Statistical methods for Detection and Attribution (D&A) in climate studies

Philippe Naveau (naveau@lsce.ipsl.fr)
Aurelien Ribes & Alexis Hannart & Francis Zwiers & Pascal Yiou

Thanks to NCAR-IMAGe, ANR-DADA, ExtremeScope,
LEFE-MULTI-RISK, Chavanah, A2C2

Outline

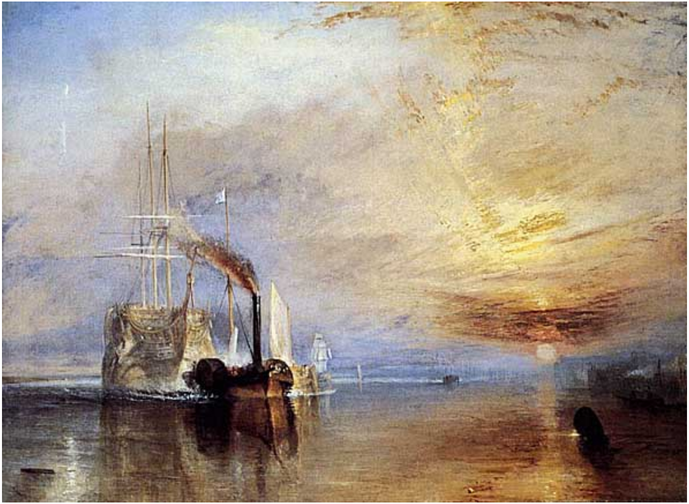
- What is D&A in climate science ?
- What are the statistical tools ?
- What are the assumptions ?
- How to deal with records & extreme events ?


STATISTICS
Detection &
Attribution
CLIMATE

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.

Antropogenic forcings



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

Detection

Demonstrating that climate or a system affected by climate has changed in some defined statistical sense [1] without providing a reason for that change.

IPCC Good Practice Guidance Paper on Detection and Attribution, 2010

1. statistically usually, significant beyond what can be explained by internal (natural) variability alone

Attribution
Evaluating the relative contributions of multiple causal factors [2] 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
- inconsistent with alternative explanations

2. 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

Examples of a "Attribution" statement (source : F. Zwiers)

Attribution results

TAR (2001)

– "most of the observed warming over the last 50 years is **likely** to have been due to the increase in greenhouse gas concentrations"

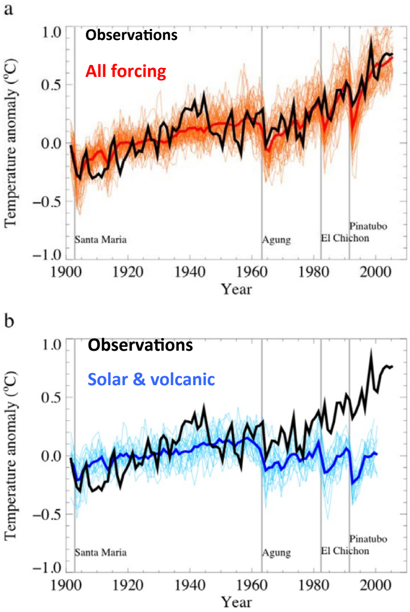
AR4 (2007)

– **likely** replaced with **very likely**
– "GHGs **likely** would have caused more warming than observed"

AR5 (2013)

– "It is **extremely likely** that human influence has been the dominant cause of the observed warming since the mid-20th century."
– "Greenhouse gases contributed a global mean surface warming **likely** to be in the range of 0.5°C to 1.3°C over the period 1951 to 2010 …"

One key idea : use climate models to generate Earth's avatars



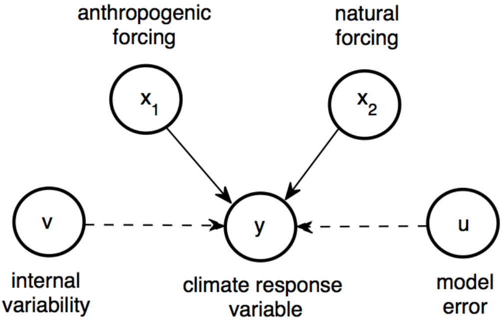
Source : Claudia Tebaldi

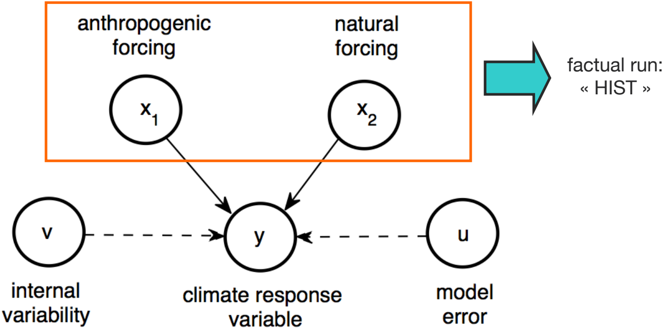
Notations

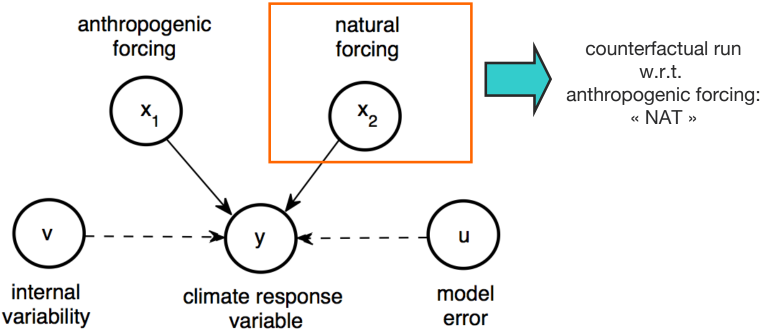
A few types of runs

- NAT (natural forcings), also called **counterfactual** runs or “world that may have been”
- ANT (anthropogenic forcings)
- ALL (anthropogenic & natural forcings), also called **factual** or HIST runs or “world that is”

Summary


anthropogenic
forcing
natural
forcing
x1
x2
v
y
u
internal
variability
climate response
variable
model
error


anthropogenic forcing
natural forcing
x_1
x_2
factual run:
« HIST »
v
y
u
internal variability
climate response variable
model error


anthropogenic forcing
natural forcing
x1
x2
v
y
u
internal variability
climate response variable
model error
counterfactual run
w.r.t.
anthropogenic forcing:
« NAT »

Two statistical and conceptual approaches in D&A

I Regression techniques : trend attribution

II Probability ratios : “event” attribution

I Regression techniques : trend attribution

What do you need in Regression techniques ?

Observations of climate indicators

Inhomogeneity in space and time (& reconstructions via proxies)

An estimate of external forcing

How external drivers of climate change have evolved before and during the period under investigation – e.g., GHG and solar radiation

A quantitative physically-based understanding

How external forcing might affect these climate indicators. – normally encapsulated in a physically-based model

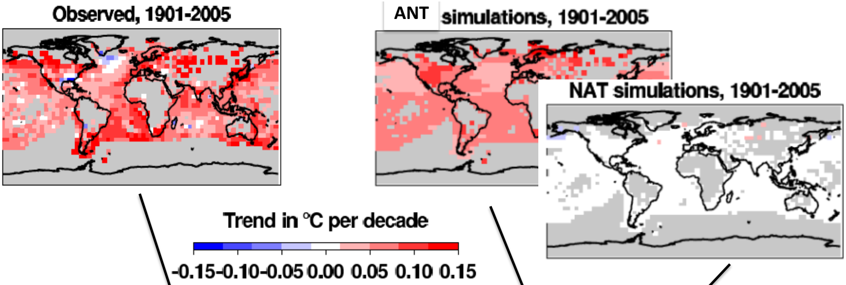
An estimate of climate internal variability Σ

Frequently derived from a physically-based model

Classical assumptions

- Key forcings have been identified
- Signals are additive
- Noise is additive
- The large-scale patterns of response are correctly simulated by climate models
- Statistical inference schemes are efficient

The basic regression scheme



$$\mathbf{Y} \qquad \mathbf{X} = (\vec{x}_{ant}, \vec{x}_{nat})$$

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

Gabi Hegerl's presentation at Geneva IPCC WG1/WG2 Meeting in Sept 2009

Error-In-Variable (EIV) (source : Ribes et al. (2016))

$$Y^* = \sum_{i=1}^{N} \beta_i X_i^*,$$

$$\begin{cases} Y &= Y^* + \varepsilon_Y, \quad \varepsilon_Y \sim N(0, \Sigma_Y), \\ X_i &= X_i^* + \varepsilon_{X_i}, \quad \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \ldots, n_f, \end{cases}$$

- Inclusion of modelling uncertainty is possible (EIV; Huntingford et al., 2006; Hannart et al., 2014)

$$\begin{cases} \Sigma_Y = \Sigma_{iv} + \Sigma_{obs}, \\ \Sigma_{X_i} = \Sigma_{iv} + \Sigma_{mod}. \end{cases}$$

- Most studies use TLS and neglect Σ_{mod} and Σ_{obs}.

Remark about climate variability

$$\boxed{\text{BIG DATA} \neq \text{ENOUGH DATA}}$$

Error-In-Variable (EIV) without β's (source : Ribes et al. (2016))

$$Y^* = \sum_{i=1}^{N} X_i^*, \tag{1}$$

$$\begin{cases} Y &= Y^* + \varepsilon_Y, \quad \varepsilon_Y \sim N(0, \Sigma_Y), & (2) \\ X_i &= X_i^* + \varepsilon_{X_i}, \quad \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \dots, n_f, & (3) \end{cases}$$

- Just remove the βs,
- Inference focuses on X_i^* (instead of β_i),
- Only additivity is assumed,
- Interpretation: models give information on each term X_i^*, then an additional constraint on the sum comes from observations.
- All inference can be made with maximum likelihood

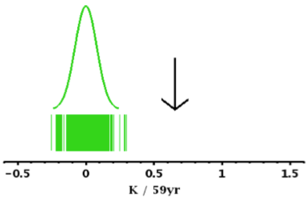
$$\widehat{X}_i^* = X_i + \Sigma_{X_i}(\Sigma_Y + \Sigma_X)^{-1}(Y - X) \sim N(X_i, \Sigma_{\widehat{X}_i^*}).$$

An example : 1951-2010 global mean temperature

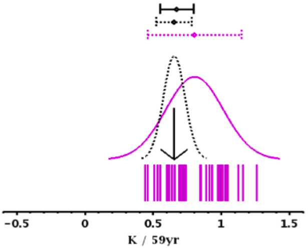
Ensembles of CMIP5 simulations used.

Climate model	Nb 60-yr PICTL seg.	Nb ALL runs	Nb NAT runs
ACCESS1-0	9	1	-
ACCESS1-3	-	1	-
bcc-csm1-1	22	3	1
bcc-csm1-1-m	-	1	-
BNU-ESM	24	1	-
CanESM2	46	5	5
CCSM4	22	6	-
CESM1-BGC	-	1	-
CMCC-CM	-	1	-
CMCC-CMS	-	1	-
CMCC-CESM	-	1	-
CNRM-CM5	47	10	6
CSIRO-Mk3-6-0	22	10	5
EC-EARTH	-	5	-
FGOALS-g2	-	2	2
FGOALS-s2	22	3	-
FIO-ESM	-	1	-
GFDL-CM3	22	-	-
GFDL-ESM2G	4	1	-
GFDL-ESM2M	-	1	-
GISS-E2-H	41	5	5
GISS-E2-R	70	5	5
HadGEM2-ES	9	4	4
HadGEM2-CC	-	1	-
inmcm4	22	1	-
IPSL-CM5A-LR	47	4	3
IPSL-CM5A-MR	-	1	-
IPSL-CM5B-LR	12	1	-
MIROC5	7	4	-
MIROC-ESM	-	1	-
MIROC-ESM-CHEM	-	1	-
MPI-ESM-LR	6	3	-
MPI-ESM-MR	22	1	-
MRI-CGCM3	22	3	-
NorESM1-M	22	3	1

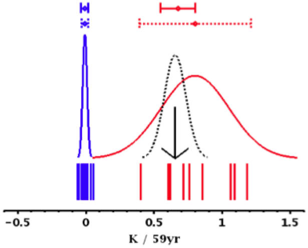
An example : 1951-2010 global mean temperature (source : Ribes et al. (2016))



Detection step



Consistency with all forcings



Obs warming: +.65K,

ALL-induced: +.67K [+.55K,+.79K],

NAT-induced: -.01K [-.03K,+.02K],

ANT-induced: +.67K [+.55K,+.80K],

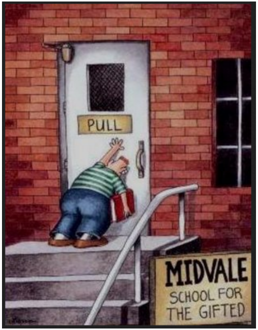
(consistent with Fig 10.5)

A short biblio

- Allen, M. R. and P. A. Stott (2003). Estimating signal amplitudes in optimal fingerprinting, part i : theory. Climate Dynamics 21, 477–491.
- Hannart, A., A. Ribes, and P. N. (2014). Optimal fingerprinting under multiple sources of uncertainty. Geophysical Research Letters 41 (4), 1261-1268.
- Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013 : Changes in temperature and precipitation extremes in the CMIP5 ensemble. Climatic Change, doi :10.1007/s10584-013-0705-8.
- Ribes A., F. Zwiers, J.M. Azais and P.N. A new statistical approach to climate change detection and attribution, Climate dynamics (in press).

II Probability ratios : “event” attribution

The FAR side



The so-called event attribution scheme

Fraction of Attributable Risk (FAR)

Relative ratio of two probabilities, p_0 the probability of exceeding a threshold in a “world that might have been (no antropogenic forcings)” and p_1 the probability of exceeding the same threshold in a “world that it is”

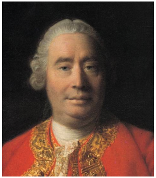
$$FAR = 1 - \frac{p_0}{p_1}.$$

(see Stott P. A., Stone D. A., Allen M. R. (2004). Human contribution to the European heatwave of 2003. Nature) + Hannart

The cornerstone of causality: counterfactual definition

- D. Hume, *An Enquiry Concerning Human Understanding*,1748

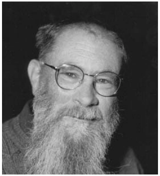
  « We may define a cause to be an object followed by another, where, if the first object had not been, the second never had existed. »



D. Hume, 18th century

- D. K. Lewis, *Counterfactuals*, 1973

  « We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects would have been absent as well. »



D. Lewis, 20th century

Coming slides : Hannart, A., Pearl J. Otto F., P. Naveau and M. Ghil. (BAMS, 2015). Counterfactual causality theory for the attribution of weather and climate-related events

Fundamental difference : necessary and sufficient causation

- Definitions:
  - “*X is a necessary cause of Y*” means that X is required for Y to occur but that other factors might be required as well.
  - “*X is a sufficient cause of Y*” means that X always triggers Y but that Y may also occur for other reasons without requiring X.
- Examples:
  - clouds are a necessary cause of rain but not a sufficient one.
  - rain is a sufficient cause for the road being wet, but not a necessary one.

Necessary and sufficient causation

- How to calculate PN, PS and PNS ?
  - difficult in general
  - closed formula under assumption of monotonicity
  - simplifies further under monotonicity and exogeneity:

$$\mathrm{PN} = 1 - \frac{p_0}{p_1}, \quad \mathrm{PS} = 1 - \frac{1 - p_1}{1 - p_0}, \quad \mathrm{PNS} = p_1 - p_0$$

FAR, « excess risk ratio »

Recall : The FAR = the relative ratio of two probabilities, p_0 the probability of exceeding a threshold in a “world that might have been (no antropogenic forcings)” and p_1 the probability of exceeding the same threshold in a “world that it is”

$$FAR = \frac{p_1 - p_0}{p_1}$$

One question

How to infer FAR (PN), PS, PNS ?

Notations

- "World that may have been" (counter-factual world)

$$(X_1, \ldots, X_m) \text{ with } G(x) = \mathbb{P}(X \leq x) \text{ \& } p_0(u) = \mathbb{P}(X > u)$$

- "World that is" (factual world)

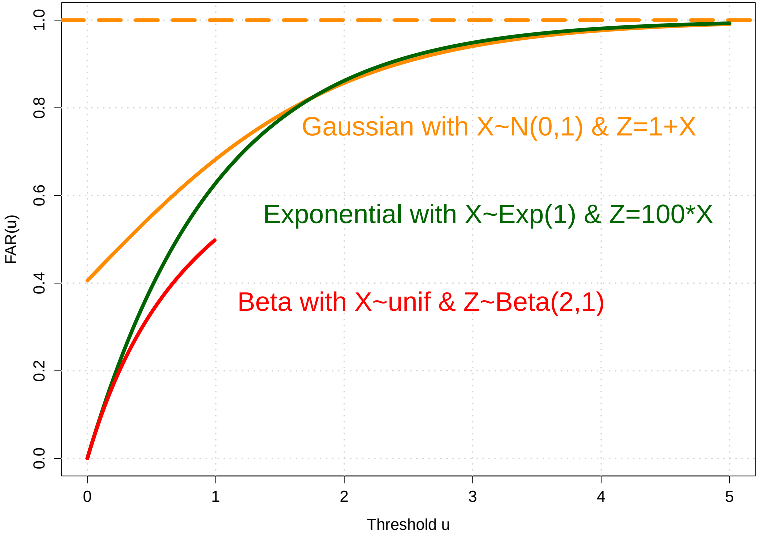
$$(Z_1, \ldots, Z_n) \text{ with } F(z) = \mathbb{P}(Z \leq z) \text{ \& } p_1(u) = \mathbb{P}(Z > u)$$

- Return level level x_r for the return period r such that $\mathbb{P}(X > x_r) = \frac{1}{r}$

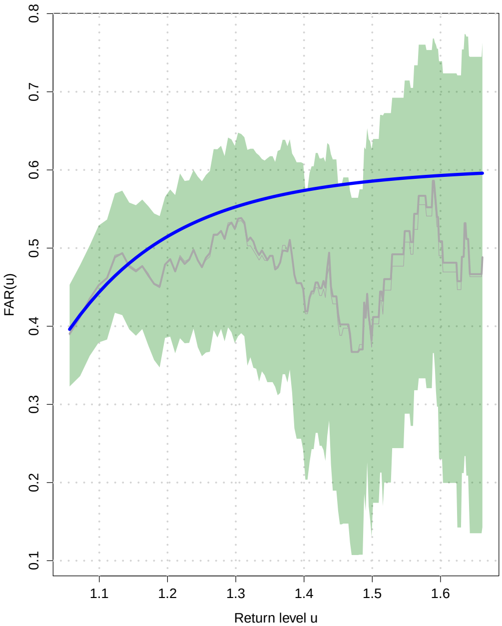
Assumptions

- **X** and **Z** two independent samples
- The two samples have the same support (values range)

Examples of $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$



Empirical inference of $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$ with $n = m = 300$

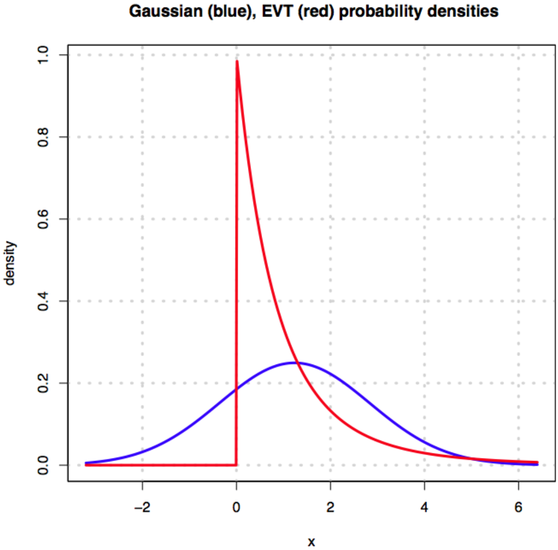


Inferential issues with $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$

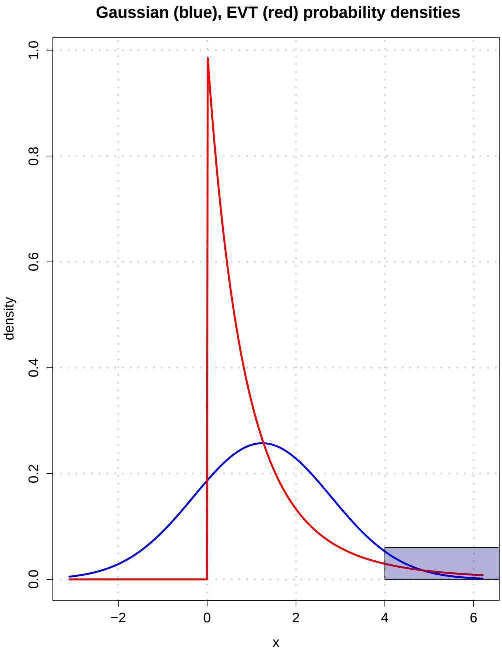
- One unknown $FAR(u)$ but two quantities ($p_0(u)$ & $p_1(u)$) to infer
- Instability of the ratio of two inferred small probabilities
- $FAR(u)$ difficult to use within a inter-model comparison assessment

“Everyone wants to be **normal**, but no one wants to be **average**”

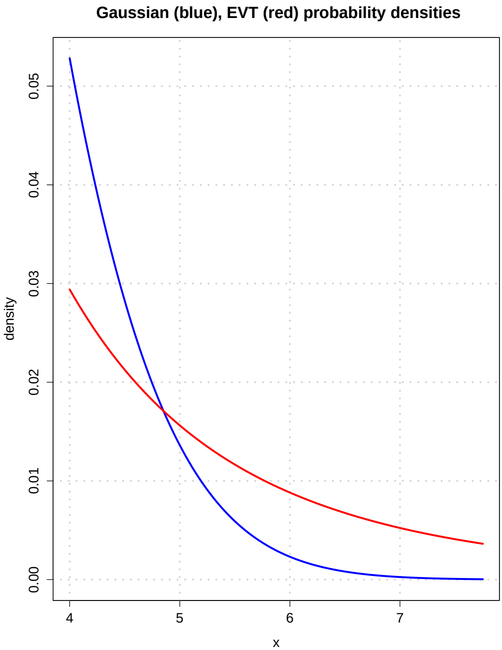
Extreme Value Theory



Extreme Value Theory

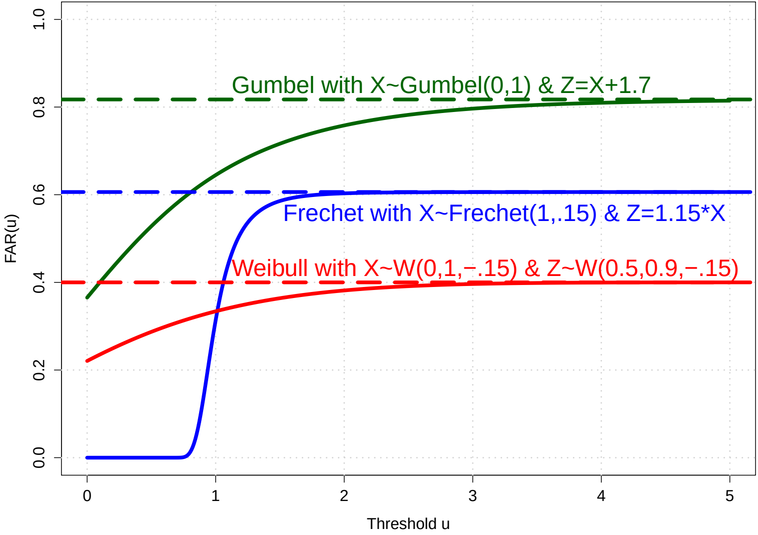


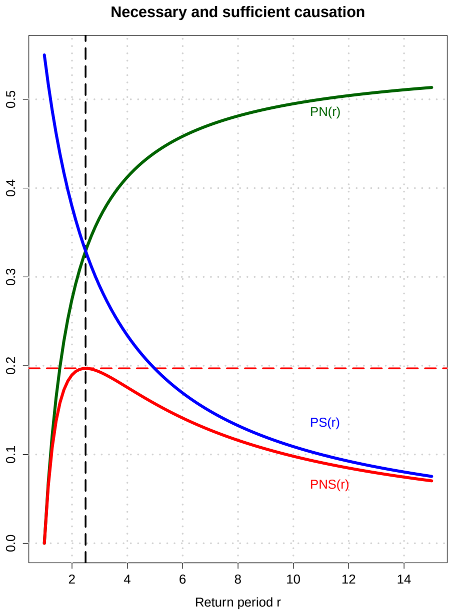
Extreme Value Theory



Bringing Extreme Value Theory in FAR

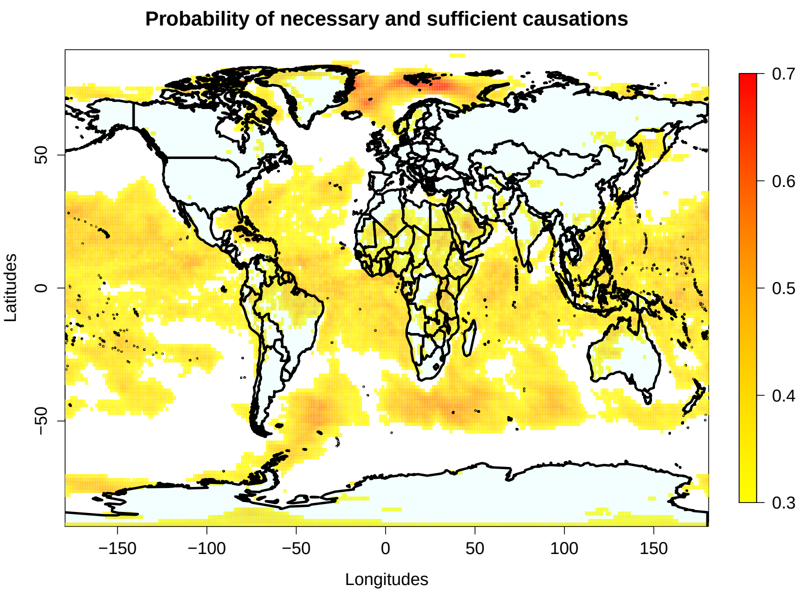
Examples of $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$





Tx Paris Montsouris : optimal return period = two years and half, apex = 0.197, PN(ra,θ)= 0.329 PS(ra,θ)= 0.329 for $\theta = .45$ & $ra = 2.491$

Annual maxima of Tx from CRMN-CM5. all forcings (1975-2005), natural forcings (1850-2012)



A short biblio

- IDAG meeting at NCAR last January https ://www2.cisl.ucar.edu/events/workshops/idag/2016/presentations
- Richard Smith, Bayesian Hierarchial Models for Extreme Event Attribution
- P.N., A. Ribes, F. Zwiers, P. Yiou. Revising return periods for record events in a climate event attribution context (in preparation)
- Hannart, A., Pearl J., Otto F., P.N., Ghil M., 2015 : Causal counterfactual theory for the attribution of weather and climate-related events. Bull. Amer. Meteor. Soc.

NCAR IMAGe TOY (Theme-Of-Year)

Do you want to give a talk on extremes (not necessarily using EVT) ?

Click on image-toy-theme-year-2015-2016

ADVANCED METHODS FOR EVALUATING WEATHER AND CLIMATE EXTREMES IN CLIMATE MODEL SIMULATIONS

Jana Sillmann and Philipe Naveau

EXTREME VALUE THEORY AND ITS APPLICATIONS

Seth Westra and Philipe Naveau

EVT & $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$

For $u > v$,

$$\frac{\mathbb{P}(X > u)}{\mathbb{P}(Z > u)} = \frac{\mathbb{P}(X > u|X > v)}{\mathbb{P}(Z > u|Z > z)} \frac{\mathbb{P}(X > v)}{\mathbb{P}(Z > v)}$$

EVT & $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$

For $u > v$ and v large

$$\frac{\mathbb{P}(X > u)}{\mathbb{P}(Z > u)} \sim \frac{\left(1 + \xi_0 \frac{u-v}{\sigma_0}\right)_+^{-1/\xi_0}}{\left(1 + \xi_1 \frac{u-v}{\sigma_1}\right)_+^{-1/\xi_1}} \frac{\mathbb{P}(X > v)}{\mathbb{P}(Z > v)}$$

FAR and RECORDS

What is a record ?

Given any return period r, the record occurs at time r if

$$X_r > \max(X_1, \ldots, X_{r-1})$$

What is a record ?

Given any return period r, the record occurs at time r if

$$X_r > \max(X_1, \ldots, X_{r-1})$$

If $X_1, \ldots, X_r$ are exchangeable,

$$\mathbb{P}(X_r > \max(X_1, \ldots, X_{r-1})) = ?$$

A magical formula

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

A magical formula

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

It is magical because

- It is always true and it is distribution free
- There is no modeling error and there is no quantity to infer
- It is very similar to return level definition

$$\mathbb{P}(X_r > x_r) = \frac{1}{r}$$

Two probabilities of records

Counterfactual world

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

Factual world

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

Two probabilities of records

Counterfactual world

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

Factual world

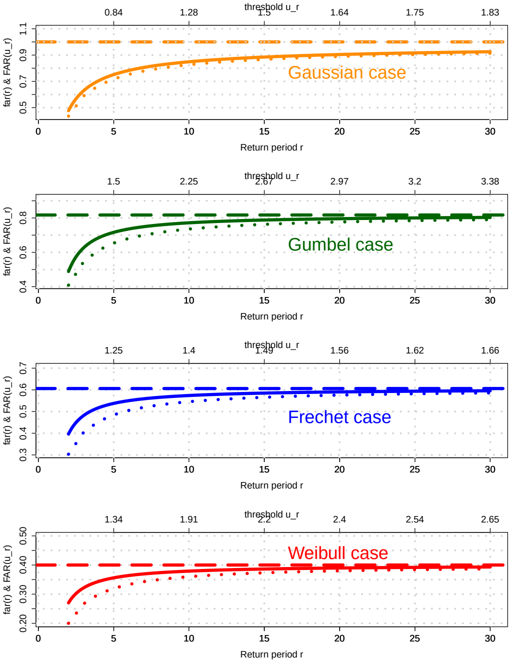
$$p_{1,r} = \mathbb{P}(Z_r > \max(X_1, \ldots, X_{r-1}))$$

A new FAR

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

How far $far(r)$ is from $FAR(u)$? Not that far

Comparing $FAR(u)$ (solid lines)& $far(r)$ (dotted lines)



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

New expression

$$p_{1,r} = \mathbb{E}\left(G(Z)^{r-1}\right)$$

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

New expression

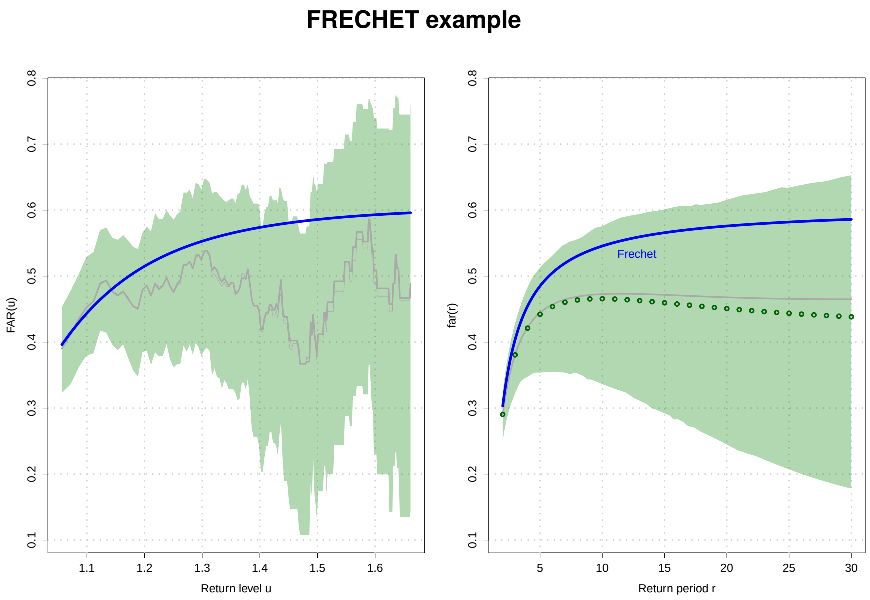
$$p_{1,r} = \mathbb{E}\left(G(Z)^{r-1}\right)$$

Empirical estimator

$$\widehat{p_{1,r}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{G}_m^{r-1}(Z_i),$$

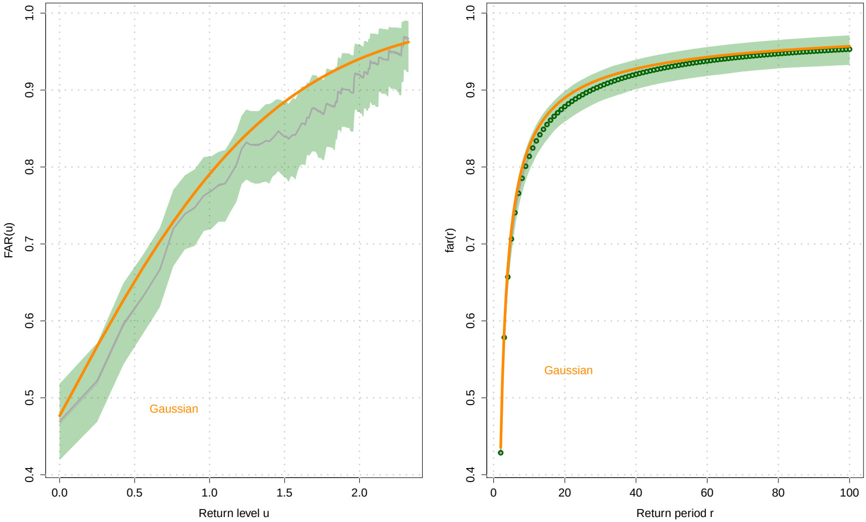
where $\mathbb{G}_m(z)$ the classical empirical cumulative distribution function

Comparing non-parametric estimators of $FAR(u)$ (left)& $far(r)$ (right)



Comparing non-parametric inference of $FAR(u)$ (left)& $far(r)$ (right)

GAUSSIAN example



Preliminary conclusion about FAR & far :

$$\text{Using } far(r) \text{ improves the non-parametric inference}$$

Only one unknown $p_{1,r} = \mathbb{P}(Z_r > \max(X_1, \ldots, X_{r-1}))$

A key relationship

$$p_{1,r} = \mathbb{E}\left(G(Z)^{r-1}\right) = \mathbb{E}\left(\exp(-(r-1)W)\right) \text{ with } W = -\log G(Z)$$

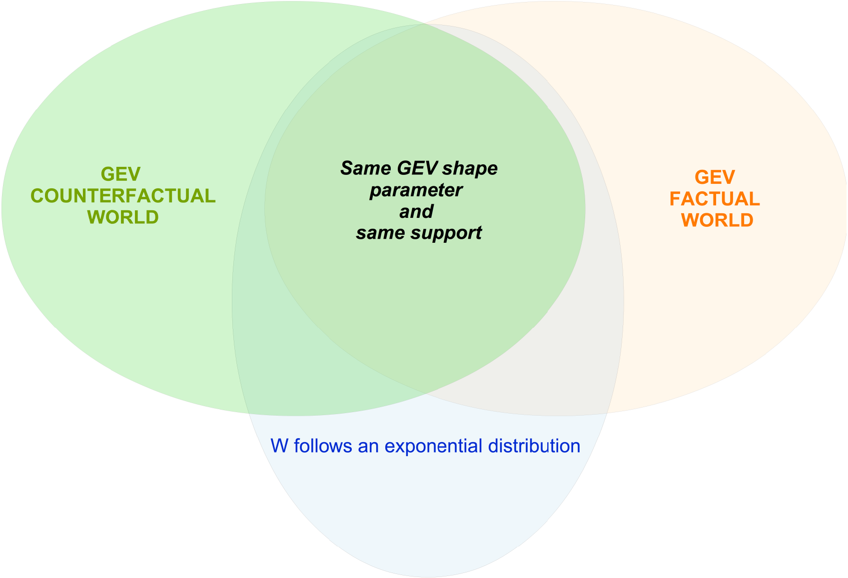
A simple idea

$p_{1,r}$ is the moment generating function of the positive random variable W and we could impose a parametric form on W

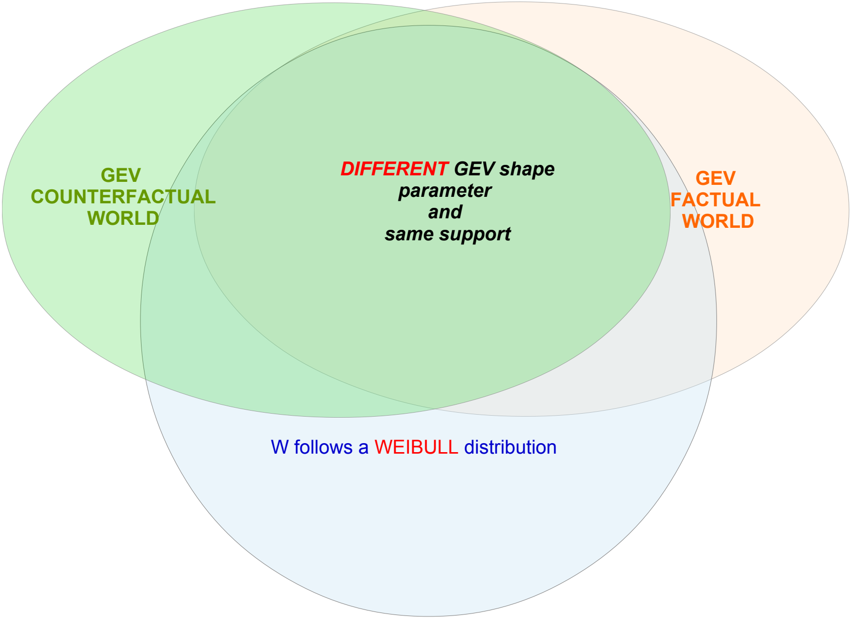
if $F = G$

Then, W follows an exponential distribution with unit mean

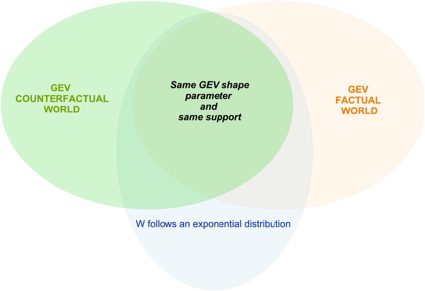
$W = -\log G(Z)$ when same GEV parameters



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



Coming back to $far(r)$ with $W = -\log G(Z)$ exponential



$$\boxed{far(r) = (1-\theta)\left(1-\frac{1}{r}\right)}$$

Inference of $far(r)$ if $W = -\log G(Z)$ is exponentially distributed

$$\widehat{far}_{\text{Exp}}(r) = \left(1 - \widehat{\theta}\right)\left(1 - \frac{1}{r}\right)$$

with $\widehat{\theta}$ fully decoupled from r and the relative error does not depend on r

$$\frac{\sqrt{var\left(\widehat{far}_{\text{Exp}}(r)\right)}}{far_{\text{Exp}}(r)} = \frac{\sqrt{var\left(1 - \widehat{\theta}\right)}}{1 - \theta}$$

Inference of $far(r)$ if $W = -\log G(Z)$ is exponentially distributed

$$\widehat{far}_{\text{Exp}}(r) = \left(1 - \widehat{\theta}\right)\left(1 - \frac{1}{r}\right)$$

with $\widehat{\theta}$ fully decoupled from r and the relative error does not depend on r

$$\frac{\sqrt{var\left(\widehat{far}_{\text{Exp}}(r)\right)}}{far_{\text{Exp}}(r)} = \frac{\sqrt{var\left(1 - \widehat{\theta}\right)}}{1 - \theta}$$

Inference of θ

$$\widehat{\theta} = 1 - 2\widehat{far}_{\text{Exp}}(2) = 1 - 2\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_m(Z_i),$$

Inference of $far(r)$ if $W = -\log G(Z)$ is exponentially distributed

$$\widehat{far}_{\text{Exp}}(r) = \left(1 - \widehat{\theta}\right)\left(1 - \frac{1}{r}\right)$$

with $\widehat{\theta}$ fully decoupled from r and the relative error does not depend on r

$$\frac{\sqrt{var\left(\widehat{far}_{\text{Exp}}(r)\right)}}{far_{\text{Exp}}(r)} = \frac{\sqrt{var\left(1 - \widehat{\theta}\right)}}{1 - \theta}$$

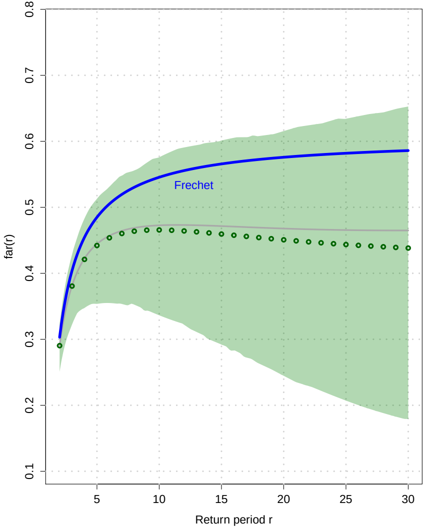
Inference of θ

$$\widehat{\theta} = 1 - 2\widehat{far}_{\text{Exp}}(2) = 1 - 2\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_m(Z_i),$$

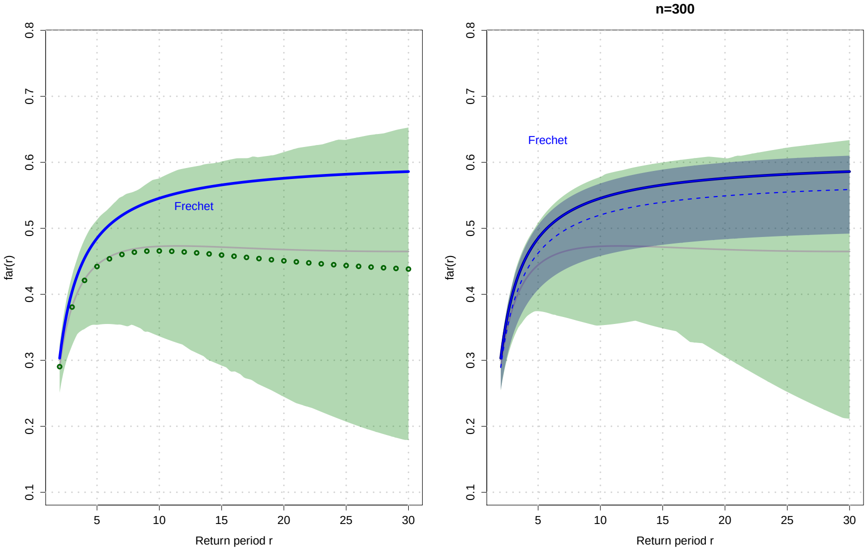
BIG BONUS

No need to estimate any EVT parameters !

Comparing parametric versus non parametric of $far_n(u)$ (green)& $far_{exp}(r)$ (blue)



Comparing parametric versus non parametric of $far_n(u)$ (green)& $far_{exp}(r)$ (blue)

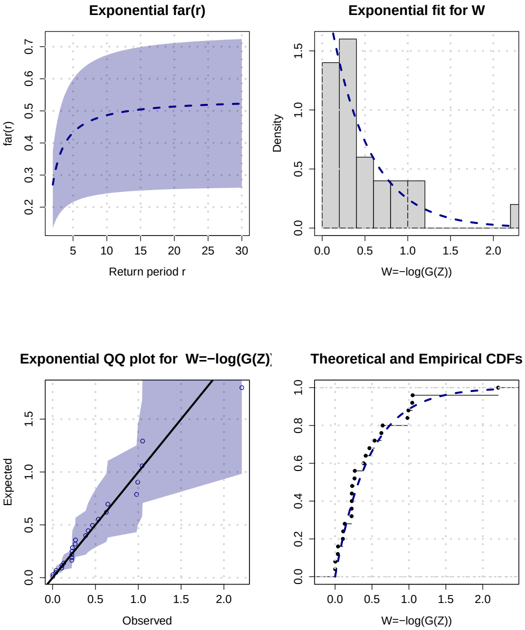


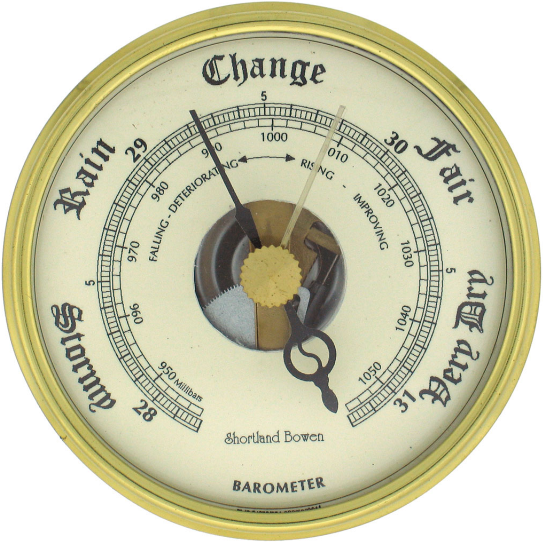
Does this work in practice ?

Does $W = -\log G(Z)$ can be exponentially or Weibull distributed ?

Examples dealing with yearly maxima of daily temperatures

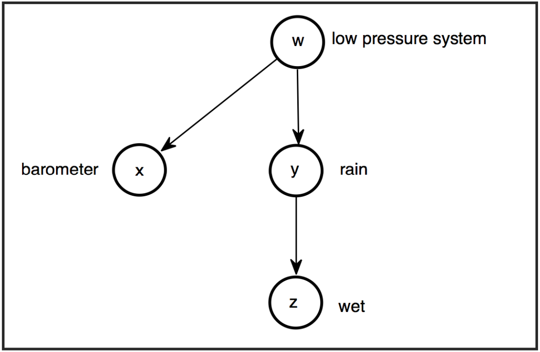
Annual max of Tx in Paris. 1900-1930 = “counterfactual” and 1990-2015=“factual” world




Change
Rain
Fair
Stormy
Very Dry
29
30
28
31
5
1000
980
970
960
950 Millibars
1020
1030
1040
1050
FALLING - DETERIORATING
RISING - IMPROVING
Shortland Bowen
BAROMETER

Oriented graphs

— visual representation of the conditional independence structure of a joint distribution



$$P(X, Y, Z, W) = P(W) \,.\, P(X \mid W) \,.\, P(Y \mid W) \,.\, P(Z \mid Y)$$

Interventional probability

- Limitation of oriented graphs
  - identifiability: several causal graphs are compatible with the same pdf (and hence with the same observations).

$$P(X,Y) = \underbrace{P(X)\,.\,P(Y \mid X)}_{X \,\rightarrow\, Y} = \underbrace{P(Y)\,.\,P(X \mid Y)}_{Y \,\rightarrow\, X}$$

  - Need for disambiguation.

$\Rightarrow$ **experimentation**

Interventional probability

- New notion:
  - intervention $do(X=x)$
  - interventional probability $P(Y \mid do(X=x)) = P(Y_x)$

the probability of rain **forcing** the barometer to decrease,
in an experimental context in which the barometer is manipulated

$$P(Y \mid do(X=x)) \neq P(Y \mid X=x)$$

the probability of rain **knowing** that the barometer is decreasing,
in a non-experimental context in which the barometer evolution is left unconstrained

Fundamental difference : necessary and sufficient causation

- Definitions:
  - **<u>Probability of necessary causality = PN</u>** = the probability that the event Y would not have occurred in the absence of the event X given that both events Y and X did in fact occur.
  - **<u>Probability of sufficient causation = PS</u>** = the probability that Y would have occurred in the presence of X, given that Y and X did not occur.
- Formalization:

$$\begin{cases} \text{PN} =_{\text{def}} P(Y_0 = 0 \mid Y = 1, X = 1) \\ \text{PS} =_{\text{def}} P(Y_1 = 1 \mid Y = 0, X = 0) \\ \text{PNS} =_{\text{def}} P(Y_0 = 0, Y_1 = 1) \end{cases}$$

Empirical inference of $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$

Worst case scenario

$p_1(u) = (1+\epsilon)p_0(u)$ with for some $\epsilon > 0$ and large u

Relative error for $FAR(x_r)$ for large r

$$\frac{\text{sdev}\left(\widehat{FAR}_n(x_r)\right)}{FAR(x_r)} \sim \frac{1}{\epsilon}\sqrt{\frac{r}{n}}\sqrt{\frac{2+\epsilon}{1+\epsilon}}$$

Empirical inference of $FAR(u) = 1 - \frac{p_0(u)}{p_1(u)} = 1 - \frac{\mathbb{P}(X>u)}{\mathbb{P}(Z>u)}$

Worst case scenario

$p_1(u) = (1+\epsilon)p_0(u)$ with for some $\epsilon > 0$ and large u

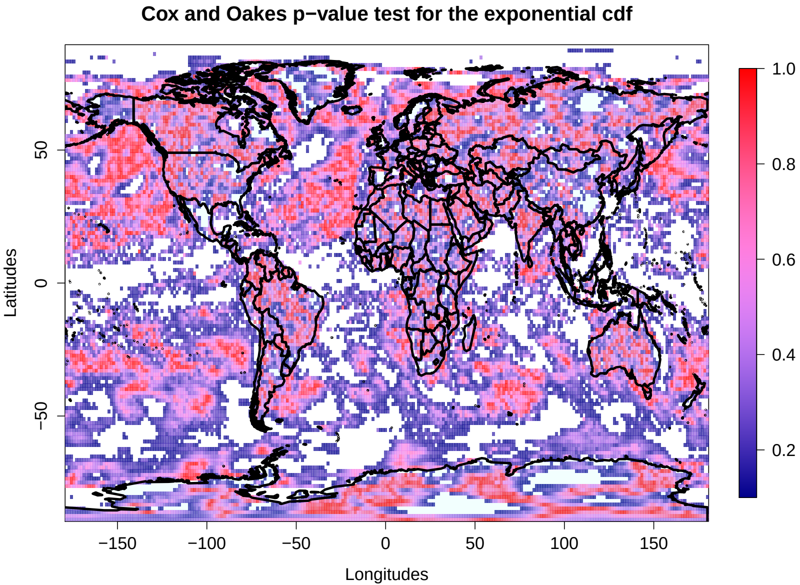
Relative error for $FAR(x_r)$ for large r

$$\frac{\text{sdev}\left(\widehat{FAR}_n(x_r)\right)}{FAR(x_r)} \sim \frac{1}{\epsilon}\sqrt{\frac{r}{n}}\sqrt{\frac{2+\epsilon}{1+\epsilon}}$$

Lessons learned

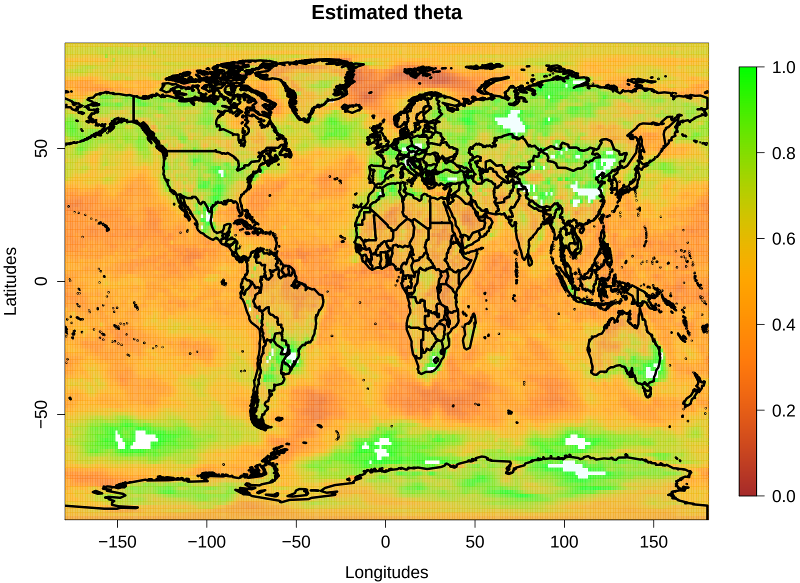
- The sample size n should be much greater than the return period r
- The relative risk explodes if ϵ near zero, i.e. small difference between the factual and counterfactual worlds, e.g. precipitation.
- Tens of thousand ensemble runs or even more are needed to estimate empirically FAR for even moderate r

Annual maxima of Tx from CRMN-CM5. all forcings (1975-2005), natural forcings (1850-2012)



Annual maxima of Tx from CRMN-CM5. all forcings (1975-2005), natural forcings (1850-2012)

$$far(r) = (1-\theta)\left(1-\frac{1}{r}\right)$$



An example of a "Detection" statement

"Warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia. The atmosphere and ocean have warmed, the amounts of snow and ice have diminished, sea level has risen, and the concentrations of greenhouse gases have increased."

IPCC-WG1-AR5 SPM

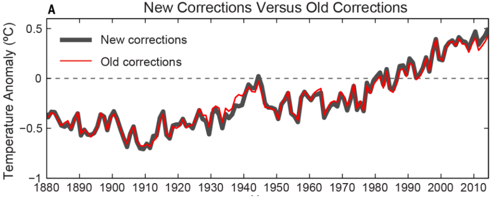


Fig. 2. Global (land and ocean) surface temperature anomaly time series with new analysis, old analysis, and with and without time-dependent bias corrections. (**A**) The new analysis (solid black) compared to the old analysis (red). (**B**) The new analysis (solid black) versus

Karl et al, 2015, Science