Record events attribution in climate studies

Julien Worms, Philippe Naveau

Environmetrics 00, 1–42

DOI: 10.1002/env.XXXX

Record events attribution in climate studies

Worms, J.[a]*, Naveau, P., [b]

Summary: Within the statistical climatology literature, inferring the contributions of potential causes with regard to climate change has become a recurrent research theme during this last decade. In particular, disentangling human induced (anthropogenic) forcings from natural causes represents a non-trivial statistical task, especially when the focal point moves away from mean behaviors and goes towards extreme events with high societal impacts. Most studies found in the field of Extreme Event Attributions (EEA) rely on Extreme Value Theory (EVT). Under this theoretical umbrella, it is often assumed that, for a given location, temporal changes in extremes can be detected in both location and scale parameters of an extreme value distribution, while its shape parameter remains unchanged over time. This assumption of constant tail shape parameters between a so-called factual world (all forcings) and a counterfactual one (without anthropogenic forcing) can be challenged due to the fact that important forcing changes could impact large scale atmospheric and oceanic circulation patterns, and consequently, the later can reshape the full distribution, including its shape parameter that drives extremal behavior.

In this paper, we study how allowing different extremal tail indices between the factual and counterfactual worlds can affect the analysis of records. In particular, we extend the work of Naveau *et al.* (2018) in which this case was not treated. We also add properties and theoretical inferential results about records in EEA and propose a procedure for model validation. A simulation study of our approach is detailed. Our method is applied on records of yearly maxima of daily maxima of near-surface air temperature issued from the numerical climate model CNRM-CM5 of Météo-France.

Keywords: Generalized Extreme Value, Weibull distribution, Causality theory, climate detection and attribution

[a] *Université Paris Saclay, UVSQ, Laboratoire de Mathématiques de Versailles, CNRS, Versailles, France*

[b] *Université Paris-Saclay, UVSQ, Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL/ESTIMR, CNRS-CEA, Gif-sur-Yvette, France*

[*] *Correspondence to: julien.worms@uvsq.fr*

This paper has been submitted for consideration for publication in *Environmetrics*

1. INTRODUCTION

The field of extreme event attribution (EEA) in climatology (see for reviews, e.g., Naveau *et al.*, 2020; Stott *et al.*, 2016) aims at comparing how the likelihood of a specific extreme climate event can change under different inputs, the so-called forcings in atmospheric sciences. The most studied driver is the increase of greenhouse gas concentrations induced by human activities. Depending on enduser needs, two questions are typically asked in this research domain. To highlight scientific evidencing (e.g., Stott *et al.*, 2004), one may want to assess if the probability of today extreme climate events could have been different without anthropogenic emissions. Besides improving scientific understanding of physical mechanisms, this setup can also be useful for litigation cases and such a framing is closely linked to causality theory (e.g., Hannart *et al.*, 2016). The thought experiment of imagining a world without anthropogenic forcings is called *counterfactual* because it did not happen in reality. In contrast, the expression *factual world* corresponds to today forcing conditions, i.e. with anthropogenic forcing included. Another example treated in EEA could be the case of a flood planner manager who is interested by revising 100-year return levels according to future forcing conditions.

In these setups, *in silico* numerical experiments are needed to provide possible climate trajectories under never observed forcings, either in the past, present or future. For example, our application in Section 5 will focus on numerical outputs from the global climate model CNRM-CM5 of Météo-France. A factual dataset will correspond to a CNRM-CM5 run with all forcings (natural and anthropogenic forcings) during the years 1975-2005, mimicking today's climate. The counterfactual data will be represented by a different CNRM-CM5 run only driven by natural forcings over the period 1850-2012, a never observed trajectory.

An important step in any EEA analysis is the choice of extreme events under study. Statistically, the extreme event of interest is classically defined as an exceedance above a large threshold, say u, for some atmospheric feature. More precisely, climatologists have compared

the survival probabilities $p_1 = \mathbb{P}(X > u)$ and $p_0 = \mathbb{P}(Z > u)$ where the random variable Z with cdf F takes its values in the factual world and the random variable X with cdf G represents the same atmospheric parameter but in the counterfactual word. In practice, the threshold u is set relatively to an exceptional event, say the 2003 European heatwave analysis in Stott *et al.* (2004), and relative ratios like p_1/p_0 or the so-called fraction of attributable risk, $FAR(u) = 1 - p_0/p_1$, are inferred from specific numerical model experiments aiming at reproducing both factual and counterfactuals worlds (see, the bibliographies in Stott *et al.*, 2016; Naveau *et al.*, 2020, for details). Still, defining extremes as exceedances above a high threshold is not the only way to capture rare events distributions. For example, one can be interested by records. The American Meteorological Society in its bulletin produces yearly special reports about EEA (see, e.g., Kew *et al.*, 2019). In such reports, the specific event under study is often chosen because a new record has been broken for a specific variable of interest (see, e.g., King, 2017). Mathematically, Y_r is said to a record at time r with respect to the sample $(Y_1, Y_2, \dots)$ if

$$Y_r \geq \max(Y_1, \dots, Y_{r-1}).$$

The advantage of defining extremes as records is that there is no need to choose a threshold. In addition, the interpretation is straightforward for the general public. It simply means that the latest recorded value is the largest. Naveau *et al.* (2018) studied record probabilities within a EEA context by introducing the following relative ratio

$$far(r) = 1 - \frac{p_{0,r}}{p_{1,r}} \quad \text{where} \quad \begin{cases} p_{0,r} &= \mathbb{P}(X_r > \max\{X_1, \dots, X_{r-1}\}), \\ p_{1,r} &= \mathbb{P}(Z_r > \max\{X_1, \dots, X_{r-1}\}), \end{cases}$$

for every integer number $r \geq 2$. Clearly, $p_{0,r} = \mathbb{P}(X_r > \max\{X_1, \dots, X_{r-1}\})$ can be interpreted as the probability that X_r is a record within the counterfactual sample $(X_1, \dots, X_r)^T$. One can also notice that $p_{0,r}$ is equivalent to set $u = \max\{X_1, \dots, X_{r-1}\}$ into $\mathbb{P}(X > u)$. By replacing $u = \max\{X_1, \dots, X_{r-1}\}$ into $\mathbb{P}(Z > u)$, we obtain $p_{1,r}$. So, $p_{1,r}$

indicates the probability of observing a factual record with respect to the counterfactual world. Concerning the computation of $p_{0,r}$, exchangeability of the sequence (X_j) implies that

$$p_{0,r} = \frac{1}{r} \quad \text{and} \quad far(r) = 1 - \frac{1}{rp_{1,r}}.$$

Then, under exchangeability of X_i's, the difficult estimation of the two small probabilities in $FAR(u) = 1 - \mathbb{P}(Z > u)/\mathbb{P}(X > u)$ is completely bypassed and replaced by the single inference of $p_{1,r}$. This greatly simplifies the statistical inference. In this work, the assumption of exchangeability appears to be reasonable in the counterfactual world[†]. The non-stationarity of anthropogenic forcing implies that exchangeability is not tenable in the factual world. This departure will be taken into account by assuming stationarity within a climatological period (1975-2005). In terms of interpretation of $FAR(u)$ and $far(r)$, we refer to Hannart and Naveau (2018) who detailed the link between these quantities and the causal counterfactual probability theory of Pearl (2000). Still, we want to remind that a $far(r)$ close to one indicates a strong probability of necessary causation, meaning that the cause under study (here increase in greenhouse gas concentrations) is required for extreme records but other factors might be required as well. A $far(r)$ between zero and one gives us the degree of such necessary causation.

Naveau *et al.* (2018) provided the basic ingredients to infer $p_{1,r}$, and a few cases were explored. In particular, the assumption of equal tail shape parameters of the factual and counterfactual distributions was used in all examples and case studies. More precisely, the generalized extreme value distribution (GEV), classically used to model block maxima (see

[†]The exchangeability assumption is satisfied in both worlds if the numerical experiments at hand are viewed as ensemble resampling of the same event within the factual and factual worlds. In our application we use transient climate runs, but not ensembles of the same event

the books of Coles, 2001; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006),

$$GEV(x;\mu,\sigma,\xi) = \exp\left(-\overline{H}_\xi\left(\frac{x-\mu}{\sigma}\right)\right), \text{ with } \overline{H}_\xi(x) = \begin{cases} (1+\xi x)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x), & \xi = 0, \end{cases} \quad (1)$$

has a shape parameter ξ that provides key information about the tail behavior[‡]. In Naveau *et al.* (2018), the GEV case where the factual and counterfactual shape parameters were assumed equal leads to a very simple expression of

$$far(r) = (1-\lambda)\,(1-1/r)\,,$$

where the unique parameter λ only depends on the scaling and location GEV parameters, but not on the common shape parameter ! This simple expression brought fast and efficient estimation schemes. Still, the assumption of equal shape parameters can be challenged in some EEA studies. For example, one can imagine that, significative changes in circulation patterns could modify the climate of regions affected by these patterns, and consequently the factual shape parameter could be different from the counterfactual one. If true, a difference in positive shape parameters is particularly important for risk assessment of extremes. Another limitation of Naveau *et al.* (2018) is that the asymptotic properties of non-parametric estimators of $far(r)$ were restricted to sample sizes of the same length. In practice, factual and counterfactual runs may have different runs and so, this was a limit in terms of applicability. Finally, the study of confidence intervals was briefly touched upon in Naveau *et al.* (2018), but not fully complete and needed to be improved.

In this work, we propose to address all the above shortcomings. Section 2 recalls the basic setup and extends the non-parametric approach to samples of different sizes. Section

[‡]A zero shape parameter corresponds to an exponential tail, while $\xi > 0$ characterizes heavy tails and $\xi < 0$ implies a bounded support. For example, block maxima of precipitation tend to be slightly heavy-tailed, (see, e.g., Katz *et al.*, 2002), while block maxima of temperatures appear to have negative ξ's, (see, e.g., Kharin *et al.*, 2007). The scale parameters σ has to be positive, $a_+ = \max(a, 0)$ and $\mu \in R$.

3 provides the main theoretical results. Under a specific class of distributions that allow for different factual and counterfactual tail behaviors, we introduce a new inference scheme and derive asymptotic convergence results. The question of model validation is also addressed. Sections 4 focuses on our simulation study, while Section 5 provides an application to temperatures maxima for the aforementioned CNRM climate model. Section 7 contains the proofs of the results and some discussion on the computational aspects of this work.

2. GENERAL SETUP AND NON PARAMETRIC APPROACH

In this work, we assume that the counterfactual and factual data samples $(X_1, \ldots, X_m)$ and $(Z_1, \ldots, Z_n)$ respectively represent iid copies of $X \sim G$ and $Z \sim F$. In addition, the time series X_i's and Z_i's are supposed to be independent. Under these conditions, $p_{1,r}$ for any integer $r \geq 2$ can be rewritten as an expectation

$$p_{1,r} = \mathbb{E}\left((G(Z))^{r-1}\right). \tag{2}$$

At this stage, one can notice that there is not need to *actually observe records at time* r to estimate $p_{1,r}$. For any given r, we simply need to estimate the expectation defined by (2). More precisely, the natural non-parametric estimator of $p_{1,r}$ can be obtained by replacing the unknown cdf G by any empirical estimator, say $\widehat{G}_m$,

$$\widehat{p}_{1,r} = \frac{1}{n}\sum_{i=1}^{n} \widehat{G}_m^{r-1}(Z_i), \tag{3}$$

where the sample sizes of the factual and counterfactual worlds are denoted n and m, respectively. In this work, $\widehat{G}_m$ is defined as the following slight modification of the usual empirical distribution function

$$\widehat{G}_m(x) = \frac{b + \sum_{j=1}^{m} \mathbb{I}_{X_j \leq x}}{m+1}, \tag{4}$$

where $\mathbb{I}_{X\leq x}$ corresponds to the indicator function, equal to one when $X \leq x$ and zero otherwise, $b \in]0,1[$ is a small constant chosen to avoid $\widehat{G}_m(x) = 0$ for practical reasons explained in Section 3.5. The corresponding non-parametric estimator of $far(r)$ is then defined as

$$\widehat{far}(r) = 1 - \frac{1}{r\widehat{p}_{1,r}}.$$

Our first proposition is the generalization of Formula (A1) in Naveau *et al.* (2018), see Appendix (A1) for elements of the proof. This is a non-parametric result that shows the asymptotic convergence in distribution of our estimator when the sample sizes can differ.

Proposition 1 *Let $\widehat{G}_m$ denote the empirical cdf defined by* (4)*. As n and m go to infinity, if $\sqrt{n/m}$ converges to some finite constant $a \geq 0$ then we have*

$$\sqrt{n}(\widehat{p}_{1,r} - p_{1,r}) \xrightarrow{d} \mathcal{N}(0, \sigma_r^2) \quad \textit{and} \quad \sqrt{n}(\widehat{far}(r) - far(r)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_r^2}{(rp_{1,r}^2)^2}\right) \tag{5}$$

where

$$\sigma_r^2 = \tau_r^2 + a\times(r-1)^2(M_r - p_{1,r}^2),$$

with

$$\tau_r^2 = \mathbb{V}ar(G^{r-1}(Z)) = p_{1,2r-1} - p_{1,r}^2 \textit{ and } M_r = \mathbb{E}\left(G^{r-2}(Z_1)G^{r-2}(Z_2)(\min(G(Z_1), G(Z_2)))\right).$$

The second term in σ_r^2 corresponds to the cost of replacing the true cdf G by its empirical estimate $\widehat{G}_m$. The price can be small (large) whenever the counterfactual sample is much larger (smaller) than the factual one. The expression of σ_r^2 leads to the following estimate of the asymptotic variance

$$\widehat{\sigma}_r^2 = (\widehat{p}_{1,2r-1} - \widehat{p}_{1,r}^2) + \sqrt{\frac{n}{m}}(r-1)^2(\widehat{M}_{r,n,m} - \widehat{p}_{1,r}^2)$$

where

$$\widehat{M}_{r,n,m} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widehat{G}_m(Z_i))^{r-2}(\widehat{G}_m(Z_j))^{r-2}\min\left(\widehat{G}_m(Z_i),\widehat{G}_m(Z_j)\right).$$

As expected from any non-parametric estimator of small probabilities, the asymptotic variance of the relative error of $\widehat{p}_{1,r}$ grows as r gets large, see the upcoming Proposition 3 for a specific example. To deal with extreme records, i.e. when r is larger than sample sizes, additional assumptions are needed, see Section 3. Still, Proposition 1 will be very useful for $r = 1$ and $r = 2$ when inferring first and second moments.

3. MAIN RESULTS

3.1. Model setup

To estimate accurately $p_{1,r}$ for large r, we need to identify an appropriate class of bivariate random variables (X, Z). To define this family, one can notice that the expectation defined by (2) can be rewritten as a Laplace transform

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

At this stage, it appears natural to define a class of (X, Z) with respect to the random variable W. This variable characterizes a type of relative coupling between X and Z, the one needed to get $p_{1,r}$. It also offers a clear link with Laplace transforms and consequently opens the door to their various tools associated with them. To make the connection with extreme value theory, we propose the following definition.

W-class definition : *Let $X \sim G$ and $Z \sim F$ be two random variables with the same support. We will say that they belong to the W-class if the random variable $W = -\log G(Z)$ is Weibull distributed with parameters (k, λ), i.e.,*

$$\mathbb{P}(W > w) = \exp\left(-(w/\lambda)^k\right), \qquad \text{for any } w > 0.$$

A first question is to wonder if either this W-class is empty or contains useful members. The following lemma answers positively to this inquiry (below, μ, σ, ξ refer to the location, scale and shape parameters of a Generalized Extreme Value distribution)

Lemma 1 *If $X \sim GEV(\mu_X, \sigma_X, \xi_X)$ and $Z \sim GEV(\mu_Z, \sigma_Z, \xi_Z)$ have the same support, then (X, Z) belongs to the W-class with parameters*

$$\begin{cases} k = \sigma_X/\sigma_Z \text{ and } \lambda = \exp\left((\mu_X - \mu_Z)/\sigma_X\right) & \text{, if } \xi_X = \xi_Z = 0, \\ k = \xi_X/\xi_Z \text{ and } \lambda = \left(k \times \sigma_Z/\sigma_X\right)^{-1/\xi_X} & \text{, if } \xi_X \xi_Z > 0. \end{cases}$$

A few remarks concerning this lemma can be made.

Remark 1 *To avoid confusion, we recall that the term "Weibull" in our W-class definition corresponds to the common denomination used in the engineering/reliability literature, not to the one used by EVT community. The "Weibull" name often corresponds to negative GEV shape parameters in EVT. Still, this case is covered by our Lemma when $\xi_X < 0$ and $\xi_Z < 0$.*

Remark 2 *Lemma 1 clearly indicates that the stringent assumption of equal GEV shape parameters used in Naveau et al. (2018) is not needed anymore. Instead, it has been replaced by a much weaker constraint: factual and counterfactual shape parameters have to have the same sign or be both equal to zero. This covers a wide variety of applied cases. In addition, if the factual shape parameter has a different sign than the counterfactual one, it means that contrasting both worlds becomes trivial, and no fancy statistics is needed then.*

Remark 3 *The assumption of a common support appears reasonable in most EEA studies. Variables with positive (or zero) shape parameters like extreme precipitation have the same support, e.g.* $(0, \infty)$ *for heavy rainfall. Variables with negative shape parameters like extreme temperatures have to be treated with more care. Having a negative* ξ *implies a finite upper bound that depends on* μ, σ *and* ξ*. For the GEV setup in Lemma 1, having the same finite upper point translates into imposing one constraint on a set of six GEV parameters* $(\mu_X, \sigma_X, \xi_X, \mu_Z, \sigma_Z, \xi_Z)$*. So, still five degrees of freedoms are available and provide enough flexibility for most applications. Finally, Proposition 6 in Appendix covers the case of different supports of* X *and* Z*. In terms of notations and number of sub cases, this extended version of Lemma 1 is rather cumbersome, and has a rather limited interest for practitioners. So, we will keep the assumption of common support in the rest of this paper.*

Remark 4 *Lemma 1 does not imply that we have to estimate six GEV parameters in our EEA studies of records. The practitioner only needs to check the validity of the Weibull assumption for* W*, not the one about the two GEVs. In terms of inference, only the two parameters* k *and* λ*, instead of six for the GEV's, need to be estimated. Overall, these elements represent an important gain in terms of models evaluation, computational time, statistical efficiency and ease of interpretation. To highlight this later point, we need to explain how to interpret* λ *and* k *for large* r*.*

As the Laplace transform of $W \sim \text{Weibull}(k, \lambda)$ is known, either as a finite integral or an infinite sum, the probability $p_{1,r}$ can be written for any (X, Z) in the W-class

$$p_{1,r} = \int_0^1 \exp(-(r-1)\lambda(-\log x)^{1/k})\, dx. \tag{6}$$

This allows us to provide the asymptotic expressions of $p_{1,r}$ and $far(r)$ for large r.

Proposition 2 *For any (X, Z) belonging to the W-class, we have*

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

and consequently,

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

This proposition highlights the fundamental role of k that compares the factual and counterfactual tail behaviors. The parameter λ is a second order parameter in terms of tail behavior in the sense that it characterizes scale changes, in particular when the factual and counterfactual shape parameters are equal.

3.2. Non-parametric inference under the W-class

Proposition 1 gave us general expressions of asymptotic variances of the non-parametric estimator $\widehat{p}_{1,r}$. Under the W-class family, it is possible, for large r, to provide precise approximations of such variances and the associated relative errors[§].

Proposition 3 *For any (X, Z) belonging to the W-class, the estimator $\widehat{p}_{1,r}$ under the conditions of Proposition 1 satisfies*

1. $\sigma_r^2 = \left(\frac{c(\lambda,k)/2^k}{r^k} + a\frac{k(c(\lambda,k))^2 - d(\lambda,k)}{r^{2k-1}} \right)(1 + o(1))$,
2. *the relative error of $\widehat{p}_{1,r}$ is of order $r^{\max(k,1)/2}$,*
3. *the relative error of $\widehat{far}(r)$ is of order*

$$\begin{cases} r^{k/2} & \text{, if } k > 1 \text{ or, if } k = 1 \text{ and } \lambda \neq 1, \\ r^{k-1/2} & \text{, if } k < 1, \end{cases}$$

[§]Note that the relative error of $\widehat{far}(r)$ makes no sense in the case $k = \lambda = 1$, since then we have exactly $p_{1,r} = 1/r$ and $far(r) = 0$.

where $c(\lambda,k)=k\Gamma(k)/\lambda^k,\ \tilde{c}(\lambda,k)=k\Gamma(2k)/\lambda^{2k},$ *and* $d(\lambda,k)=2^{1-2k}k\tilde{c}(\lambda,k)$.

As mentioned at the end of Section 2, this proposition reinforces the fact that, besides the rare case of $0<k\leq .5$, the relative errors of non-parametric estimators like $\widehat{p}_{1,r}$ and $\widehat{far}(r)$ are not made to handle large values of r. These non-parametric estimators should be only kept for small values of r. New parametric estimators that fully take advantage of the W-class are needed for large r.

3.3. Parametric inference under the W-class

Under the W-class, Equation (6) gives us a natural expression of a parametric estimator of $p_{1,r}$

$$\widehat{p}_{1,r}^{(W)}=\int_0^1 \exp\left(-(r-1)\widehat{\lambda}(-\log x)^{1/\widehat{k}}\right)\,dx \tag{7}$$

where $\widehat{\lambda}$ and $\widehat{k}$ denote any estimators of the parameters of $W=-\log G(Z)\sim \text{Weibull}(k,\lambda)$. Different estimators of Weibull parameters can be found into the literature. Here, we opt for a straightforward method-of-moment (MOM) approach because easy connections can be made between moments of $G(Z)$ and $p_{1,r}$. In particular, any bivariate random vector belonging to the W-class satisfies the system of two equations

$$\begin{cases} p_{1,2}=\mathbb{E}(G(Z))= & \int_0^1 \exp(-\lambda(-\log x)^{1/k})dx, \\ p_{1,3}=\mathbb{E}(G^2(Z))= & \int_0^1 \exp(-2\lambda(-\log x)^{1/k})dx. \end{cases} \tag{8}$$

From estimates of $p_{1,2}$ and $p_{1,3}$, it is possible to numerically infer the unknown bivariate positive vector $\theta=(\lambda,k)^T$. Then, applying (7) with $\widehat{\theta}=(\widehat{\lambda},\widehat{k})^T$ provides estimates of $p_{1,r}$ for any $r\geq 3$, in particular large ones. Our MOM can be be summarized by a three-step

approach

$$(\widehat{p}_{1,2},\widehat{p}_{1,3}) = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_m(Z_i), \frac{1}{n}\sum_{i=1}^{n}\widehat{G}_m^2(Z_i)\right) \;\rightsquigarrow\; (\widehat{\lambda},\widehat{k}) \;\rightsquigarrow\; \widehat{p}_{1,r}^{(W)}, \text{ via (7) for } r=3,4,\ldots.$$

The following proposition provides the conditions under which, for any fixed $r \geq 3$, the re-normalized estimators $\widehat{p}_{1,r}^{(W)}$ and $\widehat{far}^{(W)}(r) = 1 - \left(r\widehat{p}_{1,r}^{(W)}\right)^{-1}$ converge asymptotically in distribution towards a Gaussian law as the sample sizes of the factual and counterfactual worlds, n and m, go to infinity with $\lim\sqrt{n/m} = a \in [0,\infty)$.

Proposition 4 *For any integer $j \geq 1$ and $\theta = (\lambda,k)^T$, let $J_j(\theta)$ denote the Jacobian matrix of $g_j(\theta) = \int_0^1 \exp(-j\lambda(-\log x)^{1/k})\,dx$.*
Under the conditions of Proposition 1 (in particular $\lim\sqrt{n/m} = a \geq 0$), for any (X,Z) belonging to the W-class, and any fixed integer $r \geq 3$, we have :

(*i*) *the MOM estimator $(\hat{\lambda},\hat{k})$ is asymptotically gaussian in the sense that*

$$\sqrt{n}\left((\hat{\lambda},\hat{k})^T - (\lambda,k)^T\right) \xrightarrow{d} \mathcal{N}\left(0, (J_{1,2}(\theta))^{-1}\,\Sigma_a\,(J_{1,2}^T(\theta))^{-1}\right)$$

where $J_{1,2}(\theta)$ denotes the Jacobian matrix associated with $\theta \mapsto (g_1(\theta),g_2(\theta))^T$, and

$$\Sigma_a = \begin{pmatrix} aM_2 + p_{1,3} - (1+a)p_{1,2}^2 & 2aE_{1,2} + p_{1,4} - (1+2a)p_{1,2}p_{1,3} \\ 2aE_{1,2} + p_{1,4} - (1+2a)p_{1,2}p_{1,3} & 4aM_3 + p_{1,5} - (1+4a)p_{1,3}^2 \end{pmatrix}$$

with $E_{1,2} = \mathbb{E}\left(G(Z_1)\min(G(Z_1),G(Z_2))\right)$.

(*ii*) *the estimators $\widehat{p}_{1,r}^{(W)}$ and $\widehat{far}^{(W)}(r)$ satisfy*

$$\sqrt{n}\frac{\widehat{p}_{1,r}^{(W)} - p_{1,r}}{\sigma_r^{(W)}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{and} \quad \sqrt{n}\,rp_{1,r}^2\frac{\widehat{far}^{(W)}(r) - far(r)}{\sigma_r^{(W)}} \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\sigma_r^{(W)} = \sqrt{J_{r-1}(\theta)(J_{1,2}(\theta))^{-1}\,\Sigma_a\,(J_{1,2}^T(\theta))^{-1}\,(J_{r-1}(\theta))^T}.$$

Note that the matrix Σ_a is the limiting covariance matrix of $\sqrt{n}\left((\widehat{p}_{1,2},\widehat{p}_{1,3})^T-(p_{1,2},p_{1,3})^T\right)$. If G was known, this matrix would simply be

$$\begin{pmatrix} p_{1,3}-p_{1,2}^2 & p_{1,4}-p_{1,2}p_{1,3} \\ p_{1,4}-p_{1,2}p_{1,3} & p_{1,5}-p_{1,3}^2 \end{pmatrix}.$$

As in Proposition 1, the extra terms involving M_2, M_3 and $E_{1,2}$ indicate the cost of estimating G. This additional variability can vanish when m is much larger than n, i.e., a becomes small.

In Proposition 3, the asymptotic relative error of $\widehat{p}_{1,r}$ was found to have a polynomial form. The following proposition shows that the asymptotic relative error of $\widehat{p}_{1,r}$ has a logarithmic rate, and consequently, it should be favored for all $k > 1/2$.

Proposition 5 *Under the conditions stated in Proposition 4, we have, for large r :*

1. *the asymptotic standard deviation $\sigma_r^{(W)}$ in $\sqrt{n}(\widehat{p}_{1,r}^{(W)}-p_{1,r})$ is of order $\log(r)/r^k$,*
2. *the relative error of $\widehat{p}_{1,r}^{(W)}$ is of order $\log(r)$,*
3. *the relative error of $\widehat{far}^{(W)}(r)$ is of order*

$$\begin{cases} \log(r) & \text{, if } k>1 \text{ or, if } k=1 \text{ and } \lambda\neq 1, \\ r^{k-1}\log(r)\to 0 & \text{, if } k<1. \end{cases}$$

A sharper approximation for $\sigma_r^{(W)}$ can be found in the proof of Proposition 5. Note that the relative error of $far(r)$ in Naveau *et al.* (2018) was found to be constant. This is not the case here, even for $k = 1$. This is due to the fact that the step of estimating the unknown k, even if it is equal to one, brings a relative error that grows in $log(r)$.

3.4. Confidence intervals within the W-class

Proposition 4 contains the main ingredient for constructing asymptotic confidence interval estimates of $p_{1,r}$ and $far(r)$. In practice, we can compute two confidence interval types:

either on the original scale or on a logarithmic scale. Extensive simulations (available upon request) clearly indicate that coverage probabilities and mean square errors turn out to be superior on the log-transformed scale. For this reason, the later approach will be used in our simulation study and our application. More precisely, in this work, the asymptotic confidence intervals for $\widehat{p}_{1,r}^{(W)}$ will always be

$$\left[\widehat{p}_{1,r}^{(W)} \times \exp\left(\pm \frac{z_\alpha}{\sqrt{n}} \frac{\widehat{\sigma}_r^{(W)}}{\widehat{p}_{1,r}^{(W)}}\right)\right], \tag{9}$$

where z_α denotes the quantile of order $1-\alpha/2$ of the standard normal distribution. For $\widehat{far}^{(W)}(r)$, one possibility is to use the following confidence interval (equivalent to (9) by definition of $\widehat{far}^{(W)}(r)$)

$$\left[1 - \frac{1}{r\widehat{p}_{1,r}^{(W)}} \exp\left(\mp \frac{z_\alpha}{\sqrt{n}} \frac{\widehat{\sigma}_r^{(W)}}{\widehat{p}_{1,r}^{(W)}}\right)\right] \tag{10}$$

and another one, only available when $\widehat{far}^{(W)}(r)$ is positive, is the confidence interval on the logarithmic scale (issued from the asymptotic normality of $\log \widehat{far}^{(W)}(r)$)

$$\left[\widehat{far}^{(W)}(r). \exp\left(\pm \frac{z_\alpha \widehat{\sigma}_r^{(W)}}{\sqrt{n}\left(\widehat{p}_{1,r}^{(W)}(r\widehat{p}_{1,r}^{(W)} - 1)\right)}\right)\right]. \tag{11}$$

Simulations, see Section 4, indicate that both intervals perform well, particularly when r is large. For conciseness, (10) will be the one considered in our case study.

3.5. Checking if (X, Z) belongs to the W-class

The random variable $W = -\log G(Z)$ is at the center of our W-class definition and its Weibullity a key element. There exist various Weibull goodness-of-fit (WGOF hereafter) tests in the literature (see, e.g. Cabaña and Quiroz, 2005) which are coded with the `R` programming language (see, e.g., the EWGoF package Krit, 2019). Leveraging this package,

we focus here on one of them, the likelihood-based WGOF test implemented in the `WLK.test` function in `R`, which consists in nesting the two-parameter Weibull family in the larger family of Generalized Gamma distributions, and performing a score test of equality to 1 of some shape parameter, (see, e.g. Krit *et al.*, 2016, for mathematical details). This test uses Monte Carlo simulation for determining the p-values.

At first sight, using `WLK.test` for validating our model should be straightforward. In practice, we have at your disposal the factual sample, $(Z_1, \ldots, Z_n)^T$ and the counterfactual sample, $(X_1, \ldots, X_m)^T$. But, we cannot directly build $W_i = -\log G(Z_i)$ because the cdf G is unknown and needs to be estimated. A simple solution to this issue is to introduce "pseudo-Weibull" variables defined by

$$\widehat{W}_i = -\log \widehat{G}_m(Z_i)$$

where $\widehat{G}_m(.)$ is defined by (4). As taking the log of zero is impossible, $\widehat{G}_m(.)$ has to be always positive. This explains the constant b in the definition[¶] of $\widehat{G}_m(.)$, see Equation (4). Still, even when W_i is Weibull distributed, $\widehat{W}_i$ cannot follow exactly a Weibull distribution, and thus blind application of `WLK.test` to these pseudo-Weibull variables will be misleading.

To pinpoint the issue, we rewrite $\widehat{W}_i = -\log \widehat{G}_m(Z_i)$ as $\widehat{W}_i = -\log \mathbb{U}_m(e^{-W_i})$ where

$$\mathbb{U}_m(x) \stackrel{d}{=} \frac{1}{m+1}\left(\sum_{j=1}^{m}(\mathbb{I}_{U_j \leq x}) + b\right),$$

and $U_1, \ldots, U_m$ are independent uniform variables on $[0, 1]$. With this setup, we can easily simulate 1000 times Weibull samples (W_i) of size n for any given parameters (λ, k) and 1000 uniform samples (U_j) of size m, and then apply the function `WLK.test` to the 1000 samples of pseudo-Weibull samples of $\widehat{W}_i = -\log \mathbb{U}_m(e^{-W_i})$. For example, for $n = m = 100$ and $(\lambda, k) = (0.4, 0.8)$, instead of being uniformly distributed, the obtained 1000 p-values

[¶]We opt for the value $b = 0.05$ in our study.

are very skewed to the left with around 25% which were lower than 0.05, the expected mean fraction 5%. This example illustrates that the WLK test, applied to those pseudo-Weibull variables, reject much too often the Weibull assumption at the nominal risk 5%, and this is due to the influence of this transformation $\mathbb{U}_m$. This inflated rejection rate may be a bit milder when m is much larger than n, but this issue definitely needs to be addressed in practice.

Our procedure for determining correct p-values for the WGOF test is the following one :

1. Compute the moment estimates $(\widehat{\lambda}, \widehat{k})$ as described in Section 3.3.
2. Repeat for each $j \in \{1, \ldots, N\}$ with N a prefixed number of replicas (we advise at least $N = 500$ for precise p-value evaluation) :

   (a) simulate n iid $W_i^{(j)} \sim \text{Weibull}(\widehat{\lambda}, \widehat{k})$,

   (b) simulate an uniform sample of size m and denote by $\mathbb{U}_m^{(j)}$ its associated empirical function (see above),

   (c) compute $\widehat{W}_i^{(j)} = -\log \mathbb{U}_m^{(j)}(\exp(-W_i^{(j)}))$ for each $i = 1, \ldots, n$

   (d) compute the test statistic value from `WLK.test`, say $T^{(j)}$, associated to these pseudo-Weibull variables.

3. Approximate the test statistic p-value as the proportion $\frac{1}{N}\sum_{j=1}^{N} \mathbb{I}_{T^{(j)}>T}$ where T corresponds to the test statistic value inferred from the original pseudo-Weibull $\widehat{W}_i$.

Simulations not reported here (available on demand) indicate that this procedure provides adequate p-values for the WGOF test considered here.

4. SIMULATIONS

In this section dedicated to numerical simulations[‖] assessments in `R`, we take $X \sim GEV(\mu_X, \sigma_X, \xi_X)$ and $Z \sim GEV(\mu_Z, \sigma_Z, \xi_Z)$, i.e. (X, Z) belongs to the W-class. The shape parameter ξ_X will span eight values in the set $\{-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4\}$, the parameter $k = \xi_X/\xi_Z$ will cover the seven values in $\{4/4, 4/5, 4/6, 4/7, 4/8, 4/9, 4/10\}$, and the ratio σ_Z/σ_X will belong to the set $\{1, 2, 2.5\}$ with the fixed value $\sigma_X = 1$. From Lemma 1, we know that the Weibull scale parameter is equal to $\lambda = (k \times \sigma_Z/\sigma_X)^{-1/\xi_X}$. Hence, our simulation setup covers $8 \times 7 \times 3 = 168$ different settings of (k, λ) and spans most of the cases encountered in climate EEA. To mimic our case study in Section 5, we set $n = 30$ and $m = 150$. These factual and counterfactual sample sizes are rather small, but are typical of climate applications. Note also that

- the case $\xi_X = \xi_Z = 0$, was covered in Naveau *et al.* (2018),
- the location GEV parameters have been set to insure the common support constraint in Lemma 1,
- all values of k are less or equal to one in our 168 settings. According to Proposition 5, the relative error of $\widehat{far}^{(W)}(r)$ rapidly goes to zero when $k > 1$ for large r. So, the case $k > 1$ is trivial from a inferential point of view and it is also uninformative in terms of causality interpretation (see, e.g., Hannart *et al.*, 2016; Hannart and Naveau, 2018). If $k > 1$, *necessary* causation probability defined as $\max(0, far(r))$ quickly becomes null for large r, and the degree of evidencing is only contained in *sufficient* causation probability defined by $\max(0, 1 - (1 - p_{1,r})/(1 - p_{0,r}))$. Finally, for most EEA studies, we expect an increase in extremal tail behaviors, i.e. $k < 1$. For the few cases where the opposite is expected (e.g., snow covers), then our definition of $p_{1,r}$ could be changed. Instead of comparing $\mathbb{P}(Z_r > \max\{X_1, \ldots, X_{r-1}\})$ to $\mathbb{P}(X_r > \max\{X_1, \ldots, X_{r-1}\})$, it will make

[‖]Some R code (and a manual) for applying the contents of this paper can be found on the second author's webpage.

more sense to compare $\mathbb{P}(X_r > \max\{Z_1, \ldots, Z_{r-1}\})$ to $\mathbb{P}(Z_r > \max\{Z_1, \ldots, Z_{r-1}\})$. This study of the probability of observing a counterfactual record with respect to the factual world will imply a ratio of tail indices smaller than one.

Our main tool to assess the quality of our asymptotic intervals proposed in Section 3.4 is the coverage probability obtained by computing the intervals defined by equations (9) and (10) over 1000 replicas for each of our 168 cases.

To summarize our results for the cases with $\sigma_X = \sigma_Z = 1$, Figure 1 displays the 95% coverage probability of $p_{1,r}$ with $r = 10$ and estimated either from (3), see panel (a), or (7), see panel (b). The x-axis correspond to the value of ξ_X and the y-axis represents the inverse of k. A green color indicates a good performance, roughly between 93.5% and 96.5%, a purple/blue color a conservative interval (coverage larger than 98%), a yellow color a coverage around 90%, and orange and red colors represent worse coverages. Each scalar in the cell provides the value of the coverage. The difference between the two panels shows, as expected, that the parametric approach, see panel (a) obtained with (7), outperforms its non-parametric version, see panel (b) obtained with (9). Although the parametric approach appears to provide good coverage values in most cases, panel (b) also indicates that, when the factual shape parameter increases by more than 50%, i.e. when ξ_Z is 1.5 larger than ξ_X, the coverage appears to be too high for slightly negative ξ_X. Note the factual sample size $n = 30$ is small and this issue progressively disappears as n increases.

[Figure 1 about here.]

Concerning the confidence intervals of $far(r)$, we apply formula (10) to the parametric estimator $\widehat{far}^{(W)}(r)$ for $r = 20$. Comparing the three coverage probability maps in Figure 2 stresses the influence of the scale parameters. This time, each cell contains the value of $\lambda = ((\xi_X \times \sigma_Z)/(\xi_Z \times \sigma_X))^{-1/\xi_X}$, while the color still indicates the coverage probability. The key message from this figure is that λ is the main driver of the coverage accuracy. If λ is very large, then the coverage is underestimated, see yellow cells. In contrast, the blue cells,

overestimation of coverage probabilities, correspond to very small λ's close to zero. Overall, when λ stays in the range $[0.3, 10]$, color maps are mostly greenish indicating a good coverage probability. In the upcoming section, we will see that this situation will correspond to our application setup.

[Figure 2 about here.]

5. ANALYSIS OF ANNUAL DAILY TEMPERATURES RECORDS

In this section, we analyze yearly maxima of daily maxima of near-surface air temperature issued from the numerical climate model CNRM-CM5 of Météo-France, a participant of the CMIP5 intercomparison project. The whole world surface is represented on a 256×128 grid, each cell roughy represents a 150km$\times$150km area. Our factual dataset corresponds to the recent climatology, 1975-2005 (so $n = 31$), of an "all forcings" run (natural and anthropogenic forcings) spanning the preindustrial times (year 1850) to the year 2005. Our counterfactual data are issued from a second run with only natural forcings enabled over the period 1850-2012 (so $m = 163$). Thus, the methodology described in Section 3.1 can be applied to both factual and counterfactual time series available at each of $N = 256 \times 128 = 32,768$ gridpoints.

To assess which (X, Z) belongs to the W-class, the likelihood-based test (WLK) presented in Section 3.5 is applied to each grid point. This test at the nominal risk of 5% rejects the Weibull assumption at 6.3% of all locations. For comparison, the exponentiality assumption is rejected by the Cox-Oakes test (at risk 5%) in 27.5% of the cases, mostly located over oceans. To have a global view, Figure 3 shows the p-values of the WLK where the white zones correspond to rejecting the Weibull hypothesis. No clear spatial clusters prone to rejection can be identified over land, while small structured patches appear over ocean, but without large scale climatological patterns.

[Figure 3 about here.]

The estimated Weibull shape parameter, $\widehat{k}$ from (8), that captures the main changes in tail behavior in the W-class, is plotted in the upper panel of Figure 4. To complement this visual inspection of the fact that many k's differ from one, the hypothesis H_0: $k \geq 1$ against the alternative $k < 1$ is tested using the asymptotic normality of $\widehat{k}$ at each grid point. Around 26.5% of the gridpoints reject (at risk 5%) H_0: $k \geq 1$. This percentage is rather large, in particular by taking in consideration that the statistical power is expected to be moderate for a small sample of $n = 31$. Panel (b) of Figure 4 complements this picture by providing the map of estimated Weibull scale parameter. The red and yellow colors pinpoint regions where the factual world strongly differs from the counterfactual in terms of $\widehat{\lambda}$. In particular, a strong contrast for this parameter can be seen around the Greenland sea, the tropical band and parts of southern ocean. Note that $\widehat{\lambda}$ and $\widehat{k}$ are moderately correlated (Pearson correlation coefficient is 0.52), zones with low $\widehat{\lambda}$ more or less correspond to zones with low $\widehat{k}$.

[Figure 4 about here.]

We now focus on identifying changes in records. As the factual sample size $n = 31$ is small, we illustrate our approach with decadal records fixing, *i.e.* $r = 10$. Figure 5(a) displays the values $\widehat{p}_{1,10}^{(W)}$, *i.e.* the estimated probabilities that the factual world produces a temperature record with respect to the previous 9 counterfactual years. Again, the Greenland sea, the tropical band and parts of southern ocean appear to be the regions where an increase in decadal records can be identified. This is confirmed by the estimated map of $far(10)$ in panel (b).

[Figure 5 about here.]

Finally, Figure 6 illustrates that our approach can bring information about records beyond the observable sample size of $n = 31$. Here, we have randomly chosen the grid point corresponding to south Florida. The two lower panels show the interval estimations (based on

(9) and (10)) of $p_{1,r}$ and $far(r)$ for the values $r = 5, 10, 20, 30, 50$ (in the lower left panel, the red crosses are located at the $1/r$ values, the probability of a record in the counter-factual world). The histogram and qqplot in the upper two panels indicate that (X, Z) seems to belong to the W-class (the p-value for the WLK test at this gridpoint is 0.78). The estimate of k is 0.77, with a 95% confidence interval $[0.47, 1.08]$.

[Figure 6 about here.]

6. CONCLUSION AND DISCUSSIONS

To summarize our work, we propose a simple, fast and efficient approach to analyze records in EEA studies. The simplicity comes from the idea of checking if W_i follows a two parameter Weibull distribution. This avoids to test if both factual and counterfactual world are EVT distributed and we completely bypass the estimation of EVT parameters for each dataset. Still, our approach does not require the assumption of the equal shape parameters in the two worlds. The "efficiency" of our approach comes from our propositions 2, 3, 4 and 5 that detail the asymptotic properties of both non-parametric and parametric of all quantities of interest. For small sample sizes $n = 30$, our simulation study indicates that coverage probabilities are meaningful for a wide range of setup, in particular the ones close to our application.

Still, different shortcomings of our paper can be identified and discussed. In our application, all grids points are treated independently and consequently, spatial patterns of p-values in Figure 3 should not be interpreted as such. An unknown proportion of gridpoints showing rejection (at risk 5%) of the Weibull assumption corresponds to a wrong decision: *i.e.* the null hypothesis is erroneously rejected because no multiplicity correction test was used**. This means that the map may provide a more pessimistic image of the Weibull fit. The issue

**a direct application of the Benjamini-Hochberg correction for FDR (false discovery rate), (see, e.g. Dudoit *et al.*, 2003), drops the proportion to 12.2% of the sites for which H_0: $k \geq 1$ is rejected at 5% risk.

of multiple comparison/multiple testing is a delicate one, since it is not clear which type of spatial dependence can and should be taken into account for appropriately correcting the computed p-values. This topic can be explored in future research. This leads to a second and related point. Kiriliouk and Naveau (2020) recently leveraged both multivariate extreme value theory and counterfactual theory to assess causality into a multivariate framework. But, they did not study records and it would be of interest to extend the idea of the W-class into a multivariate setup. This could help modeling spatial dependence, and consequently improve our treatment of multiplicity correction test.

7. PROOFS AND COMPUTATIONAL DETAILS

This section contains the proofs of the various results of this paper, with the exception of Lemma 1, which is proved in a more general form in the Appendix (subsection .4). Proof of Proposition 1 is omitted, because it is only a small modification of the proof provided in Naveau *et al.* (2018), and it has similarities with the first part of the proof of Proposition 4 (there, the reader will see where the data sizes condition $\sqrt{n/m} \to a$ plays its role).

7.1. Proof of Propositions 2 and 3

Let us start by proving the development of $p_{1,r}$, *i.e.* Proposition 2. Thanks to the fact that (X, Z) belongs to the W-class, we write

$$\begin{aligned} p_{1,r} &= \mathbb{E}\left(G^{r-1}(Z_1)\right) = \mathbb{E}\left(e^{-(r-1)(-\log G(Z_1))}\right) \\ &= \int_0^\infty \exp(-(r-1)w)\,\frac{k}{\lambda^k}\,w^{k-1}\exp(-(w/\lambda)^k)\,dw \\ &= \frac{k}{((r-1)\lambda)^k}\int_0^\infty x^{k-1}e^{-x-t_r(x)}\,dx \end{aligned}$$

where $t_r(x) = \{x/((r-1)\lambda)\}^k$ is close to 0 when r is large (and λ is fixed). Writing $e^{-t_r} = 1 - t_r + (e^{-t_r} - 1 + t_r)$ (with the term in brackets being between 0 and $t_r^2/2$), it comes

$$\begin{aligned} p_{1,r} &= \frac{k}{((r-1)\lambda)^k}\left\{\int_0^\infty x^{k-1}e^{-x}\,dx \;-\; \frac{1}{((r-1)\lambda)^k}\int_0^\infty x^{2k-1}e^{-x}\,dx \;+\; R_r\right\} \\ &= \frac{c(\lambda,k)}{(r-1)^k} - \frac{\tilde{c}(\lambda,k)}{(r-1)^{2k}} + O(r^{-3k}) \quad \text{where} \quad \left\{\begin{array}{rcl} c(\lambda,k) &=& k\Gamma(k)/\lambda^k \\ \tilde{c}(\lambda,k) &=& k\Gamma(2k)/\lambda^{2k} \end{array}\right. \end{aligned}$$

since the (positive) remainder term R_r of the first line is smaller than $\Gamma(3k)/\{2((r-1)\lambda)^{2k}\}$. Consequences in terms of $far(r)$ are straightforward (since $far(r) = 1 - 1/(rp_{1,r}) = 1 - c(\lambda,k)^{-1}r^{k-1}(1+o(1))$, with $c(\lambda,k) > 0$ and $c(\lambda,1) = \frac{1}{\lambda}$), so Proposition 2 is proved.

Let us now turn our attention to Proposition 3, and first on its Statement 1, the description (for large r) of the asymptotic variance $\hat{\sigma}_r^2$ of $\sqrt{n}(\widehat{p}_{1,r} - p_{1,r})$. This variance equals

$$\sigma_r^2 = \tau_r^2 + a.(r-1)^2(M_r - p_{1,r}^2),$$

where we have

$$\tau_r^2 = p_{1,2r-1} - p_{1,r}^2 = \frac{c(\lambda,k)/2^k}{(r-1)^k}(1+O(r^{-k})) - \frac{c(\lambda,k)^2}{(r-1)^{2k}}(1+O(r^{-k}))^2 = \frac{c(\lambda,k)/2^k}{r^k}(1+O(r^{-\min(1,k)})).$$

We keep this result in mind and focus now on the study of the term M_r (and on the difference $M_r - p_{1,r}^2$). Setting $W_i = -\log(G(Z_i))$, and $r' = r-1$, $r'' = r-2$, we have

$$\begin{aligned} M_r &= \mathbb{E}\left(G^{r-2}(Z_1)G^{r-2}(Z_2)(\min(G(Z_1),G(Z_2)))\right) \\ &= 2\mathbb{E}\left(G^{r-1}(Z_1)G^{r-2}(Z_2)\mathbb{I}_{G(Z_1)<G(Z_2)}\right) \\ &= 2\mathbb{E}\left(e^{-r'W_1}e^{-r''W_2}\mathbb{I}_{W_1>W_2}\right) \\ &= \frac{2k^2}{(\lambda^2 r' r'')^k}\int\int x^{k-1}e^{-x-(x/(r'\lambda))^k}u^{k-1}e^{-u-(u/(r''\lambda))^k}\mathbb{I}_{[0,(r''/r')x]}(u)\,du\,dx \end{aligned}$$

Details about the approximation of M_r for large r are rather lengthy, so in this document

we only provide a quick picture of them. In fact, caution is necessary for the passage from the indicator of $[0,(r''/r')x]$ to the indicator of $[0,x]$. If we set $h(x,u)=x^{k-1}e^{-x}u^{k-1}e^{-u}$ and

$$f(x,u)=h(x,u)\mathbb{I}_{[0,x]}(u), \qquad \tilde{f}_r(x,u)=h(x,u)\mathbb{I}_{[0,(r''/r')x]}(u),$$
$$f_r(x,u)=\tilde{f}_r(x,u)e^{-t_r(x,u)}, \qquad t_r(x,u)=(x/(r'\lambda))^k+(u/(r''\lambda))^k,$$

we decompose the double integral in the expression of M_r above as

$$\iint f_r=\iint f-\iint f\,t_r+\iint h\,\delta_r-\iint h\,t_r\,\delta_r+\iint \tilde{f}_r(e^{-t_r}-1+t_r)$$

where $\delta_r(x,u)=\mathbb{I}_{[0,(r''/r')x]}(u)-\mathbb{I}_{[0,x]}(u)=-\mathbb{I}_{[x(1-\epsilon_r),x]}(u)$ with $\epsilon_r=1/(r-1)$. The fourth and fifth terms of this sum will be negligible, but the other first three will combine with the development of $p_{1,r}^2$ (using Proposition 2) to provide the following relation (additional details are omitted)

$$(r-1)^2(M_r-p_{1,r}^2)=\frac{k(c(\lambda,k))^2-d(\lambda,k)}{r^{2k-1}}+o(1/r^{2k-1})$$

where $d(\lambda,k)=2^{1-2k}k^2\Gamma(2k)/\lambda^{2k}$. This ends the proof of Statement 1. Note that when $k=1$, the first term of the right-hand side above is equal to $1/(2\lambda^2 r)$, which coincides with the second term $1/(2\theta^2 r)$ of the formula for the variance λ_r^2 stated in Naveau *et al.* (2018) page 3421.

Statement 2 is directly implied by Statement 1, since the relative error of $\widehat{p}_{1,r}$ is the asymptotic standard deviation of $\sqrt{n}(\widehat{p}_{1,r}-p_{1,r})/p_{1,r}$ which equals the square root of

$$\sigma_r^2/p_{1,r}^2=\left(\frac{c_1}{r^k}+\frac{c_2}{r^{2k-1}}\right)\frac{r^{2k}}{c_1^2}(1+o(1))=\left(c_1^{-1}r^k+c_2c_1^{-2}r\right)(1+o(1))\sim cst.r^{\max(k,1)}$$

where $c_1=c(\lambda,k)/2^k$ and $c_2=a(k(c(\lambda,k))^2-d(\lambda,k))$.

Finally, statement 3 describes the relative error of $\widehat{far}(r)$, which is the asymptotic standard deviation of

$$\begin{aligned}\sqrt{n}\frac{\widehat{far}(r)-far(r)}{far(r)} &= \frac{rp_{1,r}}{rp_{1,r}-1}\sqrt{n}\left(\left(1-\frac{1}{r\widehat{p}_{1,r}}\right)-\left(1-\frac{1}{rp_{1,r}}\right)\right)\\ &= \frac{1+o_{\mathbb{P}}(1)}{p_{1,r}(rp_{1,r}-1)}\sqrt{n}\left(\hat{p}_{1,r}-p_{1,r}\right)\end{aligned}$$

and is thus equal to $\sigma_r/(p_{1,r}(rp_{1,r}-1))$. Setting $c=c(\lambda,k)$ and $\tilde{c}=\tilde{c}(\lambda,k)$, we have for large r (remind that, in the case $k=1$, the exact formula $p_{1,r}=1/(1+(r-1)\lambda)$ is valid)

$$p_{1,r}(rp_{1,r}-1)=c.r^{-k}(c.r^{1-k}+\tilde{c}.r^{1-2k}-1)(1+o(1))\sim\begin{cases}c.r^{-k} & \text{if } k>1,\\ c^2.r^{1-2k} & \text{if } k<1,\\ c.(\lambda^{-1}-1)/r & \text{if } k=1 \text{ and } \lambda\neq 1.\end{cases}$$

Therefore, since $2k-1>k$ whenever $k>1$, we finally have, as announced,

$$\frac{\sigma_r}{p_{1,r}(rp_{1,r}-1)}=\begin{cases}(cst+o(1)).r^{-k/2}.r^{k}=(cst+o(1)).r^{k/2} & \text{if } k>1,\\ (cst+o(1)).r^{1/2-k}.r^{2k-1}=(cst+o(1)).r^{k-1/2} & \text{if } k<1,\\ (cst+o(1)).\sqrt{r} & \text{if } k=1 \text{ and } \lambda\neq 1.\ \square\end{cases}$$

7.2. Proof of Proposition 4

Introducing the important notations $\theta=(\lambda,k)$,

$$g_j(\theta)=g_j(\lambda,k)=\mathbb{E}_\theta[G^j(Z)]=\int_0^1\exp(-j\lambda(-\log x)^{1/k})\,dx$$

and

$$g(\theta)=\begin{pmatrix}g_1(\theta)\\ g_2(\theta)\end{pmatrix}=\begin{pmatrix}\mathbb{E}_\theta[G(Z)]\\ \mathbb{E}_\theta[G^2(Z)]\end{pmatrix},$$

our estimation procedure is to define

$$\widehat{\theta} = (\widehat{\lambda}, \widehat{k}) \text{ solving in } (\lambda, k) \text{ the non-linear system } \begin{cases} \widehat{p}_{1,2} - g_1(\lambda, k) &= 0 \\ \widehat{p}_{1,3} - g_2(\lambda, k) &= 0 \end{cases}$$

and then set

$$\widehat{p}_{1,r}^{(W)} = g_{r-1}(\widehat{\theta}) = g_{r-1}(\widehat{\lambda}, \widehat{k}) = \int_0^1 \exp(-(r-1)\widehat{\lambda}(-\log x)^{1/\widehat{k}})\, dx.$$

The first step of this proof is to obtain the asymptotic normality of the first and second moments sequence $(\widehat{p}_{1,2}, \widehat{p}_{1,3})$, *i.e.* of the sequence

$$\begin{aligned} N_n &:= \sqrt{n}\left(\begin{pmatrix} \widehat{p}_{1,2} \\ \widehat{p}_{1,3} \end{pmatrix} - \begin{pmatrix} p_{1,2} \\ p_{1,3} \end{pmatrix} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \widehat{G}_m(Z_i) - G(Z_i) \\ \widehat{G}_m^2(Z_i) - G^2(Z_i) \end{pmatrix} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} G(Z_i) - \mathbb{E}G(Z) \\ G^2(Z_i) - \mathbb{E}G^2(Z) \end{pmatrix} \end{aligned}$$

The first sum is first transformed to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ 2G(Z_i) \end{pmatrix} .(\widehat{G}_m(Z_i) - G(Z_i)) + \begin{pmatrix} 0 \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\widehat{G}_m(Z_i) - G(Z_i))^2 \end{pmatrix}$$

where the sum in the second term is $O_{\mathbb{P}}(\sqrt{n}/m) = O_{\mathbb{P}}(1/\sqrt{m})$ thanks to the fact that $\widehat{G}_m$ is defined by (4), to the Glivenko-Cantelli theorem, and to the assumption $\sqrt{n/m} \to a$. Concerning the sum in the first term now, thanks to the result (8) in Mason and Zwet (1987), if $\widehat{G}_m^{ecdf}$ is the usual empirical cdf of the counterfactual sequence $X_1, \ldots, X_m$, then there exists some brownian bridge process B on $[0, 1]$, independent of the sequences (X_j) and (Z_i), such that (almost surely)

$$\sqrt{m}(\widehat{G}_m^{ecdf}(z) - G(z)) = B(G(z)) + O\left(\frac{\log m}{\sqrt{m}}\right) \tag{12}$$

uniformly in $z \in \mathbb{R}$; and if $\widehat{G}_m$ is the estimator defined by formula (4), then it differs from the empirical cdf by the negligible term $(b - \widehat{G}_m^{ecdf})/(m+1)$, so statement (12) also holds for $\widehat{G}_m$.

Therefore, assumption $\sqrt{n/m} \to a$ on data sizes allows us to write

$$\begin{aligned} N_n &= \sqrt{\frac{n}{m}} \times \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} B(G(Z_i)) \\ 2G(Z_i).B(G(Z_i)) \end{pmatrix} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} G(Z_i) - \mathbb{E}G(Z) \\ G^2(Z_i) - \mathbb{E}G^2(Z) \end{pmatrix} + o_{\mathbb{P}}(1) \\ &= a.N_{1,n} \; + \; N_{2,n} + o_{\mathbb{P}}(1) \end{aligned}$$

Both terms $N_{1,n}$ and $N_{2,n}$ converge in distribution as $n \to \infty$, and their limits are independent thanks to the independence of the brownian bridge B with the factual world dataset Z. On one hand, the limit in distribution of the second term $N_{2,n}$ is the centered gaussian random vector with covariance matrix described after the statement of Proposition 4. On the other hand, if μ_n denotes the empirical measure of the sample $(G(Z_1), \ldots, G(Z_n))$, then $N_{1,n} = (\int B(u)\, d\mu_n(u), \int 2uB(u)\, d\mu_n(u))$ converges in distribution to the random vector $V = (\int_0^1 B(u)\, d\mu(u); \int_0^1 2uB(u)\, d\mu(u))$ where μ is the distribution of the random variable $G(Z_1)$. By the properties of the brownian bridge, this vector $V = (V_1, V_2)$ is a centered gaussian vector, with

$$\begin{aligned} Var(V_1) &= \mathbb{E}(V_1^2) = \int_0^1 \int_0^1 \mathbb{E}(B(u)B(v))\, d\mu(u) d\mu(v) \\ &= \int_0^1 \int_0^1 (\min(u,v) - uv)\, d\mu(u) d\mu(v) \\ &= \mathbb{E}(\min(G(Z_1), G(Z_2))) - \mathbb{E}(G(Z))^2 \; = \; M_2 - p_{1,2}^2 \end{aligned}$$

and similarly $Var(V_2) = 4M_3 - 4p_{1,3}^2$. It remains to compute

$$\begin{aligned} Cov(V_1, V_2) &= \mathbb{E}(V_1 V_2) = \int_0^1 \int_0^1 2v\mathbb{E}(B(u)B(v))\, d\mu(u) d\mu(v) \\ &= \int_0^1 \int_0^1 2v(\min(u,v) - uv)\, d\mu(u) d\mu(v) \\ &= 2\mathbb{E}(G(Z_1)\min(G(Z_1), G(Z_2))) - 2\mathbb{E}(G(Z))\mathbb{E}(G(Z)^2) \;=\; 2E_{1,2} - 2p_{1,2}p_{1,3}. \end{aligned}$$

We have thus proved that N_n converges in distribution to the centered gaussian distribution with covariance matrix Σ_a.

We are now ready to prove Statement (i) of Proposition 4, *i.e.* that the asymptotic distribution of $\sqrt{n}\left((\widehat{\lambda}, \widehat{k}) - (\lambda, k)\right)$ is the centered gaussian vector with covariance matrix $(J_{1,2}(\theta))^{-1}\,\Sigma_a\,(J_{1,2}^T(\theta))^{-1}$. This is in fact a consequence of the first step (standard M-estimation theory), because

$$\sqrt{n}\left((\widehat{\lambda}, \widehat{k}) - (\lambda, k)\right) = \sqrt{n}\left(g^{-1}(\widehat{p}_{1,2}, \widehat{p}_{1,3}) - g^{-1}(p_{1,2}, p_{1,3})\right)$$

where function $g : (\lambda, k) \mapsto (g_1(\lambda, k), g_2(\lambda, k))$ (which $= (p_{1,2}, p_{1,3})$ under the W-class) is locally bijective, $N_n \xrightarrow{d} \mathcal{N}_2(0, \Sigma_a)$, and $J_{1,2}(\lambda, k).J_{g^{-1}}(g(\lambda, k))$ is the identity matrix of dimension 2. Details on this jacobian matrix are provided at the end of this subsection.

Let us now deal with Statement (ii) of Proposition 4, *i.e.* that $\sqrt{n}(\widehat{p}_{1,r}^{(W)} - p_{1,r})$ converges in distribution to the centered gaussian with standard deviation $\sigma_r^{(W)}$: this is here a consequence of Statement (i) via standard delta-method, because $\sqrt{n}(\widehat{p}_{1,r}^{(W)} - p_{1,r}) = \sqrt{n}\left(g_{r-1}(\widehat{\lambda}, \widehat{k}) - g_{r-1}(\lambda, k)\right)$. Proposition 4 is thus proved, because its last statement (asymptotic normality of $\widehat{far}^{(W)}(r)$) comes from the relation

$$\sqrt{n}(\widehat{far}^{(W)}(r) - far(r)) = -\frac{1}{r}\sqrt{n}\left(\frac{1}{\widehat{p}_{1,r}^{(W)}} - \frac{1}{p_{1,r}}\right)$$

and the delta-method. □

Now, to end this subsection, we provide some details about the expressions of the involved jacobians. For any $j \geq 1$ (values $j = 1$, $j = 2$, and $j = r - 1$ are the ones that appeared in the proof above), we have $g_j(\lambda, k) = \int_0^1 h_j(x)\, dx$ where $h_j(x; \theta) = \exp(-j\lambda(-\log x)^{1/k})$, as well as

$$\begin{aligned}
\frac{dh_j}{d\lambda}(x;\theta) &= (-j)(-\log x)^{1/k} h_j(x;\theta), \\
\frac{d}{dk}(-\log x)^{1/k} &= -\frac{1}{k^2}\log(-\log x).(-\log x)^{1/k}, \\
\frac{dh_j}{dk}(x;\theta) &= \frac{j\lambda}{k^2}\log(-\log x).(-\log x)^{1/k} h_j(x;\theta).
\end{aligned}$$

Therefore we obtain

$$J_{r-1}(\theta) = \left(\frac{dg_{r-1}}{d\lambda}(\lambda,k) \quad \frac{dg_{r-1}}{dk}(\lambda,k) \right) \quad \text{and} \quad J_{1,2}(\theta) = \begin{pmatrix} \frac{dg_1}{d\lambda}(\lambda,k) & \frac{dg_1}{dk}(\lambda,k) \\ \frac{dg_2}{d\lambda}(\lambda,k) & \frac{dg_2}{dk}(\lambda,k) \end{pmatrix},$$

where

$$\begin{aligned}
\frac{dg_j}{d\lambda}(\lambda,k) &= (-j)\int_0^1 (-\log x)^{1/k} \exp(-j\lambda(-\log x)^{1/k})\, dx, \\
\frac{dg_j}{dk}(\lambda,k) &= (j\lambda/k^2)\int_0^1 \log(-\log x).(-\log x)^{1/k} \exp(-j\lambda(-\log x)^{1/k})\, dx.
\end{aligned}$$

Remarks on the way these integrals should be numerically evaluated are provided in the Appendix.

7.3. Proof of Proposition 5

Statements 2 and 3 are direct consequences of Statement 1, relying on the material contained in the proof of statements 2 and 3 of Proposition 3 some lines above. We omit the details and thus just need to prove Statement 1.

First of all, let us note in this section $r' = r - 1$ in order to lighten the notations. The variance $\left(\sigma_r^{(W)}\right)^2$ depends on r only through the jacobian $J_{r-1}(\theta)$, which we study now.

We have $J_{r-1}(\theta) = (\, I_{1,r}\,,\, I_{2,r}\,)$ where

$$\begin{aligned} I_{1,r} &= (-r') \int_0^\infty u^{1/k} \exp(-u - r'\lambda u^{1/k})\, du, \\ I_{2,r} &= \frac{r'\lambda}{k^2} \int_0^\infty \log(u)\, u^{1/k} \exp(-u - r'\lambda u^{1/k})\, du. \end{aligned}$$

Proceeding as we did for the development of $p_{1,r}$ in subsection 7.1, the change of variable $v = r'\lambda.u^{1/k}$ leads to

$$I_{1,r} \div (-r') = \frac{k}{(r'\lambda)^{k+1}} \int_0^\infty v^k e^{-v}\, dv \;-\; \frac{k}{(r'\lambda)^{2k+1}} \int_0^\infty v^{2k} e^{-v}\, dv \;+\; O(r^{-(3k+1)})$$

so

$$I_{1,r} = \frac{-c_1(\lambda,k)}{(r-1)^k} + \frac{\tilde{c}_1(\lambda,k)}{(r-1)^{2k}} + O(r^{-3k}) \;\text{ where }\; \left\{ \begin{array}{rcl} c_1(\lambda,k) &=& k\Gamma(k+1)/\lambda^{k+1} \\ \tilde{c}_1(\lambda,k) &=& k\Gamma(2k+1)/\lambda^{2k+1} \end{array} \right.$$

The treatment of $I_{2,r}$ is a bit more involved. We first have

$$\begin{aligned} I_{2,r} \div (r'\lambda/k^2) &= \int_0^\infty \log\left(\left(\frac{v}{r'\lambda}\right)^k\right) . \frac{v}{r'\lambda} . \exp\left(-(v/(r'\lambda))^k - v\right) \frac{k}{(r'\lambda)^k} v^{k-1}\, dv \\ &= \frac{k^2}{(r'\lambda)^{k+1}} \int_0^\infty v^k e^{-v} e^{-(v/(r'\lambda))^k} \log\left(\frac{v}{r'\lambda}\right)\, dv \\ &= \frac{k^2}{(r'\lambda)^{k+1}} \left\{ \int_0^\infty v^k e^{-v} \log\left(\frac{v}{r'\lambda}\right)\, dv \;-\; \frac{1}{(r'\lambda)^k} \int_0^\infty v^{2k} e^{-v} \log\left(\frac{v}{r'\lambda}\right)\, dv \;+\; O(r^{-2k}) \right\} \end{aligned}$$

Then, noticing that $\int_0^\infty x^{\alpha-1} e^{-x} \log(x)\, dx = \frac{d}{d\alpha} \int_0^\infty x^{\alpha-1} e^{-x}\, dx = \Gamma'(\alpha)$ (where Γ is the Gamma function), we have

$$\begin{aligned} \int_0^\infty v^k e^{-v} \log\left(\frac{v}{r'\lambda}\right)\, dv &= (-\log(r'\lambda)) \int_0^\infty v^k e^{-v}\, dv \;+\; \int_0^\infty v^k e^{-v} \log(v)\, dv \\ &= (-\log(r'\lambda))\Gamma(k+1) + \Gamma'(k+1) \\ &= (-\log(r'\lambda))\Gamma(k+1) \left(1 - \frac{\Psi(k+1)}{\log(r'\lambda)}\right) \\ &= (-\log(r'\lambda))\Gamma(k+1)\,(1 + o(1))\,. \end{aligned}$$

where $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function. Proceeding similarly for the second term of $I_{2,r}$, we obtain

$$\int_0^\infty v^{2k} e^{-v} \log\left(\frac{v}{r'\lambda}\right) dv \;=\; (-\log(r'\lambda))\Gamma(2k+1)\,(1+o(1))$$

(where $o(1)$ refers to the asymptotic in r). Finally, the main part of $I_{2,r}$ is

$$I_{2,r} \;\simeq\; \log((r-1)\lambda)\left\{\frac{-c_2(\lambda,k)}{(r-1)^k}(1-o_1(1)) + \frac{\tilde{c}_2(\lambda,k)}{(r-1)^{2k}}(1-o_2(1))\right\}$$

where $c_2(\lambda,k) = \Gamma(k+1)/\lambda^k$, $\tilde{c}_2(\lambda,k) = \Gamma(2k+1)/\lambda^{2k}$, and $o_i(1) := \Psi(ik+1)/\log((r-1)\lambda)$ $(i=1,2)$ converge slowly to 0 when $r \to \infty$. We notice that the term issued from $o_1(1)$ is of larger order than the term $\tilde{c}_2/(r-1)^{2k}$, so finally there is no need to take the latter into account.

Gathering all the developments obtained so far, and introducing the notations $M(\theta) = (J_{1,2}(\theta))^{-1}\,\Sigma_a\,(J_g^t(\theta))^{-1}$ and $L_r := \log((r-1)\lambda)$, we finally obtain via $\sigma_r^{(W)}$:

$$\begin{aligned}
\sigma^2_{r,par} &= \left(\frac{-c_1}{(r-1)^k} \;\; \frac{-c_2 L_r}{(r-1)^k}\right).M(\theta).\begin{pmatrix} \frac{-c_1}{(r-1)^k} \\ \frac{-c_2 L_r}{(r-1)^k} \end{pmatrix}.(1+o(1)) \\
&= \frac{L_r^2}{(r-1)^{2k}}\left(\frac{c_1}{L_r} \;\; c_2\right).M(\theta).\begin{pmatrix} \frac{c_1}{L_r} \\ c_2 \end{pmatrix}.(1+o(1)) \\
&\sim \left(\frac{L_r}{(r-1)^k}\right)^2 \times (c_2(\lambda,k))^2 \times (M(\theta))_{2,2}
\end{aligned}$$

and therefore Statement 1 is proved. □

Remark 5 *Naturally, if another estimator $(\tilde{\lambda},\tilde{k})$ is considered, which estimated variance can be computed and estimated, then properties of the relative error of the corresponding estimator $\tilde{p}_{1,r}^{(W)} = g_{r-1}(\tilde{\lambda},\tilde{k})$ can be deduced from the contents of the proof above, by replacing the matrix $M(\theta)$ by the appropriate covariance matrix in the formula for $\sigma_r^{(W)}$.*

REFERENCES

Beirlant J, Goegebeur Y, Segers J, Teugels J, 2004. *Statistics of Extremes: Theory and Applications*. Wiley, 489 pp.

Cabaña A, Quiroz A, 2005. Using the empirical moment generating function in testing the weibull and the type i extreme value distributions. *TEST* **14**(2): 417–431.

Coles SG, 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag Inc, 208 pp.

de Haan L, Ferreira A, 2006. *Extreme Value Theory: an Introduction*. Springer-Verlag Inc, 421 pp.

Dudoit S, Shaffer JP, Boldrick J, 2003. Multiple hypothesis testing in microarray experiments. *Statistical Science* **18**(1): 71–103.

Hannart A, Naveau P, 2018. Probabilities of causation of climate changes. *Journal of Climate* **31**(14): 5507–5524.

Hannart A, Pearl J, Otto FEL, Naveau P, Ghil M, 2016. Counterfactual causality theory for the attribution of weather and climate-related events. *Bull. Amer. Meteor. Soc.* **97**(99–110).

Katz RW, Parlange MB, Naveau P, 2002. Statistics of extremes in hydrology. *Advances in water resources* **25**(8-12): 1287–1304.

Kew Sf, Philip SY, Jan van Oldenborgh G, van der Schrier G, Otto FEL, Vautard R, 2019. The exceptional summer heat wave in southern Europe 2017. *Bulletin of the American Meteorological Society* **100**(1): S49–S53.

Kharin VV, Zwiers FW, Zhang X, Hegerl GC, 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate* **20**(8): 1419–1444.

King AD, 2017. Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future* **5**: 1156–1168.

Kiriliouk A, Naveau P, 2020. Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Annals of Applied Statistics* **In press**.

Krit M, 2019. *EWGoF R package reference manual*. https://cran.r-project.org/web//packages/EWGoF/index.html.

Krit M, Gaudoin O, Xie M, Remy E, 2016. Simplified likelihood based goodness-of-fit tests for the weibull distribution. *Communications in Statistics - Simul. and Comp.* **45**(3): 920–951.

Mason DM, Zwet WRV, 1987. A refinement of the kmt inequality for the uniform empirical process. *Ann. Probab.* **15**: 871–884.

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

Naveau P, Ribes A, Zwiers FW, Hannart A, Tuel A, Yiou P, 2018. Revising return periods for record events in a climate event attribution context. *Journal of Climate* **31**(3411–3422).

Pearl J, 2000. *Causality: Models, reasoning, and inference.* Cambridge University Press.

Stott PA, Christidis N, Otto FE, Sun Y, Vanderlinden JP, van Oldenborgh GJ, Vautard R, von Storch H, Walton P, Yiou P, *et al.*, 2016. Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews: Climate Change* **7**(1): 23–41.

Stott PA, Stone DA, Allen MR, 2004. Human contribution to the European heatwave of 2003. *Nature* **432**(7017): 610.

APPENDIX

.4. Proof of Lemma 1

The proposition below is a more general version of Lemma 1. It describes the situations where the two GEV distributions do not necessarily share the same support.

Proposition 6 *Suppose that the X and Z samples are issued from the cdfs G and F of two Generalized Extreme Value distributions $GEV(\mu_X, \sigma_X, \xi_X)$ and $GEV(\mu_Z, \sigma_Z, \xi_Z)$, and suppose that the shape parameters ξ_X and ξ_Z are both null or have the same sign. Consider the (possibly degenerate and non-continuous) random variable*

$$W = -\log G(Z).$$

Note that $\mathbb{P}(W \geq u) = \exp(-(u/\lambda)^k)$ for each $u > 0$.

1. *If $\xi_X = \xi_Z = 0$, then W follows a Weibull distribution. with shape parameter k and scale parameter λ with*

$$k = \frac{\sigma_X}{\sigma_Z} \quad \text{and} \quad \lambda = \exp\left(\frac{\mu_X - \mu_Z}{\sigma_X}\right).$$

2. If ξ_X and ξ_Z are non-null and have the same sign, let

$$\theta_X = \mu_X - \sigma_X/\xi_X \quad , \quad \theta_Z = \mu_Z - \sigma_Z/\xi_Z \quad , \quad \delta = \left| \frac{\xi_Z}{\sigma_Z}(\theta_X - \theta_Z) \right|$$

and

$$k = \frac{\xi_X}{\xi_Z} \quad \textit{and} \quad \lambda = \left(\frac{\xi_X}{\xi_Z} \frac{\sigma_Z}{\sigma_X} \right)^{-1/\xi_X} .$$

(a) In the heavy tail case $\xi_X > 0$ and $\xi_Z > 0$, θ_X and θ_Z are the left endpoints of the distributions G and F and we have :

(H-i) if $\theta_X = \theta_Z$, then W follows the Weibull(λ, k) distribution.

(H-ii) if $\theta_X < \theta_Z$, then $G(Z)$ is never close to 0 and W has its values in the interval $[0, u^*]$ with $u^* = \lambda\delta^{-1/\xi_X}$ and

$$\mathbb{P}(W \leq u) = \begin{cases} 0 & \textit{if} \quad u \leq 0 \\ 1 - \exp\left(-\{-\delta + (u/\lambda)^{-\xi_X}\}^{-1/\xi_Z}\right) & \textit{if} \quad u \in [0, u^*] \\ 1 & \textit{if} \quad u \geq u^* \end{cases}$$

(H-iii) if $\theta_Z < \theta_X$ (improbable case), then $G(Z)$ has a positive probability of being null, and W is degenerate with a mass at $+\infty$.

(b) In the bounded tail case $\xi_X < 0$ *and* $\xi_Z < 0$, θ_X *and* θ_Z *are the right endpoints of the distributions* G *and* F *and we have :*

(B-i) if $\theta_X = \theta_Z$, *then* W *follows the Weibull*(λ, k) *distribution.*

(B-ii) if $\theta_X < \theta_Z$, *then* $Z \geq \theta_X \Leftrightarrow G(Z) = 1 \Leftrightarrow W = 0$, *so* $G(Z)$ *has a mass at* 1, *and* W *has a positive probability of being null, with*

$$\mathbb{P}(W = 0) = 1 - \exp(-\delta^{1/|\xi_Z|})$$

and, for $u > 0$,

$$\mathbb{P}(W \leq u) = 1 - \exp\left(-\{\delta + (u/\lambda)^{|\xi_X|}\}^{1/|\xi_Z|}\right)$$

(B-iii) if $\theta_Z < \theta_X$ *(improbable case), then* $G(Z)$ *is never close to* 1 *and* W *has its values in the interval* $[u^*, +\infty[$ *with* $u^* = \lambda\delta^{1/|\xi_X|}$ *and*

$$\mathbb{P}(W \leq u) = \begin{cases} 0 & \text{if} \quad u \leq u^*, \\ 1 - \exp\left(-\{-\delta + (u/\lambda)^{|\xi_X|}\}^{1/|\xi_Z|}\right) & \text{if} \quad u > u^*. \end{cases}$$

Remark 6 *Note that the described settings, that led to* $W = -\log G(Z)$ *being Weibull, are only examples of frameworks guaranteeing the W-class assumption. Others exist : for instance, when* X *has a Gumbel distribution of cdf* $G(x) = \exp(-\exp(-(x-\mu)/\sigma)$, *and* Z *is equal to* $Z = a + bX$, *it can be easily proved that* $W = -\log G(Z)$ *is Weibull distributed with parameters* $k = 1/b$ *and* $\lambda = \exp((a + \mu(b-1))/\sigma)$.

Remark 7 *The assumption of equal support may seem very restrictive, but we warn the reader that the needed equality is only about the theoretical supports only, and in practice this does not mean that the upper values of the* X *and* Z *samples will lie in the same zone (which would mean that our model is not very realistic). As an illustration (among many others), in the situation* $\xi_X = -0.2$, $\mu_X = 10$, $\sigma_X = 1.5$, $\xi_Z = -0.25$ *and* $\sigma_Z = 1.25$, *the value*

of μ_Z that makes the upper endpoints of F and G to be equal is $\mu_Z = 12.5$, and therefore $\theta_X = \theta_Z = 17.5$. But this equality of upper endpoints does not strike the eye when looking at the two GEV densities, represented in Figure A.1. Note that, for negative tail / shape parameters, the situation $k < 1$ corresponds to ξ_Z being larger than ξ_X.

[Figure A.1 about here.]

Proof of Proposition 6

The proof is tedious, but not complicated. In case 1 ($\xi_X = \xi_Z = 0$), we have

$$G(x) = \exp\left\{-\exp\left(-\frac{x-\mu_X}{\sigma_X}\right)\right\} \quad \text{and} \quad F(z) = \exp\left\{-\exp\left(-\frac{z-\mu_Z}{\sigma_Z}\right)\right\}.$$

The supports of F and G are $\mathbb{R}$, so $W = -\log G(Z)$ is $]0, +\infty[$ valued, with (for $u > 0$)

$$\begin{aligned}
\mathbb{P}(W \leq u) &= \mathbb{P}(Z \geq \mu_X - \sigma_X \log(u)) \\
&= 1 - \exp\left(-\exp\left(\log(u)\frac{\sigma_X}{\sigma_Z} + \frac{\mu_Z - \mu_X}{\sigma_Z}\right)\right) \\
&= 1 - \exp\left(-(u/\lambda)^k\right)
\end{aligned}$$

because $\exp((\mu_Z - \mu_X)/\sigma_Z) = 1/\lambda^k$, where the values of k and λ are given in statement 1 of the Proposition.

Suppose now that we are in case 2, where

$$G(x) = \exp\left\{-\left(1 + \xi_X \frac{x-\mu_X}{\sigma_X}\right)_+^{-1/\xi_X}\right\} \quad \text{and} \quad F(z) = \exp\left\{-\left(1 + \xi_Z \frac{z-\mu_Z}{\sigma_Z}\right)_+^{-1/\xi_Z}\right\}$$

with $u_+ = \max(u, 0)$. The values $\theta_X = \mu_X - \sigma_X/\xi_X$ and $\theta_Z = \mu_Z - \sigma_Z/\xi_Z$ are the respective values at which F and G are not differentiable.

In the heavy tail case $\xi_X > 0$ and $\xi_Z > 0$, the quantities in the inner brackets of the definition of G and F are positive whenever (respectively) $x > \theta_X$ and $z > \theta_Z$. So θ_X and θ_Z are the lower endpoints of G and F, and consequently statement *(H-iii)* for $\theta_Z < \theta_X$ holds

true (*i.e.* $G(Z)$ is null with probability > 0, so W is degenerate with a mass at $+\infty$). We suppose now $\theta_X \leq \theta_Z$, so that W is positive but can never be infinite, and $Z \geq \theta_X$ (*a.s.*). For a given $u > 0$, we have (the second equality is due to $\xi_X > 0$ and $Z \geq \theta_X$)

$$\begin{aligned}
\mathbb{P}(W \leq u) &= \mathbb{P}\left(\left(1 + \frac{\xi_X}{\sigma_X}(Z - \mu_X)\right)_+^{-1/\xi_X} \leq u \right) \\
&= \mathbb{P}\left(Z \geq \mu_X + \frac{\sigma_X}{\xi_X}(u^{-\xi_X} - 1) \right) \\
&= \mathbb{P}\left(\frac{Z - \mu_Z}{\sigma_Z} \geq v(u) \right) = 1 - \exp\left(-(1 + \xi_Z v(u))_+^{-1/\xi_Z} \right)
\end{aligned}$$

where (after some computations)

$$1 + \xi_Z v(u) = \frac{\xi_Z}{\sigma_Z}(\theta_X - \theta_Z) + \frac{\sigma_X}{\sigma_Z}\frac{\xi_Z}{\xi_X}u^{-\xi_X} = -\delta + (u/\lambda)^{-\xi_X}.$$

for the values δ, λ and k stated in the Proposition. Therefore, if $\theta_X = \theta_Z$ (*i.e.* F and G have the same support) then the first term $(-\delta)$ of $1 + \xi_Z v(u)$ vanishes and we have $\mathbb{P}(W \leq u) = 1 - \exp(-(u/\lambda)^k)$, statement *(H-i)* is proved. If $\theta_X < \theta_Z$ (*i.e.* X can take low values that Z cannot), then statement *(H-ii)* follows, W being bounded with the upper endpoint $u^* = \lambda\delta^{-1/\xi_X}$ (this upper endpoint may be rather large when θ_X is not far from θ_Z).

Finally, in the bounded tail case $\xi_X < 0$ and $\xi_Z < 0$, the quantities in the inner brackets of the definition of G and F are positive whenever (respectively) $x < \theta_X$ and $z < \theta_Z$, so now θ_X and θ_Z are the upper endpoints of G and F. Let us first say that in this setting we have

in general

$$\begin{aligned}
\mathbb{P}(W \leq u) &= \mathbb{P}\left(\left(-\frac{|\xi_X|}{\sigma_X}(Z-\theta_X) \right)_+ \leq u^{|\xi_X|} \right) \\
&= 1 - \mathbb{P}\left(\left(-\frac{|\xi_X|}{\sigma_X}(Z-\theta_X) \right) > u^{|\xi_X|} \right) \\
&= \mathbb{P}\left(Z \geq \theta_X - \frac{\sigma_X}{|\xi_X|} u^{|\xi_X|} \right) \\
&= 1 - \exp\left(- \left(\frac{\xi_Z}{\sigma_Z}(\theta_X - \theta_Z) + \frac{\sigma_X}{\sigma_Z}\frac{\xi_Z}{\xi_X} u^{|\xi_X|} \right)_+^{1/|\xi_Z|} \right)
\end{aligned}$$

In the inner bracket of the last line, the second term equals $(u/\lambda)^{|\xi_X|}$, and the first term is zero when $\theta_X = \theta_Z$ (so statement *(B-i)* is proved), is equal to $-\delta$ when $\theta_Z < \theta_X$, and to δ when $\theta_Z > \theta_X$. Let us deal with $\theta_Z > \theta_X$ first. In this case, Z can be greater than θ_X with positive probability, so $G(Z) = 1$ and thus $W = 0$ is a non negligible event : statement *(B-ii)* is thus proved ($\mathbb{P}(W = 0)$ is given by the formula above with $u = 0$). We finally deal with the $\theta_Z < \theta_X$ (improbable in practice) case : statement *(B-iii)* should be clear to the reader, by identifying the value u^* such that the inner bracket above is zero for $u \leq u^*$. Note that this value u^* is rather small when θ_Z is not much greater than θ_X (*i.e.* when the "equal support condition" of assumption (MA) is only slightly violated). □

.5. Computational issues and details

A number of quantities described in this paper require numerical procedures for computing them. Among them :

- various integrals such as $p_{1,r}$, $\widehat{p}_{1,r}^{(W)}$, M_r or $E_{1,2}$ (described in Propositions 1 and 4), or the jacobians $J_j(\lambda, k)$ appearing in Proposition 4 ;
- computation of the estimator $\widehat{\theta} := (\widehat{\lambda}, \widehat{k})$ by solving a non-linear system of 2 equations (involving integrals evoked above).

Appropriate procedures have been coded in R with care so as to minimize numerical uncertainty regarding these issues (see the codes on the second author's webpage). We wanted in this section to point out and explain why the computation of the integrals involved with this work is a particularly sensitive issue.

Let us look for instance at the computation of the estimator $\widehat{p}_{1,r}^{(W)}$. It turns out that, generally, for a moderate or large value of r (and/or some values of the couple $(\hat{\lambda}, \hat{k})$), the function we need to integrate in the expression

$$\widehat{p}_{1,r}^{(W)} := \int_0^1 \exp(-(r-1)\widehat{\lambda}(-\log x)^{1/\widehat{k}})\, dx$$

can be very or *extremely flat* in almost all the interval $]0,1[$, and then grows very or *extremely steeply* in the left neighborhood of 1 : this makes the concrete numerical evaluation of these integrals particularly problematic from the computation accuracy point of view. This is why we used in our R codes an alternative equivalent formula

$$\widehat{p}_{1,r}^{(W)} = \frac{1}{c}\int_0^1 \exp(-r'(-\log u)^{1/\widehat{k}})u^{1/c-1}\, du$$

where $r' = (r-1)\widehat{\lambda}/c^{1/k}$ and the (generally large) real value $c > 0$ is chosen so that the function inside this second integral is less flat, and the integral is thus easier to numerically compute (a possibility is to choose c so that r' is equal to 1) . One may argue that approximations to Laplace transforms of Weibull variables are available in the literature, in the form of series representations : we have however observed that they are particularly unreliable *in our setting*, because of the great variability of situations and values of the parameters involved, and therefore we decided not to use them.

Concerning the other integrals evoked at the start of this subsection (*i.e.* those involved in the formula of the asymptotic variance of $\widehat{p}_{1,r}^{(W)}$), they must also be computed with similar caution. For instance, noting $W_i = -\log G(Z_i)$ and considering the uniform variables V_i such

that $W_i = \lambda(-\log V_i)^{1/k}$, we have

$$\begin{aligned} M_r &= \mathbb{E}\left(G^{r-2}(Z_1)G^{r-2}(Z_2)(\min(G(Z_1),G(Z_2)))\right) = \mathbb{E}\left(e^{-(r-1)(W_1+W_2)}e^{\min(W_1,W_2)}\right) \\ &= 2\mathbb{E}\left(e^{-(r-1)(W_1+W_2)}e^{W_2}\mathbb{I}_{W_1\geq W_2}\right) = 2\mathbb{E}\left(e^{-(r-2)W_2}e^{-(r-1)W_1}\mathbb{I}_{V_1\leq V_2}\right) \\ &= 2\int_0^1 \left\{\exp(-(r-2)\lambda(-\log y)^{1/k})\int_0^y \exp(-(r-1)\lambda(-\log x)^{1/k})\,dx\right\} dy \end{aligned}$$

In this double integral, the "flatness problem" is even more important in the inner integral (in x) than for the evaluation of $p_{1,r}$, since here we only integrate between 0 and y : so for small values of y the integrated function takes extremely small and similar values. Therefore, for computing M_r accurately , we advocate to rely on the following formula, where $r' := (r-2)\lambda/c^{1/k}$ for some value c to be chosen (for instance such that $r' = 1$)

$$M_r = \frac{2}{c}\int_0^1 \exp(-r'(-\log v)^{1/k})\left(\int_0^{v^{1/c}} \exp(-(r-1)\lambda(-\log x)^{1/k})\,dx\right) v^{1/c-1}\,dv.$$

Concerning the value $E_{1,2}$ also appearing in the matrix Σ_a,

$$\begin{aligned} \mathbb{E}\left(G(Z_1)\min(G(Z_1),G(Z_2))\right) &= \mathbb{E}\left(e^{-W_1}e^{-\max(W_1,W_2)}\right) \\ &= \mathbb{E}\left(e^{-2W_1}\mathbb{I}_{W_1\geq W_2}\right) + \mathbb{E}\left(e^{-(W_1+W_2)}\mathbb{I}_{W_1\leq W_2}\right) \\ &= A + \frac{1}{2}p_{1,2}^2 \end{aligned}$$

where, setting $\phi(v) = \mathbb{E}(\exp(-2\lambda(-\log v)^{1/k})\mathbb{I}_{v\leq V_2}) = (1-v)\exp(-2\lambda(-\log v)^{1/k})$, we have

$$\begin{aligned} A &= \mathbb{E}\left(\phi(V_1)\right) = \mathbb{E}\left(\exp(-2\lambda(-\log V_1)^{1/k})\right) - \mathbb{E}\left(V_1\exp(-2\lambda(-\log V_1)^{1/k})\right) \\ &= \mathbb{E}\left(e^{-2W_1}\right) - \int_0^1 u^2\exp(-2\lambda(-\log u)^{1/k})\,\frac{du}{u} \\ &= p_{1,3} - \frac{1}{c}\int_0^1 \exp(-(2\lambda/c^{1/k})(-\log x)^{1/k})\,x^{2/c-1}\,dx \end{aligned}$$

so that finally (for some constant c to be chosen)

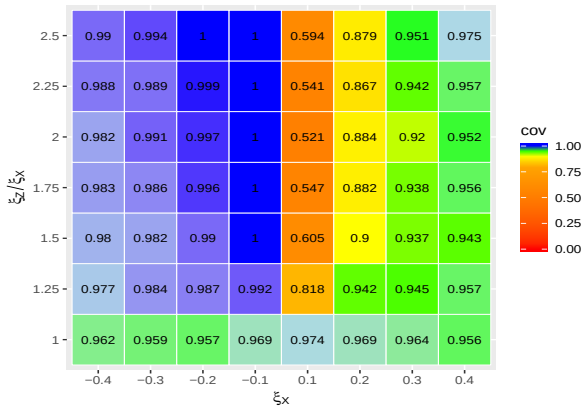
$$E_{1,2} \;=\; p_{1,3} + \tfrac{1}{2}p_{1,2}^2 - \frac{1}{c}\int_0^1 \exp\{-(2\lambda/c^{1/k})(-\log x)^{1/k}\}\, x^{2/c-1}\, dx.$$

Similarly, we provide here the expressions that we used in our `R` routines for evaluating the partial derivatives with respect to λ and k of the functions $g_j(\theta)$ (for $j = 1, 2$ or r, described in subsection 7.2), and thus the jacobians involved in the asymptotic variance of our parametric estimator :

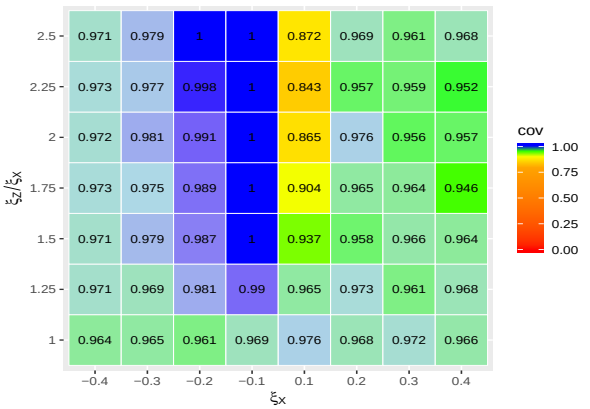
$$\begin{aligned}
\frac{dg_j}{d\lambda}(\lambda,k) &= (-j)c^{-1-1/k}\int_0^1 (-\log u)^{1/k}\exp(-j'(-\log u)^{1/k}).\, u^{1/c-1}\, du \\
\frac{dg_j}{dk}(\lambda,k) &= (j\lambda/k^2)c^{-1-1/k}\int_0^1 \log(-(1/c)\log u)(-\log u)^{1/k}\exp(-j'(-\log u)^{1/k}).\, u^{1/c-1}\, du
\end{aligned}$$

where $j' := j\lambda/c^{1/k}$ for some value c to be chosen.

FIGURES

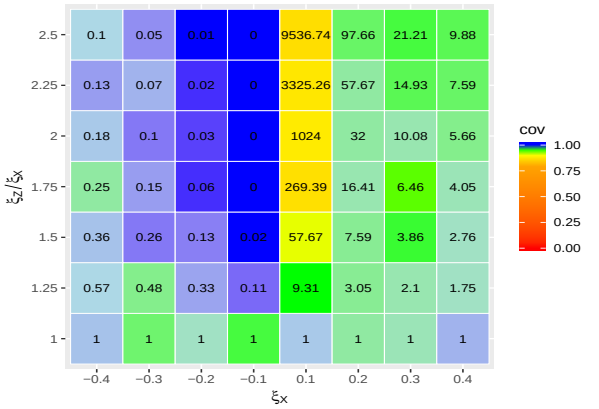


(a) Coverage probabilities based on (9) for $\log(\widehat{p}_{1,r})$

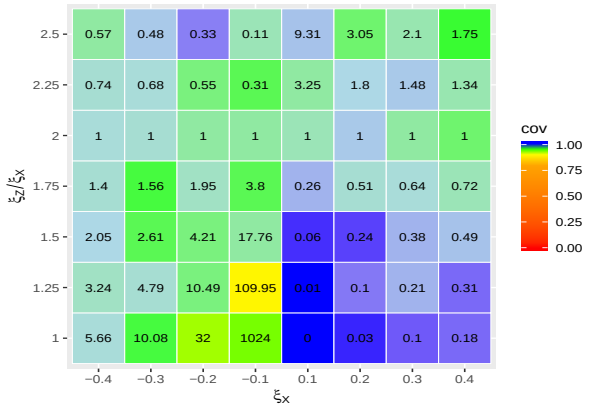


(b) Coverage probabilities based on (9) for $\log(\widehat{p}^{(W)}_{1,r})$

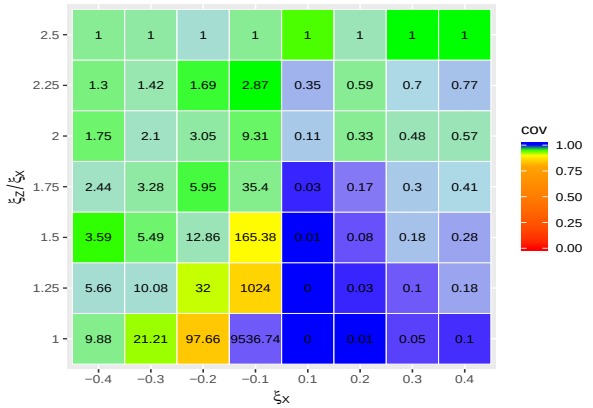
Figure 1. Heatmaps of 95% coverage probabilities of confidence intervals for $p_{1,r}$ when $r = 10$. Values inside cells are the coverage probabilities obtained from 1000 estimated confidence intervals based on Equation (9) with $X \sim GEV(\mu_X, 1, \xi_X)$ and $Z \sim GEV(\mu_Z, 1, \xi_X/k)$ with sample sizes $n = 30$ and $m = 150$. Colors also correspond to the value of the coverage. The x and y axis indicate the shape parameter ξ_X and the ratio $1/k = \xi_Z/\xi_X$, respectively. Panels (a) and (b) correspond the estimates defined by (3) and (7), respectively.



(a) $\sigma_X = 1$, $\sigma_Z = 1$



(b) $\sigma_X = 1$, $\sigma_Z = 2$



(c) $\sigma_X = 1$, $\sigma_Z = 2.5$

Figure 2. Same simulation setup than in Figure 1, but coverage probabilities of confidence intervals for $far(r)$ based on formula (10) when $r = 20$ and with changing scale parameters. Each cell contains the rounded value of the Weibull parameter $\lambda = ((\xi_X/\xi_Z) \times (\sigma_Z/\sigma_X))^{-1/\xi_X}$. The y-axis corresponds to the inverse of the Weibull parameter $k = \xi_X/\xi_Z$.

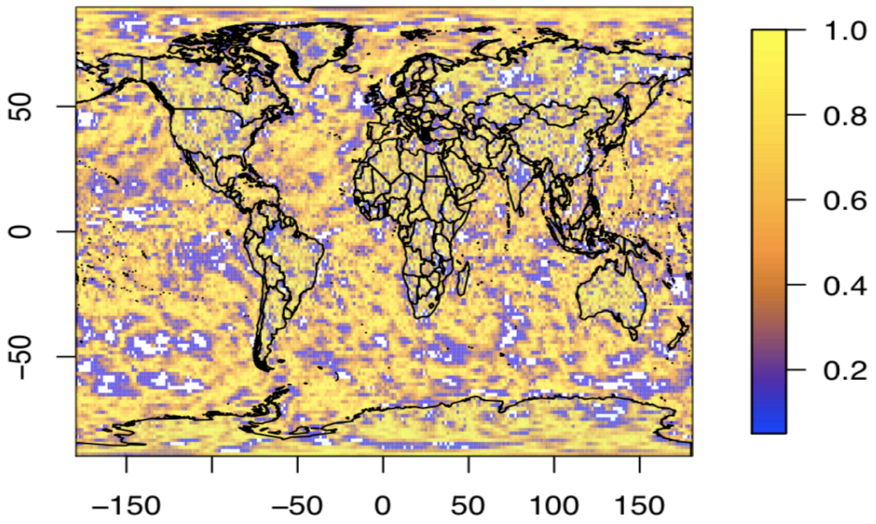
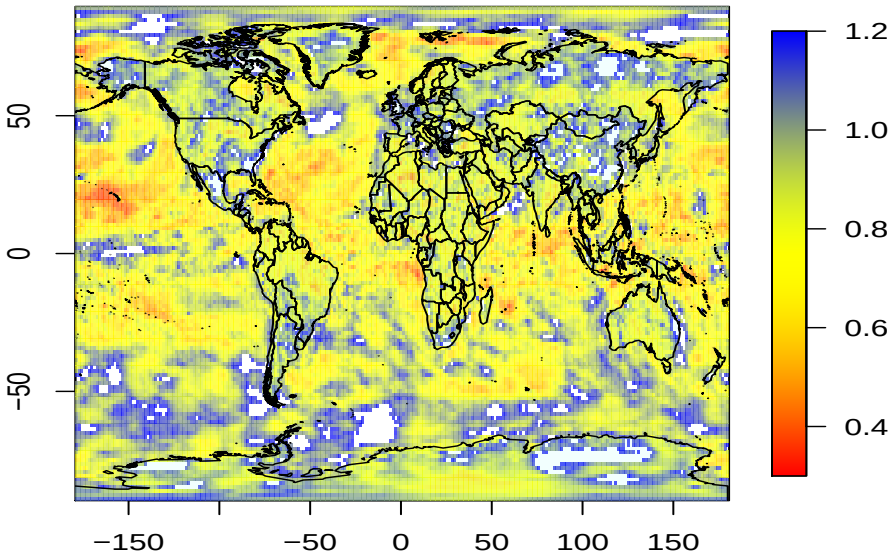
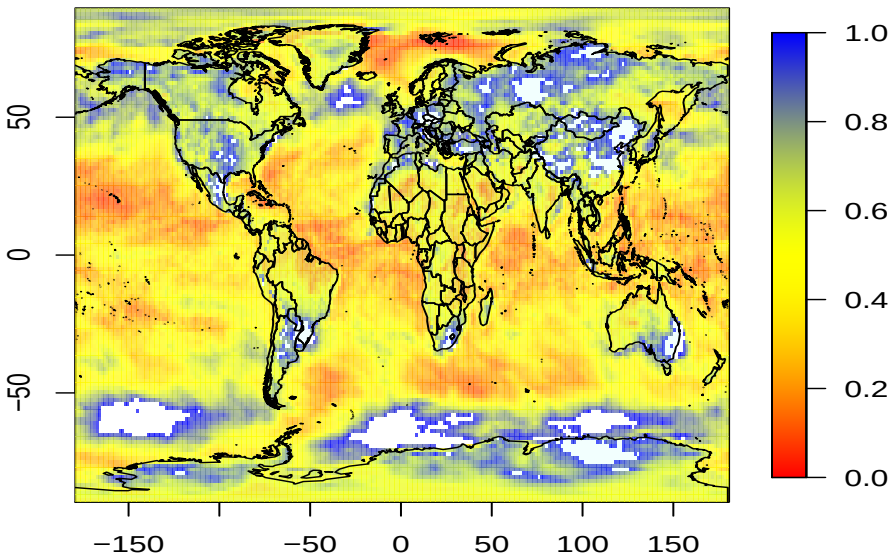


Figure 3. Records analysis of yearly maxima of daily maxima of near-surface air temperature issued from the numerical climate model CNRM-CM5 of Météo-France (the factual run corresponds to "'all forcings" run in 1975-2005 and the counterfactual one to a "natural forcing" run over 1850-2005: This map displays p-values of the likelihood-based Weibullity test, see Section 3.5, white areas corresponding to the p-values lower than 5%

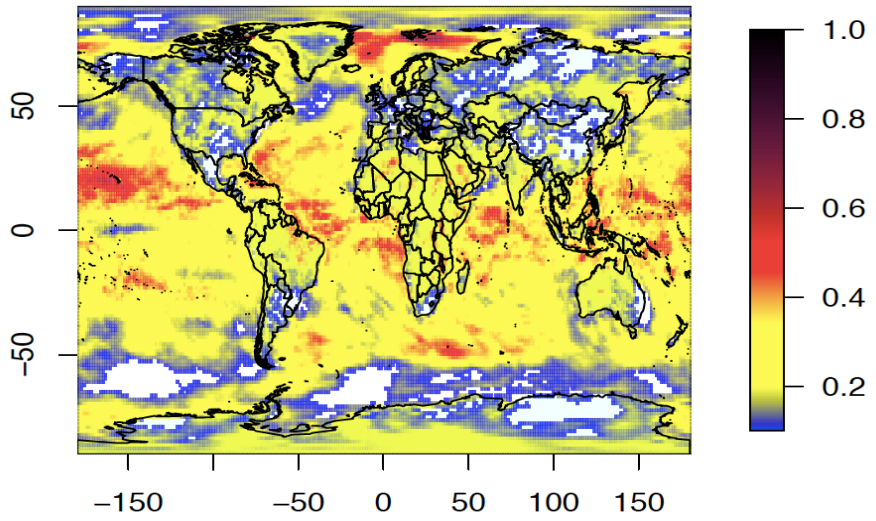


(a) Map of shape parameter estimates (white zones for $\widehat{k} > 1.2$)

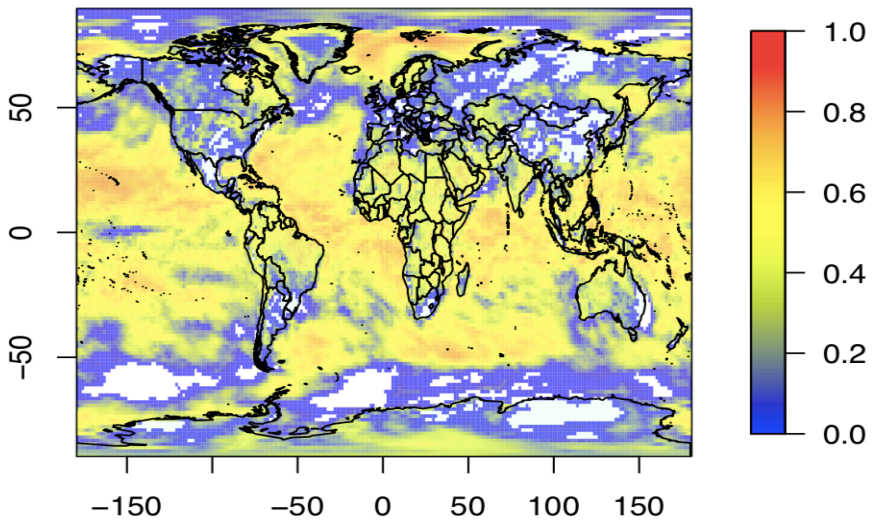


(b) Map of scale parameter estimates (whites zones for $\widehat{\lambda} > 1$)

Figure 4. Same setup as in Figure 3: Weibull parameters estimated from (7)



(a) Map of $\widehat{p}_{1,10}^{(W)}$ (white zones when $\widehat{p}_{1,10}^{(W)} < 1/r$; maximum is at 0.729, 95th percentile at 0.409)



(b) Map of $far(10)$ confidence interval 95% lower bound (white zones when this lower bound is ≤ 0)

Figure 5. Same setup as in Figure 3: Probability that the factual world produces a temperature record with respect to the previous 9 counterfactual years. White zones correspond to the gridpoints at which $\widehat{far}^{(W)}(10)$ or this lower bound is nonpositive (in the second case, this means that the hypothesis "$far(10) \leq 0$" is not rejected at the 2.5% risk).

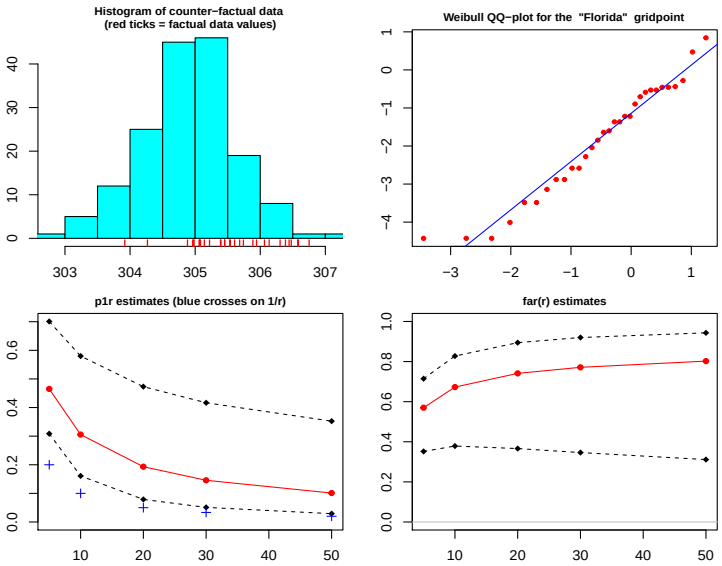


Figure 6. A local analysis at the gridpoint near south Florida. Point estimates (red solid) and confidence intervals (black dotted) of $p_{1,r}$ and $far(r)$ are shown in the two lower panels for the values $r = 5, 10, 20, 30, 50$. In the upper left panel, the histogram of the counter-factual data (with blue ticks on the horizontal axis being the factual data values). In the upper right panel, the qqplot of values $\widehat{W}_1, \ldots, \widehat{W}_{31}$ indicates that (X, Z) seems to belong to the W-class.

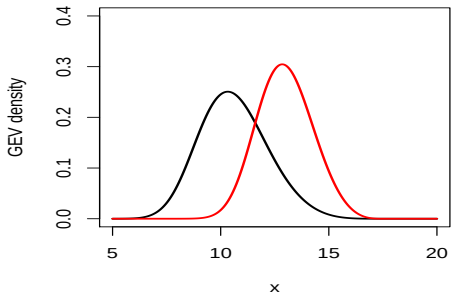


Figure A.1. Densities of GEV distributions with equal support : $\xi_X = -0.2$, $\mu_X = 10$, $\sigma_X = 1.5$ (black line), $\xi_Z = -0.25$, $\mu_Z = 12.5$, $\sigma_Z = 1.25$ (red line), Weibull shape parameter $k = 0.8 < 1$.