ for the factual climate

... and models have biases...

- **Counterfactual** world from model m

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

- **Factual** world from model m

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

Statistical bias correction

Linking factual and counterfactual truth with their numerical approximations

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

and

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

where $G^{\leftarrow}(.)$ the inverse of G, i.e. the quantile function. All variables are assumed to be continuous.

A fundamental condition in D&A

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

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

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

Climatological interpretation

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

Mathematical consequence of A

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

Change of lenses: relative comparison of probabilities

Instead of studying exceedance probabilities like

$$p_0(t) = \mathbb{P}(X_t > u) \;\; \textbf{and} \;\; p_1(t) = \mathbb{P}(Z_t > u),$$

we will estimate and interpret record probabilities like

$$p_{0,r}(t) = \mathbb{P}(X_t > \max(X_{t-1}, X_{t-2}, \ldots, X_{t-r+1}))$$

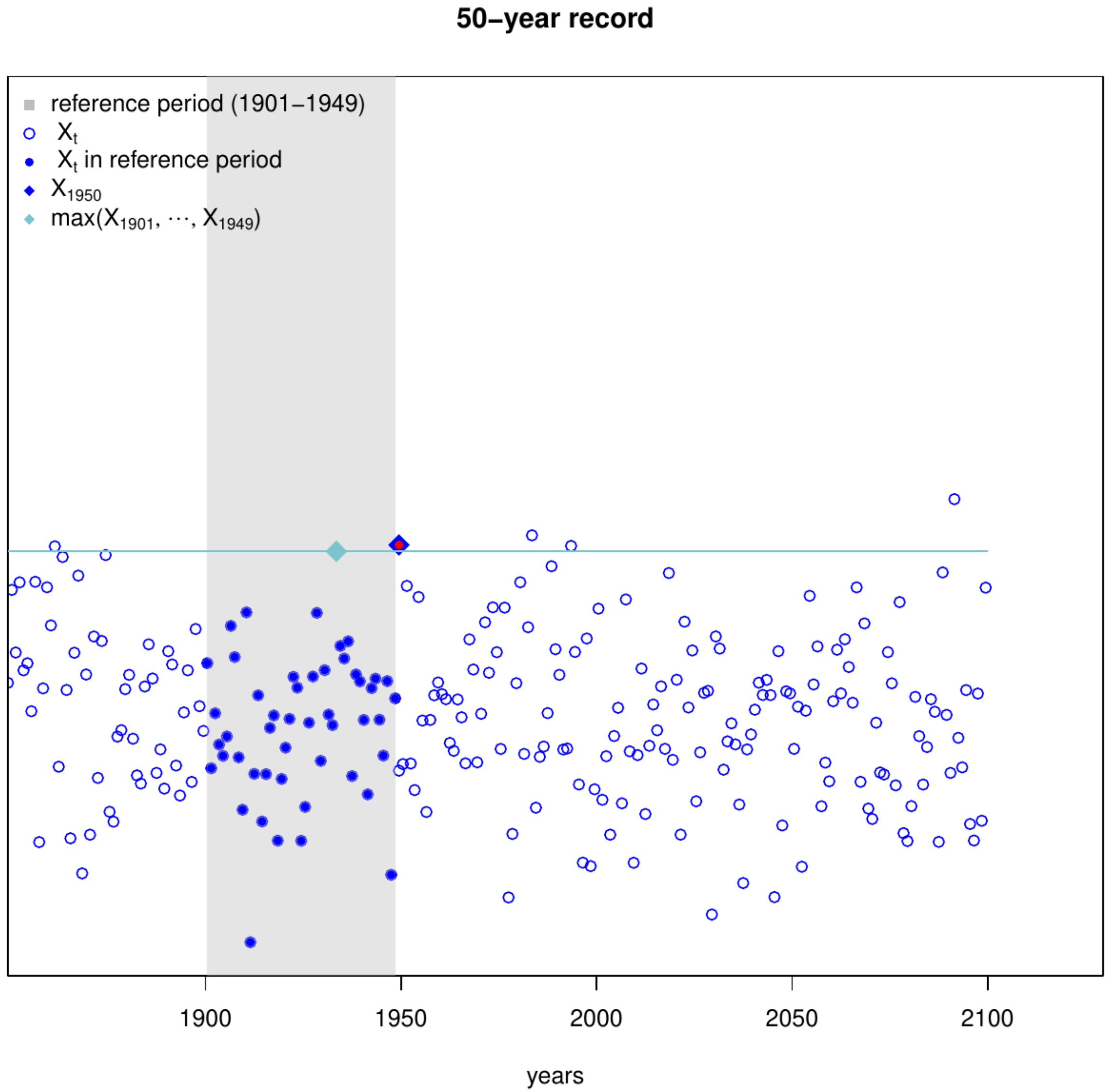
and

$$p_{1,r}(t) = \mathbb{P}(Z_t > \max(X_{t-1}, X_{t-2}, \ldots, X_{t-r+1}))$$

where the usual threshold u has been replaced by $\max(X_{t-1}, X_{t-2}, \ldots, X_{t-r+1})$.

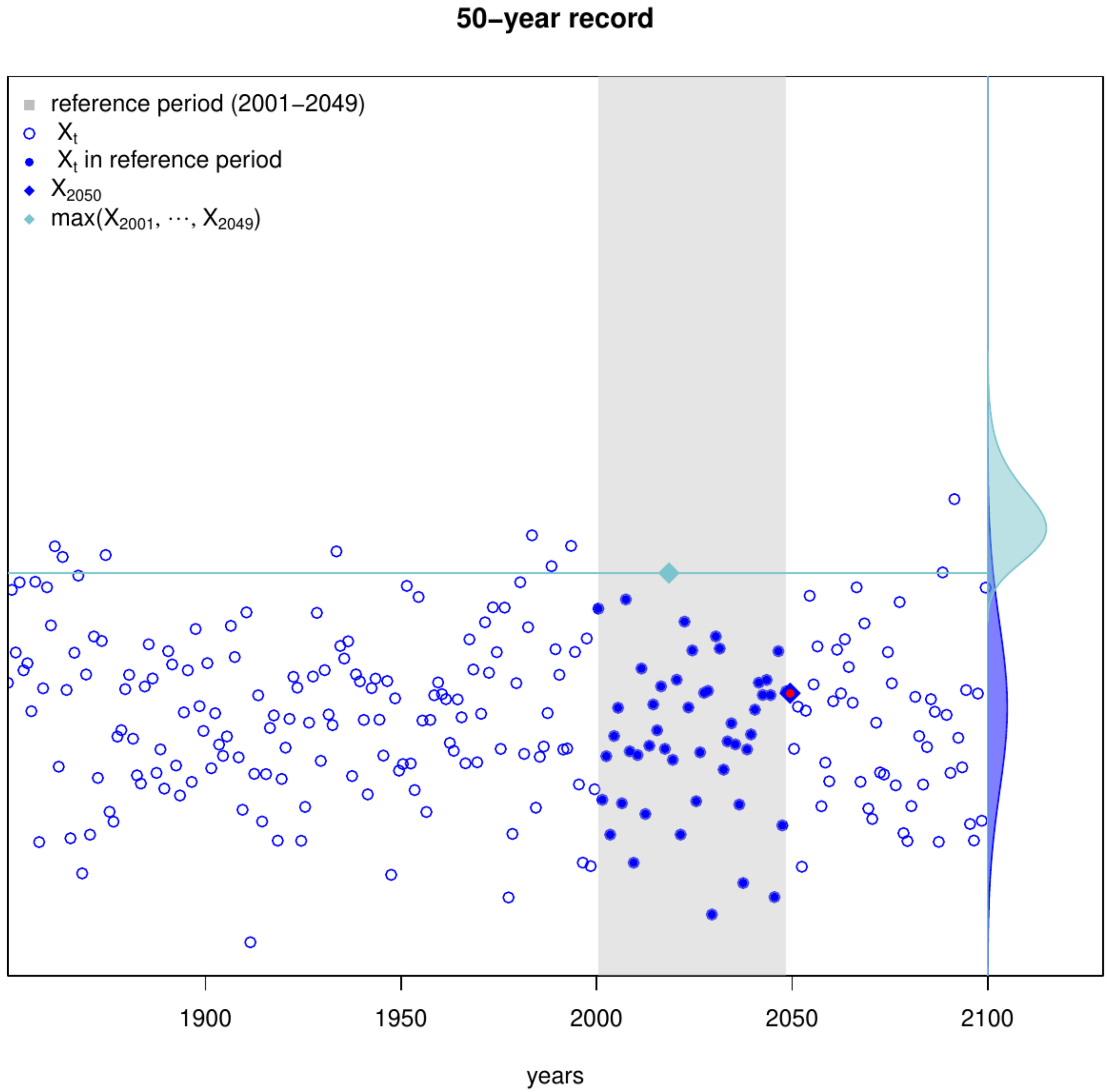
Records in the counterfactual world:

$$p_{0,r}(t) = \mathbb{P}(X_t > \max(X_{t-1}, \ldots, X_{t-r+1}))$$



Records in the counterfactual world:

$$p_{0,r}(t) = \mathbb{P}(X_t > \max(X_{t-1}, \ldots, X_{t-r+1}))$$



Exchangeable random sequences

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

- No need to view data to compute this probability of record: *universal yardstick*
- This equality is *relative* and does not depend on the marginal type: *bypassing bias-correction*

Can a factual realization be a record in the counterfactual world?

$p_{1,r}(t) = \mathbb{P}(Z_t > \max(X_{t-1}, \ldots, X_{t-r+1}))$

Under exchangeability of X_t and assumption A

$$p_{1,r}(t) = p_{1,r}(t)^{(m)}.$$

Under A, there is no need to correct Model m. Huge gain!!!
We can just average estimates from the different models that respect assumption A,

Under exchangeability of X_t and assumption A

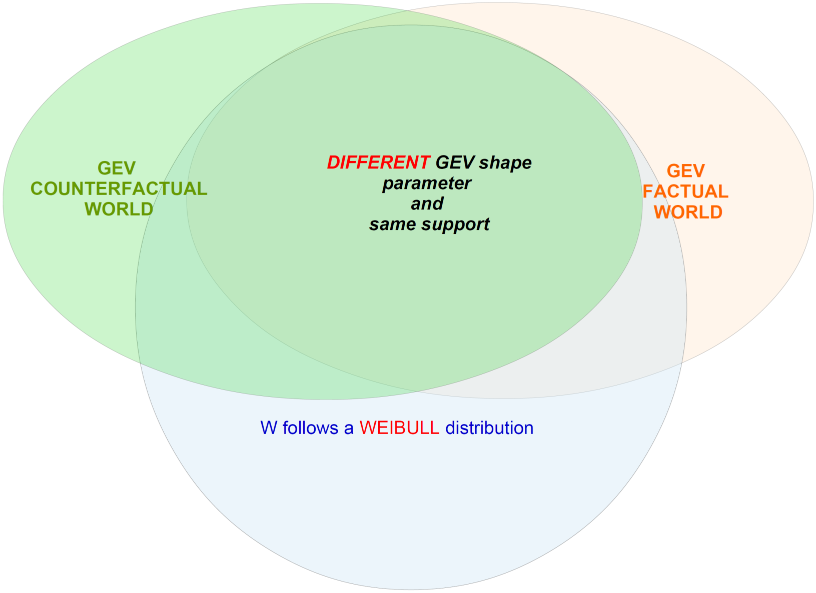
$$p_{1,r}(t) = p_{1,r}(t)^{(m)}.$$

Under A, there is no need to correct Model m. Huge gain!!!
We can just average estimates from the different models that respect assumption A,

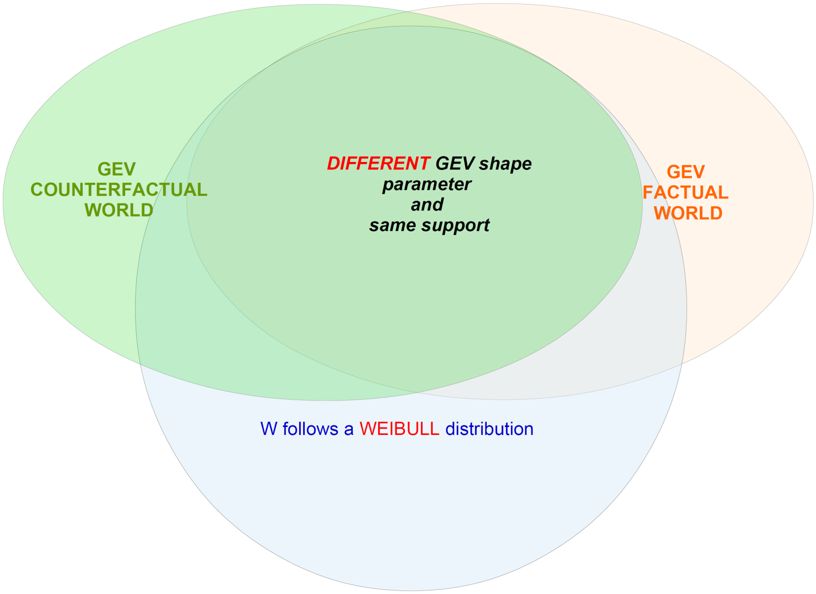
Indeed,

$$\begin{aligned}
p_{1,r}(t)^{(m)} &= \mathbb{P}(Z_t^{(m)} > \max(X_{t-1}^{(m)}, \ldots, X_{t-r+1}^{(m)})) \\
&= \mathbb{P}(Z_t^{(m)} > X_{t-1}^{(m)}, \ldots, Z_t^{(m)} > X_{t-r+1}^{(m)}) \\
&= \mathbb{P}\left(G^{\leftarrow} \circ G^{(m)}(Z_t^{(m)}) > G^{\leftarrow} \circ G^{(m)}(X_{t-1}^{(m)}), \ldots, G^{\leftarrow} \circ G^{(m)}(Z_t^{(\ldots}\right. \\
&\overset{A}{=} \mathbb{P}\left(F^{\leftarrow} \circ F^{(m)}(Z_t^{(m)}) > G^{\leftarrow} \circ G^{(m)}(X_{t-1}^{(m)}), \ldots, F^{\leftarrow} \circ F^{(m)}(Z_t^{\ldots}\right. \\
&= \mathbb{P}(Z_t > X_{t-1}, \ldots, Z_t > X_{t-r+1}) \\
&= \mathbb{P}(Z_t > \max(X_{t-1}, \ldots, X_{t-r+1})) = p_{1,r}(t)
\end{aligned}$$

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



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



If $X \sim GEV(\mu_X, \sigma_X, \xi_X)$ and $Z \sim GEV(\mu_Z, \sigma_Z, \xi_Z)$ such that

$$\mu_X - \frac{\sigma_X}{\xi_X} = \mu_Z - \frac{\sigma_Z}{\xi_Z},$$

then

$$W = -\log G(Z) \sim \mathrm{Weibull}(k, \lambda)$$

with $k = \xi_X / \xi_Z$ and $\lambda = (k \times \sigma_Z / \sigma_X)^{-1/\xi_X}$.

Estimation algorithm of $p_{1,r}(t)$

Under Weibull assumption,

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

Estimation algorithm of $p_{1,r}(t)$

Under Weibull assumption,

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

Estimation algorithm:

- Step 1: $\forall\, t$,

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

Estimation algorithm of $p_{1,r}(t)$

Under Weibull assumption,

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

Estimation algorithm:

- Step 1: $\forall\, t$,

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

- Step 2: for a chosen t,

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

Estimation algorithm of $p_{1,r}(t)$

Under Weibull assumption,

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

Estimation algorithm:

- Step 1: $\forall\, t$,

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

- Step 2: for a chosen t,

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

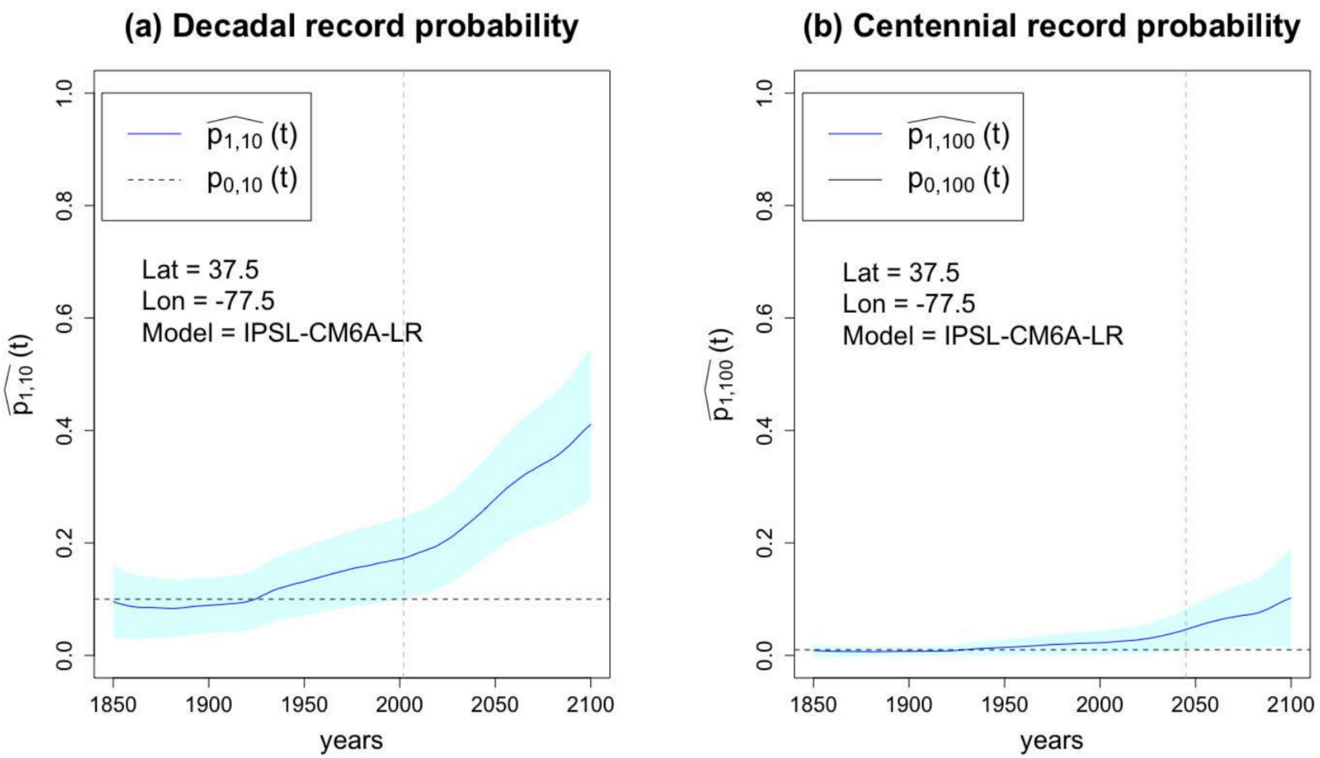
- Step 3:

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

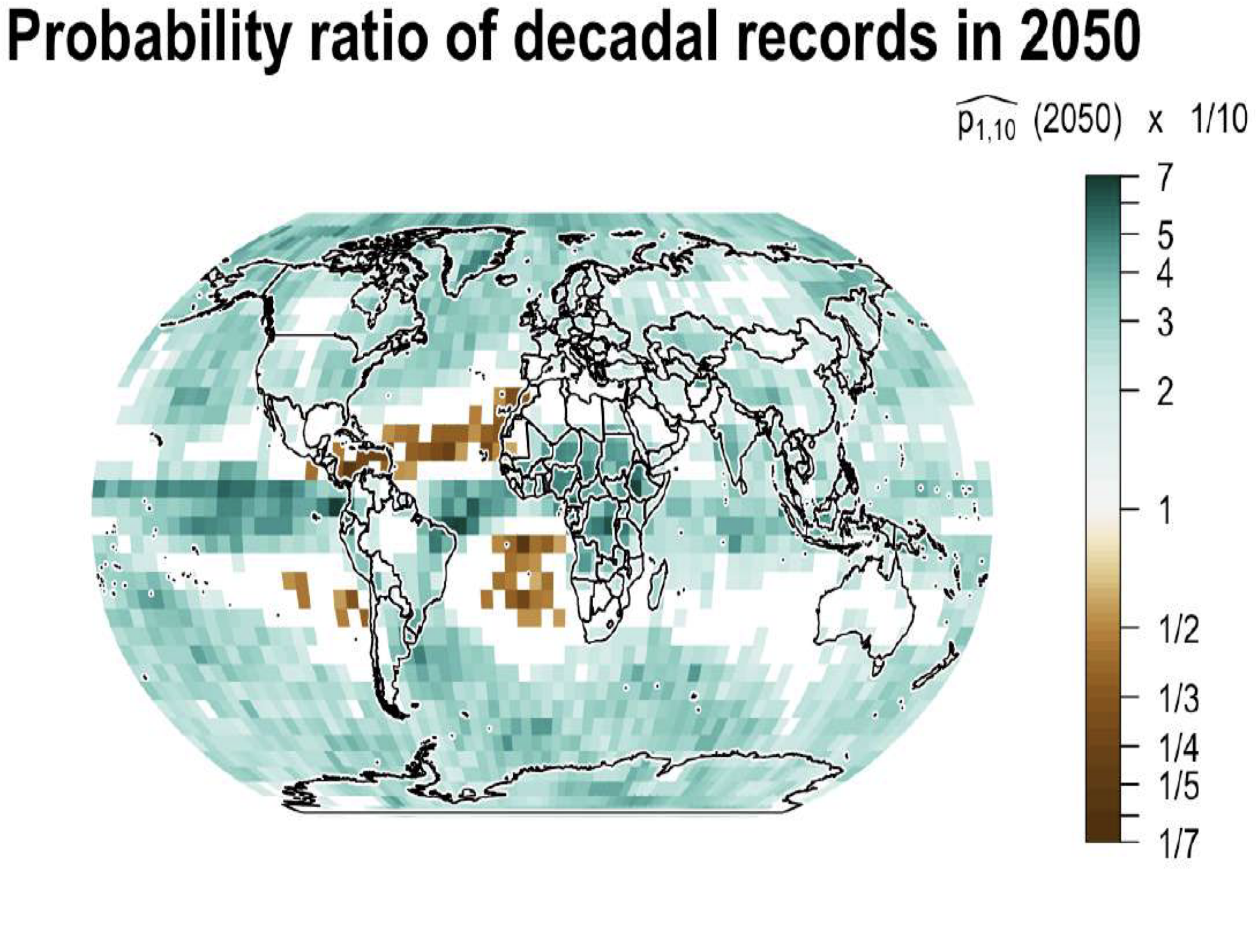
Practical facts

- GEV distributions are never fitted in this study
- The Weibull goodness of fit can be checked
- We don't need to observe a record to compute a record probability
- Extrapolation can be done, i.e. take r larger than the sample size

Yearly maxima of daily precipitation at Richmond grid-point (IPSL, SSP585)



Yearly maxima of daily precipitation (IPSL, SSS85)



How to check A

If we assume that there is a period of time, T, when the **factual climate** is like the **counterfactual climate** , i.e.

$$F_t \stackrel{d}{=} G \text{ for } t \in T$$

then, for the period T, it implies that under A

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

and that

$$p_{1,t}^{(m)} = 1/2 \text{ for } t \in T$$

Test for equality of distribution between samples $\{X_t^{(m)} : t \in T\}$ and $\{Z_t^{(m)} : t \in T\}$.

How to check A

If we assume that there is a period of time, T, when the **factual climate** is like the **counterfactual climate** , i.e.

$$F_t \overset{d}{=} G \text{ for } t \in T$$

then, for the period T, it implies that under A

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

and that

$$p_{1,t}^{(m)} = 1/2 \text{ for } t \in T$$

Test for equality of distribution between samples $\{X_t^{(m)} : t \in T\}$ and $\{Z_t^{(m)} : t \in T\}$.

How to use the observations in the setup ?

Conclusion

Conclusion

- Uncertainties in observations and climate models: using only one source of data may not give robust results.

Conclusion

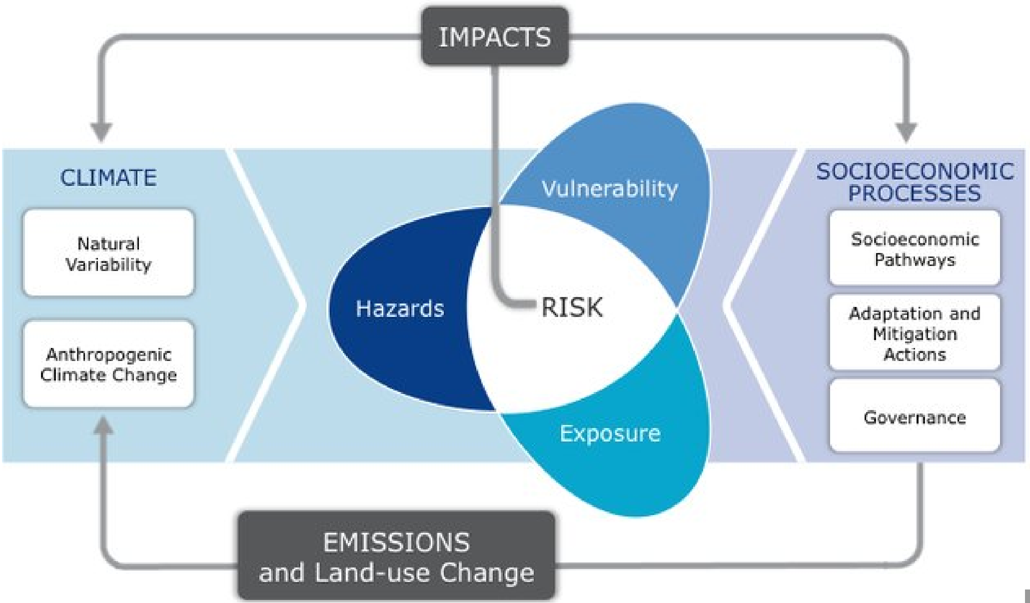
- Uncertainties in observations and climate models: using only one source of data may not give robust results.
- How to combine models and observations depends on assumptions we make about the relationship between models and observations, e.g.:
  - for threshold exceedances, climate simulations are used as a prior distrbution,
  - for records, we use assumption $A: \quad F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$ that simplifies the inference.

Conclusion

- Uncertainties in observations and climate models: using only one source of data may not give robust results.
- How to combine models and observations depends on assumptions we make about the relationship between models and observations, e.g.:
  - for threshold exceedances, climate simulations are used as a prior distrbution,
  - for records, we use assumption $A: \quad F^{\leftarrow} \circ F^{(m)} = G^{\leftarrow} \circ G^{(m)}$ that simplifies the inference.
- Extreme Event Attribution studies are limited by the quality of the data, e.g. observations availability and homogeneity or the resolution of climat models.

Some limits of the methods presented today

- Only for univariate yearly time-series.
- Further developments are needed for multivariate or coumpounds events, e.g. temporal and spatial dependencies.
- Only tackle the climate hazard part of the risk.
- Not usable for impact models that require time-series as inputs.



Thank you for your attention!

Questions, comments, suggestions?

LSCE
Laboratoire des sciences du climat & de l'environnement