



**HAL**  
open science

## A statistical method to model non-stationarity in precipitation records changes

Paula Gonzalez, Philippe Naveau, Soulivanh Thao, Julien Worms

► **To cite this version:**

Paula Gonzalez, Philippe Naveau, Soulivanh Thao, Julien Worms. A statistical method to model non-stationarity in precipitation records changes. 2024. hal-04006516v2

**HAL Id: hal-04006516**

**<https://hal.science/hal-04006516v2>**

Preprint submitted on 10 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A statistical method to model non-stationarity in precipitation records changes

Paula Gonzalez<sup>1</sup>, Philippe Naveau<sup>1</sup>, Soulivanh Thao<sup>1</sup>, and Julien Worms<sup>2</sup>

<sup>1</sup>Université Paris-Saclay, UVSQ, Laboratoire des Sciences du Climat et de l'Environnement,

LSCE/IPSL/ESTIMR, CNRS-CEA, Gif-sur-Yvette, France.

<sup>2</sup>Université Paris Saclay, UVSQ, Laboratoire de Mathématiques de Versailles, CNRS, Versailles, France.

## Key Points:

- This work proposes a simple definition of non-stationary records and offers a method to assess the likelihood of record event changes.
- We analyze annual maxima of daily precipitation, whose statistical features strongly depart from a Gaussian probability distribution.
- IPSL-CM6A-LR climate model highlights that by 2023 signals of rainfall yearly maxima decadal records have emerged on the half of the globe.

---

Corresponding author: Paula Gonzalez, [paula.gonzalez@lsce.ipsl.fr](mailto:paula.gonzalez@lsce.ipsl.fr)

## Abstract

In the context of climate change, assessing how likely a particular change or event was caused by human influence is important for mitigation and adaptation policies. In this work we propose an extreme event attribution (EEA) methodology to analyze yearly maxima records, key indicators of climate change that spark off media attention and research in the EEA community. Although they deserve a specific statistical treatment, algorithms tailored to record analysis are lacking. This is particularly true in a non-stationarity context. This work aims at filling this methodological gap by focusing on records in transient climate simulations. We apply our methodology to study records of yearly maxima of daily precipitation issued from the numerical climate model IPSL-CM6A-LR. Illustrating our approach with decadal records, we detect in 2023 a clear human induced signal in half the globe, with probability mostly increasing, but decreasing in the south and north Atlantic oceans.

## Plain Language Summary

The increase of frequency and strength of climate extremes raises the interest in quantifying the extent to which these changes are influenced by climate change. In this work we propose an Extreme Event Attribution (EEA) methodology allowing us to assess whether climate records are attributable to climate change. Records have been typically studied by considering climate unvarying in some time span, despite the fact that climate is constantly changing. This work aims at filling this methodological gap by focusing on records in time-varying climate simulations. We apply our methodology to study records of yearly maxima of daily precipitation issued from the latest version of the Institute Pierre Simon Laplace climate model. Illustrating our approach with decadal records, we detect in 2023 a clear human induced signal in almost half of the globe. Even though decadal record probability mostly increases, we observe a decrease of record probability in the south and north Atlantic oceans.

## 1 Introduction

In its recent media release on January 23rd 2023, the European Copernicus program highlighted that *2022 was a year of climate extremes, with record-high temperatures and rising concentrations of greenhouse gases*. This statement underlines the current interest in records changes and records breaking. This can be explained by their high societal and economic impacts, the question of mitigation and the attribution to anthropogenic forcings. Assessing how likely a particular extreme event has been caused by human influence has been an active field of research (IPCC, 2014). Changes in various climate events have been well documented (Gulev et al., 2021), for example concerning the frequency and intensity of extreme precipitation at continental to global scales (Dong et al., 2021; Alexander, 2016).

To attribute changes in any extreme climate variable, the field of extreme event attribution (EEA) (see, e.g., Stott et al., 2016; Naveau et al., 2020) specifically aims at comparing the probability of the same extreme climate event within two different realities: a *factual* world that mimics the conditions observed around the time of the event (i.e, a world that contains the effect of human influence on climate) and a *counterfactual* world, in which anthropogenic emissions have never occurred (Angéilil et al., 2017). This design of experiment can only be tested via numerical climate models as a world without anthropogenic forcing does not exist (Hegerl & Zwiers, 2011). This numerical setup seeks to address the following question: can we attribute the change of likelihood of a particular extreme event to the difference between the factual and counterfactual worlds? Mathematically, most EEA studies compare the following two probabilities of exceeding some

62 high threshold  $u$  for a given year  $t$ ;

$$p_{0,u}(t) = P(X_t > u) \quad \text{and} \quad p_{1,u}(t) = P(Z_t > u), \quad (1)$$

63 where the notations  $X_t$  and  $Z_t$  represent the same real-valued variable of interest (e.g.,  
 64 annual maxima of daily precipitation in our application) but  $X_t$  corresponds to its counter-  
 65 factual version while  $Z_t$  denotes its factual one. The temporal index  $t$  will correspond  
 66 to the years from 1850 to 2100 in our application, see Section 3. The choice of the thresh-  
 67 old  $u$  that defines the extreme event is usually delicate and depends on the case study  
 68 at hand.  $1 - \frac{p_{1,u}(t)}{p_{0,u}(t)}$  has been called the fraction of attributable risk by Stott et al. (2016).  
 69 This type of ratio can be interpreted within Pearl's counterfactual theory of causality  
 70 (Hannart et al., 2016; Hannart & Naveau, 2018). By leveraging multivariate extreme value  
 71 theory (EVT), this relative ratio can be optimized to highlight causality (see, e.g., Kir-  
 72 iliouk & Naveau, 2020).

73 It is important to notice that, given a fixed  $u$  and a year of interest  $t$ , the proba-  
 74 bilities defined by (1) do not directly provide relevant information concerning records.  
 75 In this article we provide a methodology to analyze record events in a non-stationary con-  
 76 text. The meaning of *records* is not based on a fixed threshold. Instead, it is rooted in  
 77 the comparison between the current value and past observations. For example, the state-  
 78 ment that 2016 was the warmest global temperature on record can only be understood  
 79 with respect to a reference period, here since the moment when reliable instrumental mea-  
 80 surements were available. Mathematically, the probability of being a record with respect  
 81 to a given period  $\mathfrak{R}$  can be generically defined as

$$P(Y > \max(X_t : t \in \mathfrak{R})), \quad (2)$$

82 where the event  $\{Y > \max(X_t : t \in \mathfrak{R})\}$  means that the value  $Y$  is larger than any  
 83 value from the sample  $X$  obtained during the reference period  $\mathfrak{R}$ . As highlighted pre-  
 84 viously, EEA is rooted in the comparison between factual and counterfactual data, the  
 85 later being considered as the baseline. Regarding this yardstick, it is natural to wonder  
 86 what is the probability of observing a record in the counterfactual world, i.e. taking  $Y =$   
 87  $X_{t+1}$  in (2) with respect to a given reference period  $\mathfrak{R} = \{1, \dots, T\}$ . A similar but more  
 88 complex question is to estimate the probability that the *factual* observation at time  $T+$   
 89  $1$ , i.e. taking  $Y = Z_{t+1}$  in (2), would have been a record in the *counterfactual* world.  
 90 This leads to our two new definitions of record probabilities  
 91

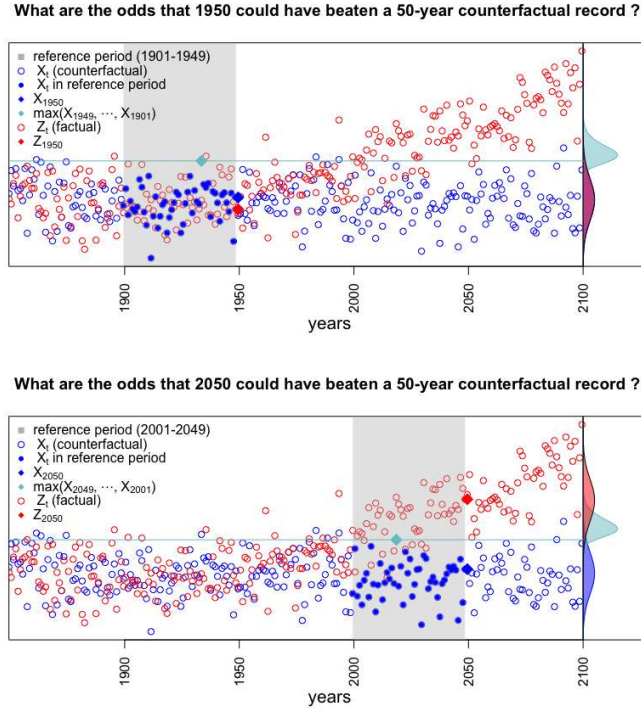
$$p_{0,r}(t) := P(X_t > \max(X_{t-1}, \dots, X_{t-r+1})), \quad (3)$$

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

92 where, given the  $r-1$  counterfactual observations before the year  $t$ ,  $p_{0,r}(t)$  represents  
 93 the probability of the counterfactual value being the largest at time  $t$ , and  $p_{1,r}(t)$  is the  
 94 same quantity but when the last value comes from the factual world. This setup is sim-  
 95 ilar to return level computations. For example, the 100-year return period in risk anal-  
 96 ysis is given to a risk manager and the task is to estimate the corresponding 100-year  
 97 return level from a sample of data. The record length  $r$  can be arbitrarily set in our study.  
 98 The simplest form of this definition arises when setting  $r = 2$ . This special case was  
 99 studied in (Naveau & Thao, 2022). It comes down to comparing the central part of the  
 100 factual and counterfactual distributions with one statistic. In this paper, our goal is to  
 101 estimate  $p_{1,r}(t)$  for any positive integers  $r$ . We would like to highlight that it is possi-  
 102 ble to find two factual worlds,  $Z_t$  and  $Z_t^*$  such that  $P(Z_t > X_{t-1}) = P(Z_t^* > X_{t-1})$   
 103 and  $P(Z_t > \max(X_{t-1}, X_{t-2})) \neq P(Z_t^* > \max(X_{t-1}, X_{t-2}))$ . So checking  $p_{1,2}(t)$  is  
 104 not sufficient to capture other changes in  $r$ , (see example in the Supplementary mater-  
 105 ial, section 1.  
 106  
 107  
 108

109 To understand the difference between  $p_{0,r}(t)$  and  $p_{1,r}(t)$  in a non-stationarity con-  
 110 text, we compare them in Figure 1 for  $r = 50$  years and two different years,  $t = 1950$

111 in panel (a) and  $t = 2050$  in panel (b). The sequence of blue-colored dots indicates a  
 112 simulated example of a counterfactual time series  $X_t$  with  $t$  varying from 1850 to 2100,  
 113 while the sequence of red-colored points represents a factual trajectory. The upper panel  
 114 highlights the year  $t = 1950$  with the grey vertical band enlightening its associated refer-  
 115 ence period from 1900 to 1949. The probability  $p_{1,50}(1950)$  assesses how often the red  
 116 diamond of the year 1950 could have been above the maximum of the blue solid dots.  
 117 During the time window 1900–1950, the effect of the anthropogenic forcing is not very  
 118 apparent in this example, and  $p_{1,50}(1950)$  should be close to the value of  $p_{0,50}(1950)$ . In  
 119 contrast, observing a factual record in 2050 with respect to the counterfactual world of  
 120 2000–2050 should be much more probable, i.e.  $p_{1,50}(2050)$  is much greater than  $p_{0,50}(1950)$   
 121 in this artificial example. The bottom panel (b) highlights this phenomenon as the non-  
 122 stationarity of  $Z_t$  increases the likelihood of the red diamond of 2050 to be above the largest  
 123 value of  $X_t$  with  $t$  spanning 2000–2050. In this simulated example, this can be clearly  
 124 seen with the probability density functions (pdf) displayed on the right side of Figure  
 125 1. The red pdf in panel (a) that corresponds to the pdf of  $Z_t$  for year  $t = 1950$  has been  
 126 switched up in year  $t = 2050$ . This non-stationarity in the factual world explains the  
 127 change in 50-year records and highlights the necessity of interpreting records with re-  
 128 spect to a chosen time window.



**Figure 1.** Schematic example to interpret the 50-year record probabilities (i.e.  $r = 50$  in Equation (3)) in a non-stationary context. The upper panel highlights the year  $t = 1950$  and the lower panel, the year  $t = 2050$ . The blue and red colors represent a simulated counterfactual trajectory,  $X_t$ , and a factual one,  $Z_t$ , respectively. The solid blue dots correspond to  $(X_{t-1}, \dots, X_{t-49})$ . On the right side of the plot, the probability density functions of  $Z_t$ ,  $X_t$  and  $\max(X_{t-1}, \dots, X_{t-49})$  are displayed in red, blue and light blue, respectively.

129 As already pointed out, one advantage of preferring records over exceedances is that  
 130 there is no need to choose a threshold  $u$  like in (1), but this is not the only one. Record

131 analysis relies on their relative nature. For each climate model the reference value of com-  
 132 parison is derived from the model's own outputs, allowing us to easily compare the re-  
 133 sults. In some instances, this allows to bypass the bias correction steps in multi-model  
 134 error analysis (Naveau & Thao, 2022). More importantly, the interpretation of  $p_{0,r}(t)$   
 135 corresponds to the classical understanding of records for the general public. The expres-  
 136 sion of  $p_{0,r}(t)$  is also very simple

$$137 \quad p_{0,r}(t) = \frac{1}{r}, \quad \text{for all years } t \text{ in a reference period of length } r, \quad (4)$$

138 under the assumption of exchangeability in the counterfactual world (Chow & Teicher,  
 139 2003). For example, yearly maxima of daily precipitation, due to their high temporal vari-  
 140 ability at the yearly scale, can be considered independent and identically distributed (iid)  
 141 at the yearly scale in the counterfactual world, and therefore exchangeable.

142 (4) 2 For illustration purposes, Figure 2, panel (a) shows how Equation (4) can be  
 143 derived from the exchangeability assumption. It displays the bivariate pdf of exchange-  
 144 able max-stable variables  $(X_t, X_{t-1})$  (see Supplementary material, section 2) (Beirlant  
 145 et al., 2005; Coles, 2001). Since  $X_t$  and  $X_{t-1}$  are exchangeable,  $P(X_t > X_{t-1}) = P(X_{t-1} >$   
 146  $X_t)$ , and consequently  $p_{0,2}(t) = 1/2$ .

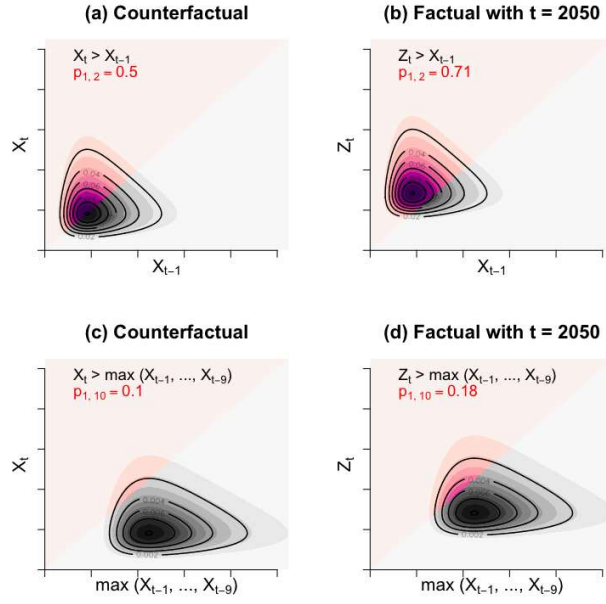
147 This can be visually understood by noticing the symmetric nature of panel (a) around  
 148 the diagonal. The zone highlighted in red represent the event  $\{X_t > X_{t-1}\}$ , which un-  
 149 der the exchangeability assumption, represents half of the mass of the pdf. In contrast,  
 150 (b) focuses on the couple  $(Z_t, X_{t-1})$  for  $t = 2050$ , which represents the case of  $r = 2$   
 151 in the factual world, here  $p_{1,2}(t) = P(Z_t > X_{t-1}) = .71$ . We notice that the bivari-  
 152 ate pdf is no longer symmetric with respect to the  $\{Z_t = X_{t-1}\}$  line. The event  $\{Z_t >$   
 153  $X_{t-1}\}$ , represented in red contains more mass than in panel (a), meaning that the prob-  
 154 ability of this event is higher. Then, for  $r = 2$  exchangeability of the counterfactual world  
 155 allow us to do attribution by comparing  $p_{1,2}(t)$  to  $1/2$ . For  $r = 3$ , exchangeability pro-  
 156 vides  $P(X_t > \max(X_{t-1}, X_{t-2})) = P(X_{t-1} > \max(X_t, X_{t-2})) = P(X_{t-2} > \max(X_t, X_{t-1}))$ ,  
 157 and leads to  $p_{0,3}(t) = 1/3$ . This argument can be repeated for any  $r$  in an exchange-  
 158 able counterfactual world. 2 In panel (c), we now consider the decadal records proba-  
 159 bility  $p_{0,10}$  by looking at a bivariate density of the vector  $(X_t, \max(X_{t-1}, \dots, X_{t-9}))$ . The  
 160 zone in red represents the event  $X_t > \max(X_{t-1}, \dots, X_{t-9})$  and contains  $1/10$  of the mass  
 161 of the bivariate pdf. In contrast, panel (d) shows  $p_{1,10}(t) = .18$  for  $t = 2050$  as proba-  
 162 bility for the factual value  $Z_t$  to become a decadal record with respect to  $\max(X_{t-1}, \dots, X_{t-9})$ .  
 163 Therefrom, the main problem we would like to address in this work is how to efficiently  
 164 and rapidly estimate  $p_{1,r}(t)$  for any given year  $t$  and for any given record length  $r$  in a  
 165 non-stationary context.

166 This article is organized as follows. In Section 2, we propose a new methodology  
 167 that handles non-stationary situations when attributing records. In Section 3, we apply  
 168 this transient record approach to analyze yearly maxima of daily precipitation issued from  
 169 climate model IPSL-CM6A-LR from the CMIP6 inter-comparison projet. Finally, in Sec-  
 170 tion 4, we summarize the added value of this methodology and discuss the results.

## 171 2 Inference of non-stationary record probabilities

172 Our inference goal is to estimate  $p_{1,r}(t)$  for any record period length  $r$ . This means  
 173 that  $r$  can even be larger than the length of time series under study, i.e. our approach  
 174 should be able to extrapolate beyond the largest record ever observed in either the fac-  
 175 tual or counterfactual worlds. Similarly to the computation of high return levels in hy-  
 176 drology (Katz et al., 2002), the developed approach here is unconditional in the sense  
 177 that there is not need to observe a realization of the event of interest (a record) to com-  
 178 pute its probability of occurrence. Performing such unconditional extrapolation implies  
 179 that a parametric model needs to be imposed, theoretically justified and tested.

Joint distribution of  $(X_t$  or  $Z_t)$  and  $\max(X_{t-1}, \dots, X_{t-r+1})$



**Figure 2.** Joint distribution of  $X_t$  and  $\max(X_{t-1}, \dots, X_{t-r+1})$  and of  $Z_t$  and  $\max(X_{t-1}, \dots, X_{t-r+1})$  in 2050. (a) and (b) illustrate the case  $r = 2$ , where our probability of interest is  $p_{1,2}(t) = P(Z_t > X_{t-1})$ , (c) and (d) illustrate the case  $r = 10$ , where our probability of interest is  $p_{1,10}(t) = P(Z_t > \max(X_{t-1}, \dots, X_{t-9}))$ , the reddish zones represent our events of interest and the intersection between these zones and the pdf its probability.

180 The variable of interest in our study corresponds to annual block maxima (of daily  
 181 rainfall). According to EVT, the classical three-parameter extreme generalized distri-  
 182 bution (GEV) (Coles, 2001; Beirlant et al., 2005) should represent a mathematically sound  
 183 distribution for such variables. Within this framework, one modeling possibility would  
 184 be to fit a three parameter GEV to the counterfactual time series, and a different three  
 185 parameter GEV to the factual temporal sequence. By noticing that records are relative  
 186 quantities, Worms and Naveau (2022) showed, that under the conditions stated in their  
 187 Lemma 1, the estimation of the six GEV parameters can be reduced to inferring only  
 188 the two parameters of the following Weibull random variable defined as

$$189 \quad W_t = -\log G(Z_t) \sim \text{Weibull}(k_t, \lambda_t), \quad (5)$$

190 where  $G(x) = P(X_t \leq x)$  corresponds to the cumulative distribution function of  $X_t$ .

Weibull  $(k_t, \lambda_t)$  denotes a Weibull distribution, with positive parameters  $k_t$  and  $\lambda_t$ , that can be defined by its Laplace transform

$$E[\exp(-uW_t)] = \int_0^\infty e^{-ux} \frac{k_t}{\lambda_t} \left(\frac{x}{\lambda_t}\right)^{k_t-1} e^{-(x/\lambda_t)^{k_t}} dx.$$

191 A bivariate vector  $(X_t, Z_t)$  satisfying (5) is said to belong to the so-called *W-class*. Un-  
 192 der this class, we can make the link between the computation of  $p_{1,r}(t)$  and the Weibull  
 193 Laplace transform.

As previously mentioned, in our application the sequence  $(X_1, \dots, X_t)$  can be assumed independent, and consequently

$$p_{1,r}(t) = E(\exp(-(r-1)W_t)),$$

whenever  $Z_t$  is independent of  $(X_1, \dots, X_{t-1})$ . This info and a reparametrization of the Laplace integral lead to the following expression of  $p_{1,r}(t)$

$$p_{1,r}(t) = \int_0^1 \exp(-(r-1)\lambda_t(-\log x)^{1/k_t}) dx. \quad (6)$$

A key element of our approach is that, for a given year  $t$ , record probabilities can be deduced from the Weibull distribution with parameters  $(k_t, \lambda_t)$ . Hence, checking the Weibull hypothesis has to be done only once. If not rejected, any  $p_{1,r}(t)$  can be inferred for any  $r$ . So, our inference strategy is to first infer these two parameters and then plug their estimates in (6). Concerning the first step, it can be implemented by coupling a Nadaraya-Watson kernel regression method with a method of moments to estimate  $\hat{\lambda}_t$  and  $\hat{k}_t$  (see Section 3 of the Supplementary material and Naveau & Thao, 2022). We call  $\hat{p}_{1,r}(t)$  the estimator obtained by this method. Its theoretical properties can be found in Section 4 of the supplementary material. It is noteworthy that proposition 3 and 5 from Worms & Naveau (2022) states that the parametric estimator based on the Weibull assumption has lower relative error than the non-parametric estimator. Thus, whenever the Weibull assumption can be checked, a parametric approach is preferable for the inference of record probabilities associated to large values of  $r$ .

### 3 Analysis of yearly maxima of daily precipitation

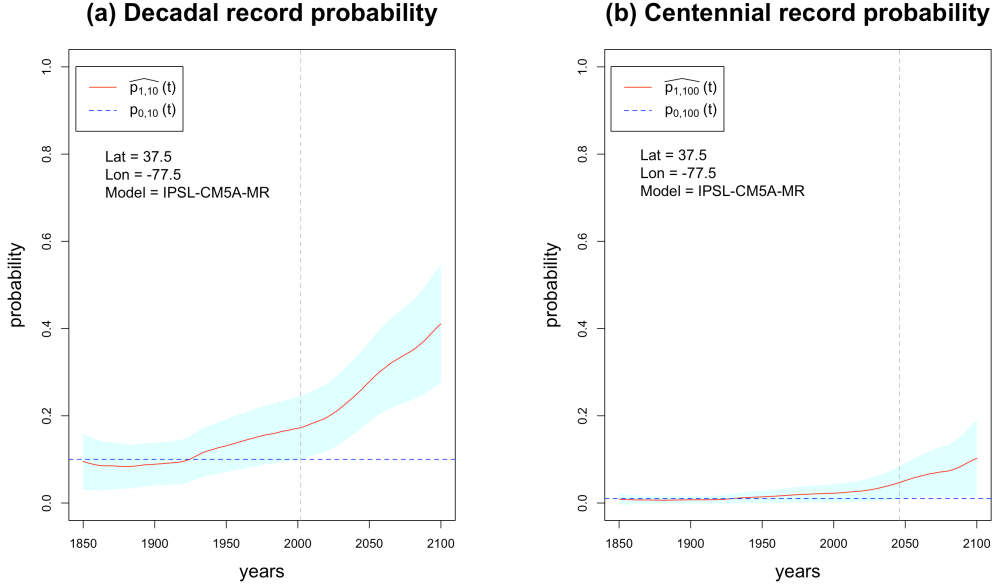
We use our methodology to study annual maxima of daily precipitation, a variable for which the W-class assumption is usually reasonable (see Supplementary material, Section 5). The climate model used here is the IPSL-CM6A-LR from the CMIP6 inter-comparison project. Our factual trajectory of yearly maxima of daily precipitation corresponds to the historical global run over the period 1850 - 2014 combined with the SSP5-8.5 scenario over the period 2015 - 2100. Our counterfactual trajectory is represented by a global run with only natural forcings over the period 1850 - 2020.

To illustrate our approach, we first focus on the analysis of decadal and centennial record probability evolution, i.e.  $p_{1,10}(t)$  and  $p_{1,100}(t)$  for  $t \in \{1850, \dots, 2100\}$ , at a randomly selected grip point near Richmond in Virginia (USA). From Equation (4), we expect to have  $\hat{p}_{1,10}(t)$  near 1/10 and  $\hat{p}_{1,100}(t)$  near 1/100 during the pre-industrial period. This is confirmed by Figure 3 that displays the decadal (panel (a)) and centennial (panel (b)) record probability estimates of  $p_{1,10}(t)$  and  $p_{1,100}(t)$  as a function of the year (x-axis). From this grid point near Richmond, a clear climate change signal emerges on decadal records from the year 2002, year from which the confidence interval no longer contains 1/10. By the year 2100, decadal record are almost four times more likely than in a world without climate change.

For the centennial record period, see panel (b), observing a record in 2100 is about ten times more likely than in world without climate change, with a clear climate change signal emerging in 2045. It's noteworthy that this centennial signal emerges more than 40 years later than the signal observed for decadal records. This is due to the fact that the relative error increases as a function of  $r$  (see Proposition 5 from Worms and Naveau, 2022 for details). This results in wider confidence intervals and a later detection of a significant signal. This is consistent with our calculations (see Supplementary material, section 4). Smoothing bandwidth is  $h = 60.5$  as previously defined by (Naveau & Thao, 2022).

This analysis at a specific location begs the question of when and where a significant attributable signal emerges at the global scale. To answer this question, we leverage Equation (4) and we define record emergence time associated with a given record



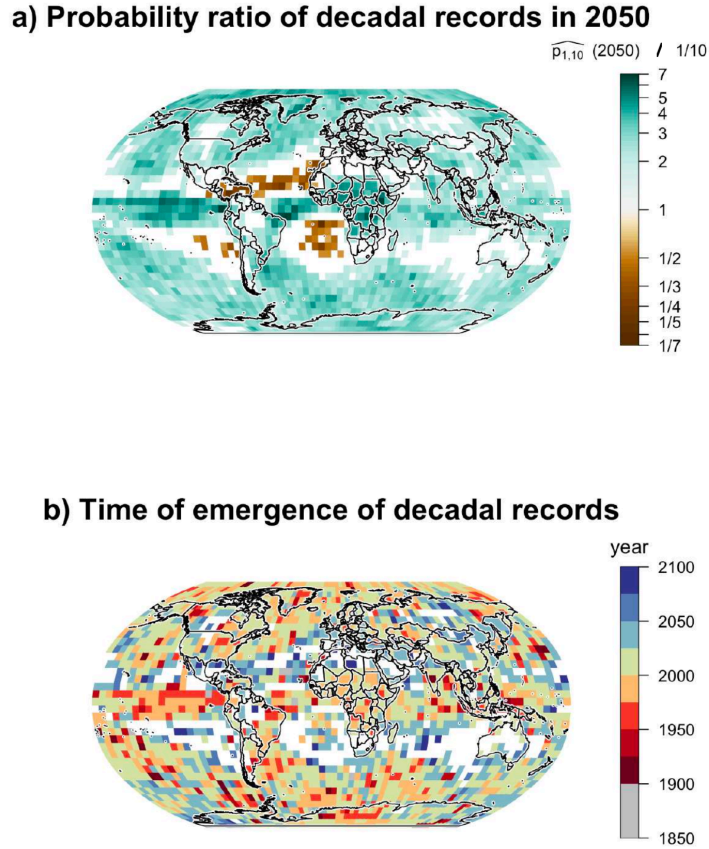


**Figure 3.** Decadal (a) and centennial (b) record probability of yearly maxima of daily precipitation at Richmond, Virginia grid-point (latitude 37.5, longitude -77.5), using the IPSL-CM6A-LR climate model and the scenario SSP5-8.5, with spatial resolution of 5 x 5 degrees. The light blue zones represent the asymptotic confidence intervals of confidence level 95%.

length  $r$  as the first year when  $\hat{p}_{1,r}(t)$  is significantly different from its counterfactual value, i.e. different from  $p_{0,r}(t) = 1/r$ , with confidence level of 95%. Mathematically, this brings the following definition

$$\tau_{0.95}(r) = \min \left\{ t \text{ such that for all } t' \geq t, \frac{1}{r} \notin [\hat{p}_{1,r}(t') \pm 1.96 \hat{\sigma}_{rt'}] \right\}, \quad (7)$$

where  $\hat{\sigma}_{rt}$  represents the estimation of the asymptotic standard deviation and 1.96 corresponds to the Gaussian significance level 0.95 (see Section 4 of the supplementary material for details). Equation (7) allows us to identify the emergence year of any gridpoint. Panel (a) of Figure 4 highlights decadal record probability ratio on 2050 in the zones where by that year there is already a clear signal of climate change. By 2050, there is a clear signal on 80% of the globe, we expect decadal records on tropical latitudes to be up to seven times more likely than in a world without anthropogenic forcing. We do not only observe an increase of decadal record probability, we can also identify a clear decrease in the south and north Atlantic ocean and the south Pacific ocean, which is consistent with previous studies showing a decline of precipitation in these zones (Pfahl et al., 2017). Between these zones of increasing and decreasing probability we observe transition zones, where climate change effect is not clear. When analyzing centennial records, even if climate change signal is less clear, the increase and decrease of probability patterns remain unchanged. Equation (7) also brings out climate change signal timeline. Panel (b) of Figure 4 shows decadal record emergence times, i.e.  $\tau_{0.95}(10)$  over the globe. 26% of climate change signal emerged between 2000 and 2023, adding up to this last year 57% of the globe. An equivalent analysis for centennial records can be found in section 6 of the supplementary material.



**Figure 4.** (a) Decadal record probability ratio on 2050 with respect to the counterfactual world, the white zones represent the gridpoints where by 2050 climate change signal has still not emerged, using (7) as criteria. (b) Emergence record times defined by (7), the white zones represent the gridpoints where by 2100 climate change signal has still not emerged, the grey points represent the gridpoints where  $\widehat{p}_{1,10}(t)$ 's confidence interval does not contains 1/10 during the pre-industrial period, these points are left out of our analysis as considered poorly represented. Results obtained using the IPSL-CM6A-LR climate model and the scenario SSP5-8.5, with a spatial resolution of 72 x 36 grid points.

265

## 4 Conclusion and discussion

266

267

268

269

270

271

272

273

To summarize, we proposed and studied a new EEA record analysis in a transient setup to estimate record probability at each time step  $t$  and for any record length  $r$ . Our approach accounts for the non-stationarity of the factual world without constraining the shape parameter of the GEV distributions to be constant. This represents an added value compared to methods based on return periods where non-stationarity is not fully captured. Additionally, it fills a methodological gap for record analysis, as records have usually been studied in a stationary context, with limited advances on non-stationary times series. Furthermore, our approach has a straightforward interpretation and it bypasses

274 the separate estimation of both distributions  $X_t$  and  $Z_t$  separately. By focusing on relative  
 275 changes between the counterfactual and factual worlds (see Equation (1) of (Naveau  
 276 & Thao, 2022)), we mitigate the underestimation bias of climate models on the climate  
 277 change signal associated with extreme precipitation (Min et al., 2011; Fischer & Knutti,  
 278 2015). In our framework ( $X_1, \dots, X_t$ ) are considered independent. Yet, our approach is  
 279 still valid in a dependent scenario, if the sequence corresponds to any max-stable time  
 280 series.

281 Our analysis of yearly maxima of daily precipitation obtained from the IPSL-CM6A-  
 282 LR (scenario SSP5-8.5) indicates that precipitation records are affected all over the world,  
 283 with a clear climate change signal on decadal records before the year 2050. The trop-  
 284 ical latitudes and the polar circles appear to be the zones where record probabilities will  
 285 increase the most and the north and south Atlantic ocean those where they clearly de-  
 286 crease. This result is consistent with previous studies on changes of precipitation (Pfahl  
 287 et al., 2017; Tandon et al., 2018; Dong et al., 2021). However, our conclusions are only  
 288 valid for the IPSL-CM6A-LR climate model. As improvements of this methodology, it  
 289 would be interesting to incorporate multi-model climate error. For example, the tech-  
 290 nique proposed by Naveau and Thao (2022) may be used to handle this type of error.  
 291 In addition, our analysis was made independently for each grid and it is likely that the  
 292 signal in record emergence times will be enhanced by incorporating spatial information  
 293 at a regional scale. Although more statistically complex, multivariate EVT used in EEA  
 294 (Kiriliouk & Naveau, 2020) could be implemented to perform this task.

## 295 Acknowledgments

296 Part of this work was supported by the French national program (80 PRIME CNRS-INSU),  
 297 and the European H2020 XAIDA (Grant agreement ID: 101003469). This work bene-  
 298 fitted from state aid managed by the National Research Agency under France 2030 bear-  
 299 ing the reference ANR-22-EXTR-0005 (TRACCS-PC4-EXTENDING project). The au-  
 300 thors also acknowledge the support of the French Agence Nationale de la Recherche (ANR)  
 301 under reference ANR-20-CE40-0025-01 (T-REX), and the ANR EXSTA.

## 302 Open Research

303 The simulations of the IPSL-CM6A-LR model used in this paper come from the  
 304 CMIP6 inter-comparison project (Boucher et al., 2021). They are available through the  
 305 Earth System Grid Federation. An overview of the guidelines for CMIP6 Data users is  
 306 available through the Program for Climate Model Diagnosis & Intercomparison (Eyring  
 307 et al., 2016). The presented methodology was implement using R. Figures were made  
 308 using R. The code developed in this work is available on GitHub [https://github.com/  
 309 PaulaFlorescia/A-statistical-method-to-model-non-stationarity-in-precipitation-  
 310 records-changes](https://github.com/PaulaFlorescia/A-statistical-method-to-model-non-stationarity-in-precipitation-records-changes)

## 311 References

- 312 Alexander, L. V. (2016). Global observed long-term changes in temperature  
 313 and precipitation extremes: A review of progress and limitations in IPCC  
 314 assessments and beyond. *Weather and Climate Extremes*, 11, 4–16. doi:  
 315 <https://doi.org/10.1016/j.wace.2015.10.007>
- 316 Angéllil, O., Dáithí, S., Wehner, M., Paciorek, C. J., Krishnan, H., & Collins, W.  
 317 (2017). An independent assessment of anthropogenic attribution statements for  
 318 recent extreme temperature and rainfall events. *Journal of Climate*, 5-16. doi:  
 319 <https://doi.org/10.1175/JCLI-D-16-0077.1>
- 320 Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2005). *Statistics of Extremes:  
 321 Theory and Applications*. doi: <https://doi.org/10.1002/0470012382>

- 322 Boucher, O., Denvil, S., Levvasseur, G., Cozic, A., Caubel, A., Foujols, M.-A.,  
 323 ... Marchand, M. (2021). *IPSL IPSL-CM5A2-INCA model output pre-*  
 324 *pared for cmip6 AerChemMIP*. Earth System Grid Federation [Dataset]. doi:  
 325 <https://doi.org/10.22033/ESGF/CMIP6.13641>
- 326 Chow, Y. S., & Teicher, H. (2003). *Probability Theory: Independence, Interchange-*  
 327 *ability, Martingales*. New York: Springer New York. doi: [https://doi.org/10](https://doi.org/10.1007/978-1-4612-1950-7)  
 328 [.1007/978-1-4612-1950-7](https://doi.org/10.1007/978-1-4612-1950-7)
- 329 Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London:  
 330 Springer. doi: <https://doi.org/10.1007/978-1-4471-3675-0>
- 331 Dong, S., Sun, Y., Li, C., Zhang, X., Min, S.-K., & Kim, Y.-H. (2021). Attribu-  
 332 tion of Extreme Precipitation with Updated Observations and CMIP6 Sim-  
 333 ulations. *Journal of Climate*, *34*, 871–881. doi: [https://doi.org/10.1175/](https://doi.org/10.1175/JCLI-D-19-1017.1)  
 334 [JCLI-D-19-1017.1](https://doi.org/10.1175/JCLI-D-19-1017.1)
- 335 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., &  
 336 Taylor, K. E. (2016). *Overview of the coupled model intercomparison project*  
 337 *phase 6 (cmip6) experimental design and organization* [dataset]. Geoscientific  
 338 Model Development. Retrieved from [https://gmd.copernicus.org/](https://gmd.copernicus.org/articles/9/1937/2016/)  
 339 [articles/9/1937/2016/](https://gmd.copernicus.org/articles/9/1937/2016/) doi: <https://doi.org/10.5194/gmd-9-1937-2016>
- 340 Fischer, E. M., & Knutti, R. (2015, June). Anthropogenic contribution to global  
 341 occurrence of heavy-precipitation and high-temperature extremes. *Nature Cli-*  
 342 *mate Change*, *5*(6), 560–564. doi: <https://doi.org/10.1038/nclimate2617>
- 343 Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland,  
 344 S., ... Vose, R. S. (2021). Changing state of the climate system [Book Sec-  
 345 tion]. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical*  
 346 *science basis. contribution of working group I to the sixth assessment report of*  
 347 *the intergovernmental panel on climate change* (chap. 2). Cambridge, United  
 348 Kingdom and New York, NY, USA: Cambridge University Press.
- 349 Hannart, A., & Naveau, P. (2018). Probabilities of causation of climate changes.  
 350 *Journal of Climate*, *31*, 5507–5524. doi: [https://doi.org/10.1175/JCLI-D-17](https://doi.org/10.1175/JCLI-D-17-0304.1)  
 351 [-0304.1](https://doi.org/10.1175/JCLI-D-17-0304.1)
- 352 Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., & Ghil, M. (2016). Counterfactual  
 353 causality theory for the attribution of weather and climate-related events. *Bul-*  
 354 *letin of the American Meteorological Society*, *97*, 99–110. doi: [https://doi.org/](https://doi.org/10.1175/BAMS-D-14-00034.1)  
 355 [10.1175/BAMS-D-14-00034.1](https://doi.org/10.1175/BAMS-D-14-00034.1)
- 356 Hegerl, G., & Zwiers, F. (2011). Use of models in detection and attribution of cli-  
 357 mate change. *Wiley interdisciplinary reviews: climate change*, *2*(4), 570–591.  
 358 doi: <https://doi.org/10.1002/wcc.121>
- 359 Härdle, W. (1991). *Smoothing techniques : With implementation in s*. Springer-  
 360 Verlag, New York: Springer Series in Statistics. doi: [https://doi.org/10.1007/](https://doi.org/10.1007/978-1-4612-4432-5)  
 361 [978-1-4612-4432-5](https://doi.org/10.1007/978-1-4612-4432-5)
- 362 IPCC. (2014). *Climate change 2014: Mitigation of climate change. contribution of*  
 363 *working group III to the fifth assessment report of the intergovernmental panel*  
 364 *on climate change* (O. Edenhofer et al., Eds.). Cambridge University Press,  
 365 Cambridge and New York, NY.
- 366 Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hy-  
 367 drology. *Advances in Water Resources*, *25*(8-12), 1287. doi: [https://doi.org/10](https://doi.org/10.1016/S0309-1708(02)00056-8)  
 368 [.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)
- 369 Kiriliouk, A., & Naveau, P. (2020). Climate extreme event attribution using mul-  
 370 tivariate peaks-over-thresholds modeling and counterfactual theory. *An-*  
 371 *als of Applied Statistics*, *14*(3), 1342-1358. doi: [https://doi.org/10.1214/](https://doi.org/10.1214/20-AOAS1355)  
 372 [20-AOAS1355](https://doi.org/10.1214/20-AOAS1355)
- 373 Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in an-  
 374 nual extremes of daily temperature and precipitation in cmip6 models. *Journal*  
 375 *of climate*, *.*, 3441–3460. doi: <https://doi.org/10.1175/JCLI-D-19-1013.1>
- 376 Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011, February). Human

- 377 contribution to more-intense precipitation extremes. *Nature*, 470(7334), 378–  
378 381. doi: <https://doi.org/10.1038/nature09763>
- 379 Naveau, P., Hannart, A., & Ribes, A. (2020). Statistical Methods for Extreme Event  
380 Attribution in Climate Science. *Annual Reviews of Statistics and its Applica-*  
381 *tion*. doi: <https://doi.org/10.1146/annurev-statistics-031219-041314>
- 382 Naveau, P., Ribes, A., Zwiers, F., Hannart, A., Tuel, A., & Yiou, P. (2018, May).  
383 Revising Return Periods for Record Events in a Climate Event Attribution  
384 Context. *Journal of Climate*, 31(9), 3411–3422. doi: <https://doi.org/10.1175/JCLI-D-16-0752.1>
- 385
- 386 Naveau, P., & Thao, S. (2022). Multimodel Errors and Emergence Times in Climate  
387 Attribution Studies. *Journal of Climate*, 35(14), 4791–4804. (Publisher: Amer-  
388 ican Meteorological Society Section: Journal of Climate) doi: [https://doi.org/](https://doi.org/10.1175/JCLI-D-21-0332.1)  
389 [10.1175/JCLI-D-21-0332.1](https://doi.org/10.1175/JCLI-D-21-0332.1)
- 390 Pfahl, S., O Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional  
391 pattern of projected future changes in extreme precipitation. *Nature Climate*  
392 *Change*, 7, 423–427. doi: <https://doi.org/10.1038/nclimate3287>
- 393 Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J., van Old-  
394 enborgh, G. J., . . . Zwiers, F. W. (2016). Attribution of extreme weather  
395 and climate-related events. *WIREs Climate Change*, 7(1), 23–41. doi:  
396 <https://doi.org/10.1002/wcc.380>
- 397 Tandon, N. F., Zhang, X., & Sobel, A. H. (2018). Understanding the Dynamics of  
398 Future Changes in Extreme Precipitation Intensity. *Geophysical Research Let-*  
399 *ters*, 45, 2870–2878. doi: <https://doi.org/10.1002/2017GL076361>
- 400 Worms, J., & Naveau, P. (2022). Record events attribution in climate studies. *Envi-*  
401 *ronmetrics*, 33(8). doi: <https://doi.org/10.1002/env.2777>