Integration of physical bound constraints to alleviate shortcomings of statistical models for extreme temperatures

Robin Noyelle, Yoann Robin, Philippe Naveau, Pascal Yiou, Davide Faranda

Integration of physical bound constraints to alleviate shortcomings of statistical models for extreme temperatures

Robin Noyelle[1], Yoann Robin[1], Philippe Naveau [1], Pascal Yiou[1], and Davide Faranda[1,2,3]

[1]Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL & Université Paris Saclay, Gif-sur-Yvette, 91191,France
[2]London Mathematical Laboratory, 8 Margravine Gardens London, W6 8RH, London, United Kingdom
[3]Laboratoire de Météorologie Dynamique/IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, 75005, Paris, France

Key Points:

- Statistical models based on extreme value theory underestimate tail risks for extreme temperatures.
- Estimates of physical upper bounds for extreme surface temperatures are 3–8°C higher than the statistical upper bounds.
- We propose a method to physically constrain extreme value theory estimates of very extreme temperature events.

Corresponding author: Robin Noyelle, robin.noyelle@lsce.ipsl.fr

Abstract

Heatwaves have devastating impacts on societies and ecosystems. Their frequencies and intensities are increasing globally with anthropogenic climate change. Statistical models using Extreme Value Theory (EVT) have been used for quantifying risks of extreme temperatures but recent very intense events have cast doubt on their ability to represent the tail probabilities of temperatures. Using outputs from a large ensemble of a climate model, we show that physics-based estimates of the upper-bound of temperatures in the mid-latitudes are 3–8°C higher than suggested by EVT-based models. We propose a new method to bridge the gap between the physical and statistical estimates by forcing the EVT-based models to have an upper bound coherent with the bound provided by the instability of the air column. We show that our method reduces the underestimation of tail risks while not deteriorating the performance of the statistical models on the core of the distribution of extreme temperatures.

Plain Language Summary

The usual application of extreme value theory to determine the probabilities of extreme temperatures tends to underestimate the risks. We propose a method to reduce this underestimation by imposing a physically relevant upper bound estimate based on the stability of the air column. The coupling between extreme value theory and physics knowledge alleviates many of the shortcomings of usual statistical approaches used in heatwave attribution.

1 Introduction

Heatwaves and their impacts have been the focus of extensive attention in the context of global climate change (National Academies of Sciences Engineering and Medicine, 2016; Pörtner et al., 2022). There is indeed a clear signal pointing towards increasing frequency and intensity for these events worldwide (Seneviratne et al., 2021). Human bodies (Campbell et al., 2018; Breshears et al., 2021; Huang et al., 2022; Masselot et al., 2023), plants (Hatfield & Prueger, 2015; Brás et al., 2021), ecosystems (Bastos et al., 2021) and infrastructures (Zuo et al., 2015) have a limited capacity to cope with extreme temperatures and can suffer large damages once certain thresholds are reached. The capacity to predict and anticipate future maximum intensities of heatwaves is therefore of primary importance for adaptation to climate change.

A simple way to quantify the risks associated with extreme temperatures is to run a long simulation with a climate model to sample more extreme events than those observed in the past. Apart from the inherent limitations of models to represent correctly the entire temperature distribution — especially at its tail (Naveau et al., 2018; Vautard et al., 2020; Van Oldenborgh et al., 2022; Patterson, 2023; Vautard et al., 2023) — this method is limited by the fact that estimating precisely small probabilities requires extremely long simulations (Wouters & Bouchet, 2016), which is too costly for most complex — and therefore realistic — models. Various approaches have been suggested to address this problem, like rare events algorithms (Ragone et al., 2018; Yiou & Jézéquel, 2020; Ragone & Bouchet, 2021) or ensemble boosting (Gessner et al., 2021; E. M. Fischer et al., 2023), but, although less costly, they still require extensive climate simulations.

An other classical approach is to measure the upper tail probabilities of temperatures distribution — typically the risk of the yearly maximum of daily maximum 2-m air temperature reaching a certain level — using results from Extreme Value Theory (EVT). EVT is a mathematical theory based on a convergence principle of the probability distribution of block maxima (Coles et al., 2001; Beirlant et al., 2005) towards a Generalized Extreme Value (GEV) distribution. It has been used to devise statistical models

for maxima (or minima) of climate variables of interest (Ghil et al., 2011; Katz et al., 2002), such as temperatures. It is extensively used by attribution methods to compare the probabilities of reaching extreme levels between a counterfactual world without climate change and a factual world with climate change (Hannart & Naveau, 2018; S. Philip et al., 2020; Naveau et al., 2020; Kiriliouk & Naveau, 2020; Van Oldenborgh et al., 2021; Worms & Naveau, 2022).

A GEV distribution has three parameters: the location μ, scale σ and shape ξ parameters. Fitting such a distribution to extreme temperatures usually gives estimates of the shape parameter that are negative (Van Oldenborgh et al., 2022; Auld et al., 2021). This implies that the distribution of annual maximum temperatures is bounded upwards by a value:

$$B(\mu, \sigma, \xi) = \mu - \frac{\sigma}{\xi} \tag{1}$$

which depends on the parameters of the distribution. Observing a record temperature above this upper bound should then have a null probability, i.e. it should be impossible. The upper bound is an important quantity insofar as it provides an ultimate worst case scenario for societal adaptation (Palmer, 2002; Fischbacher-Smith, 2010; Scher et al., 2021). An underestimation of this quantity may therefore have daunting consequences in impact studies.

Recent intense heatwaves such as in 2010 in Russia (Di Capua et al., 2021), 2019 in Western Europe (Mitchell et al., 2019) and 2021 in the Pacific Northwest (S. Y. Philip et al., 2021; Thompson et al., 2022) have nonetheless challenged the reliability of the estimation of the statistical upper bound by breaking it sometimes by a large margin (E. Fischer et al., 2021). Several issues indeed arise when applying EVT to empirical data: the pre-asymptotic nature of these data (Gomes, 1984; Coles et al., 2001), the parametric hypotheses made for taking into account the non-stationarity induced by global warming and the limited number of points to fit the distribution. For extreme temperatures, the latter is probably the most problematic insofar as one usually has no more than 70-80 historical years to fit a GEV distribution. The mis-specification of the statistical GEV model, upper bound included, may therefore lead to an underestimation of the tail risks of very intense events (Diffenbaugh, 2020; Zeder et al., 2023).

The physical mechanisms leading to heatwaves in the mid-latitudes occur at various spatial and temporal scales (Perkins, 2015; Horton et al., 2016; Domeisen et al., 2022; Barriopedro et al., 2023). In the mid- and upper-level troposphere, these events are associated to a quasi-stationary high-amplitude Rossby wave (Petoukhov et al., 2013) which can be embedded in a hemispheric pattern (Coumou et al., 2014; Kornhuber et al., 2020; Di Capua et al., 2021). A blocking anticyclone is situated above the heatwave region sustaining the poleward advection of warm air along its western flank and causing adiabatic warming by subsidence and clear skies at its center. The latter causes warming through short-wave insolation of the lower layers of the atmosphere, which can be amplified if soils are dessicated and water is limited for plants evapotranspiration. This results in the allocation of incoming solar energy towards sensible rather than latent heat, increasing further surface air temperature (Seneviratne et al., 2006; E. M. Fischer et al., 2007; Seneviratne et al., 2010; Hirschi et al., 2011; Miralles et al., 2012, 2014; Rasmijn et al., 2018; Dirmeyer et al., 2021).

If the anticyclone stays stationary, extreme temperatures can be reached. High temperature increases are fundamentally limited by the moist convective instability of the air column (Zhang & Boos, 2023), which defines a relevant upper bound for surface temperatures in the mid-latitudes. During heatwaves, surface air temperatures reach their maximal values when the air column is neutrally stratified. Accordingly, temperatures drop when precipitations are triggered by the convective instability of the boundary layer.

As a consequence, one can derive an estimate of the maximum reachable temperature at the surface from characteristics of the free-troposphere (temperature and geopotential at 500hPa) and surface specific humidity. This allows to propose an estimate of the physical upper bound B_ϕ of temperature for mid-latitudes regions, for which Noyelle et al. (2023) showed that it is 5–10°C higher than statistical estimates of the upper bound in a case study for Western Europe. Additionally, as shown by Zhang and Boos (2023) this upper bound increases with global warming.

In this paper, we address the issue of the underestimation of tail risks for extreme temperatures by bridging the gap between the physical and statistical theories of extreme temperatures. We propose to estimate statistical GEV models on time series of yearly maxima where the upper bound of the distribution is imposed by the physical knowledge of the system:

$$B(\mu, \sigma, \xi) = B_\phi. \tag{2}$$

This reduces the space parameter dimension by one. We show that this method estimates correctly the GEV distribution parameters and reduces the underestimation bias of traditional fitting methods.

This paper is organized as follows. Section 2 presents the climate data and the statistical models used. In this section we detail how we compute an estimate of the physical upper bound for surface temperatures. The results of the analysis are presented in section 3. We first demonstrate the inadequacy of the estimation of the upper bound with traditional methods. Then we show the relevance of our approach on synthetic GEV data and on a large ensemble of climate model outputs. Finally, we discuss the results in section 4.

2 Data and methods

2.1 Climate model data

Here we use the large ensemble of the state-of-the-art IPSL-CM6A-LR model (Boucher et al., 2020) with the CMIP6 configuration under an extended historical simulations spanning historical forcing over the period 1850-2014 and RCP2-4.5 forcing over the period 2014-2059. The ensemble encompasses 31 independent members. For each member and each year, the GMST is computed as the area weighted global mean 2-m air temperature. For each member and each grid point over the Northern Hemisphere (30°N-80°N) we extract the yearly maximum of daily maximum 2-m air temperature. To compute the physical upper bound of surface temperature B_ϕ we additionally consider the yearly maximum over the months June, July, August and September of the daily mean air temperature at 500hPa (T_{500}), daily mean geopotential height at 500hPa (Z_{500}) and yearly minimum of daily mean 2-m air specific humidity. For the period 1850-1949, only 10 members are available for the variables T_{500} and Z_{500}. At each grid point, we therefore resample randomly from these 10 members to create data series of yearly maxima of T_{500} and Z_{500}.

The use of climate data that we make here is not intended to give precise and relevant values for adaptation purposes but merely as an illustration of our method on physically relevant data points.

2.2 Statistical models

Classically, the distribution of yearly maxima of daily maximum 2-m air temperature (TXx) are modelled using a member of the so-called Generalized Extreme Value

(GEV) family of distributions. The probability of a yearly maxima Z to be below a threshold z is expressed as $P(Z \leq z) = G(z)$ where:

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi(\frac{z-\mu}{\sigma})\right]^{-1/\xi}\right\}, & \text{if } \xi \neq 0, \\ \exp\left\{-\exp\left[-(\frac{z-\mu}{\sigma})\right]\right\}, & \text{if } \xi = 0, \end{cases} \tag{3}$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. We denote this model $\mathcal{M}_0$. The use of this family of distributions is motivated by the extremal theorems which state that at a suitable limit of an infinite number of samples, the distribution of maxima converges towards a member of the GEV family (Coles, 2001). It is therefore implicitly assumed that the pre-asymptotic distribution of yearly maxima can be correctly represented by this distribution.

To account for non-stationarity in the temperatures distribution, one can first let the location parameter depend linearly on GMST:

$$\mu = \mu_0 + \mu_1 \text{GMST}. \tag{4}$$

and the scale parameter σ is constant. We denote this model $\mathcal{M}_1$. We also consider a statistical model $\mathcal{M}_2$ where the scale parameter is also dependent on GMST:

$$\begin{cases} \mu = \mu_0 + \mu_1 \text{GMST} \\ \sigma = \log\left(1 + \exp(\sigma_0 + \sigma_1 \text{GMST})\right). \end{cases} \tag{5}$$

The functional form for the scale parameter assures that it stays positive and – although it is non-linear – for typical values that we consider here the scale parameter depends almost linearly on GMST. Importantly, for all these models if the shape parameter ξ is negative, the distribution is bounded upwards and the bound B depends on the three parameters (cf. Eq.1). We use the maximum likelihood method to provide estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ (and correspondingly for non-stationary ones) of the parameters of the GEV distribution. The estimator for the bound is then:

$$\hat{B} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}. \tag{6}$$

Note that in the non-stationary cases, the estimate $\hat{B}$ depends on the covariate GMST and therefore on time.

In the following, we mimic the practice of attribution studies (S. Philip et al., 2020) by fitting the statistical models with one ($\mathcal{M}_1$) and two non-stationary parameters ($\mathcal{M}_2$) on 70 TXx randomly resampled from the pool of the 31 climate model members over the period 1945-2014 using the ensemble mean GMST as a covariate (see Fig. B1a for its time evolution). The choice to use only 70 points was made in reference to the average number of data points available for most weather stations (which were installed after 1945). To obtain distributions of the quantities of interest displayed below, we repeat the fitting procedure 1000 times.

2.3 Computation of the physical upper bound

As in Zhang and Boos (2023) and Noyelle et al. (2023), the maximum reachable surface temperature is computed as the surface temperature for which the moist static energy of the surface air parcel is equal to the saturated moist static energy of the free-

troposphere (considered to be at 500hPa):

$$T_{s,max} = T_{500} + \frac{L_v}{c_p}(Q_{sat}(T_{500}) - Q) + \frac{g}{c_p}(Z_{500} - Z_s). \tag{7}$$

where T_{500} is the air temperature at 500hPa, Q the surface specific humidity of the air parcel, Z_{500} the geopotential height at 500hPa and Z_s the elevation of the surface. $T_{s,max}$ is computed from these four variables using the latent heat of vaporization L_v, the specific heat of air at constant pressure c_p, the gravitational constant g and the equation $Q_{sat}(T_{500}) \simeq \frac{\epsilon e_{sat}(T_{500})}{500\text{hPa}}$ where $e_{sat}(T_{500})$ is the saturation vapor pressure given by the Clausius-Clapeyron relation.

The relation 7 is maximal when T_{500} and Z_{500} are maximized and Q is minimized. However, by plugging the yearly maximum value of the two first quantities and the yearly minimum of the last one, it is possible that we either overestimate the physical upper bound B_ϕ — because there could not be a meteorological situation that combines the extremized values of these three quantities at the same time — or that we underestimate it — because the yearly natural variability may not provide sufficiently extreme values of these three quantities.

To propose a reasonable value of the physical upper bound B_ϕ for each resampling of n values of TXx that may be computable in practice, we also resample the corresponding n yearly maxima of T_{500} ($TX_{500,i}$) and Z_{500} ($ZX_{500,i}$), and n yearly minima of Q (QN_i). For each year $1 \leq i \leq n$ resampled we compute the physical upper bound $B_{\phi,i}$ using equation 7 with the corresponding yearly maxima of T_{500} and Z_{500}. However, we treat differently the air specific humidity Q variable which, as shown by Noyelle et al. (2023), is critical to have a relevant estimation of the upper bound. For this variable we use the same value for each of the n resampled years. This value is defined as the covariate-trend adjusted minima $\widetilde{QN}_{min}$ of the $(QN_i)_{1\leq i\leq n}$ over the n resampled years. If the trend on QN is non significant, we use simply the minima of Q over the n resampled years. This procedure gives us n values of the physical upper bound $B_{\phi,i}$:

$$B_{\phi,i} = T_{s,max}(TX_{500,i}, ZX_{500,i}, \widetilde{QN}_{min}). \tag{8}$$

We then regress the $B_{\phi,i}$ on the covariate GMST (or RMST) which gives us a linear estimate for the physical upper bound

$$B_\phi(\text{GMST}) = A + B \times \text{GMST}. \tag{9}$$

During the fit of the statistical models presented in the preceding subsection, we then impose that the statistical upper bound be equal to the physical upper bound. Doing so is equivalent to imposing a constraint on the GEV parameters.

This procedure is straightforward to apply and gives a reasonable estimate for the physical upper bound. It should nonetheless be stated that this value constitutes only a rough estimate of the true upper bound and is subject to uncertainties. We note that our procedure can be more easily conceptualized in a bayesian context for which one could define a prior distribution on B_ϕ(GMST) (Robin & Ribes, 2020).

3 Results

3.1 Shortcomings of the traditional fitting approach

We first show the inadequacy of the traditional fitting approach for estimating tail probabilities of yearly maximum of daily maximum temperature (TXx) with a GEV distribution (see also E. M. Fischer et al. (2023); Zeder et al. (2023)). To do so, we use the

outputs of the large ensemble (31 members). We fit the statistical models with one ($\mathcal{M}_1$) and two non-stationary parameters ($\mathcal{M}_2$) on 70 TXx randomly resampled from the pool of the 31 members over the period 1945-2014 using the ensemble mean GMST as a covariate. We then investigate whether the upper bounds B_1 and B_2 estimated with the two models are "true" upper bounds, i.e. whether they are exceeded on the full dataset over the period 1850-2059 for the 31 members. By doing so 1000 times, we have an estimate of the probability that the estimated upper bounds of the two models will be exceeded at least once (see Appendix A for the detail of the computation).

Figure 1ab shows the result of this computation for every land grid points in the Northern Hemisphere mid-latitudes (30N-80N). For most of the grid points, the probability to exceed the upper bound estimated on the period 1945-2014 is above 40%. Remarkably, there is not a single grid point over the region studied for which the probability to exceed the upper bound is equal to 0 after 1000 resamplings. Model $\mathcal{M}_1$ (first column) performs better for this test than model $\mathcal{M}_2$ (second column) but both statistical models largely underestimate the value of the upper bound. This shows that, at least on climate model outputs, the estimation of the upper bound of the TXx distribution is not reliable. Figure 1cd gives an estimation of the empirical return time that corresponds to the values of the bounds B_1 and B_2. The return time — which should be infinite — is in practice closer to 2000 years for model $\mathcal{M}_1$ (Fig. 1c) and to 200 years for model $\mathcal{M}_2$ (Fig. 1d). We also note that, for model $\mathcal{M}_1$, large regions in the eastern side of continental masses show a large return time of the upper bound (above 6000 years, colored in light blue), which suggests that for these regions the upper bound is only slightly underestimated. Finally, the ratio between the GEV and the empirical probabilities of a centennial event is depicted in Figure 1ef. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate (see Appendix A for the detail of the computation). For both models, the intensity of events which are predicted to be centennial according to the GEV estimate actually happen every 25 to 50 years, which make them 2 to 4 times more likely than predicted. Apart for a few grid points, this underestimation occurs almost everywhere.

Increasing the size of the sample for fitting the models only slightly improves the situation (see Fig. B2 for 100 years and Fig. B3 for 150 years) and for most grid points the probability to exceed the upper bound estimated is above 20% even with 150 years of data. Similarly, changing the fitting period towards a period with more variation in the covariate does not change the results significantly (Fig. B4 with 70 points resampled over the period 1956-2025). Finally, using the Regional Mean Surface Temperature (RMST) rather than GMST as a covariate only marginally changes the results (Fig. B1bcd for the regional covariates, and Fig. B5 for the results of the fits). Figure B6 shows that the statistical estimation of the upper bound is lower than the physical estimation of the upper bound by a margin of 3 to 8°C. Except for a small region in Western Russia, the statistical upper bound is systematically lower than the physical upper bound.

3.2 Imposing the upper bound on synthetic GEV data

We first demonstrate the interest of the method of imposing the upper bound in the GEV fit using synthetic data with parameters typical of the parameters found when fitting a GEV on TXx data (see Appendix A for the detail). We simulate 1000 series of 50, 100 and 200 points distributed according to the model $\mathcal{M}_1$, i.e. with a non-stationarity on the location parameter only. We then fit the full model $\mathcal{M}_1$ and the model $\mathcal{M}_1^B$ where the (correct) upper bound is imposed during the fit. For comparison purposes, we additionally fit three other models derived from $\mathcal{M}_1$ where either the (correct) location ($\mathcal{M}_1^\mu$), scale ($\mathcal{M}_1^\sigma$) or shape parameter ($\mathcal{M}_1^\xi$) is imposed to the model.

Figure B8 shows the results for the estimation of the different parameters (Fig. B8abcd) and several metrics (Fig. B8efghi). All models estimate correctly the location param-

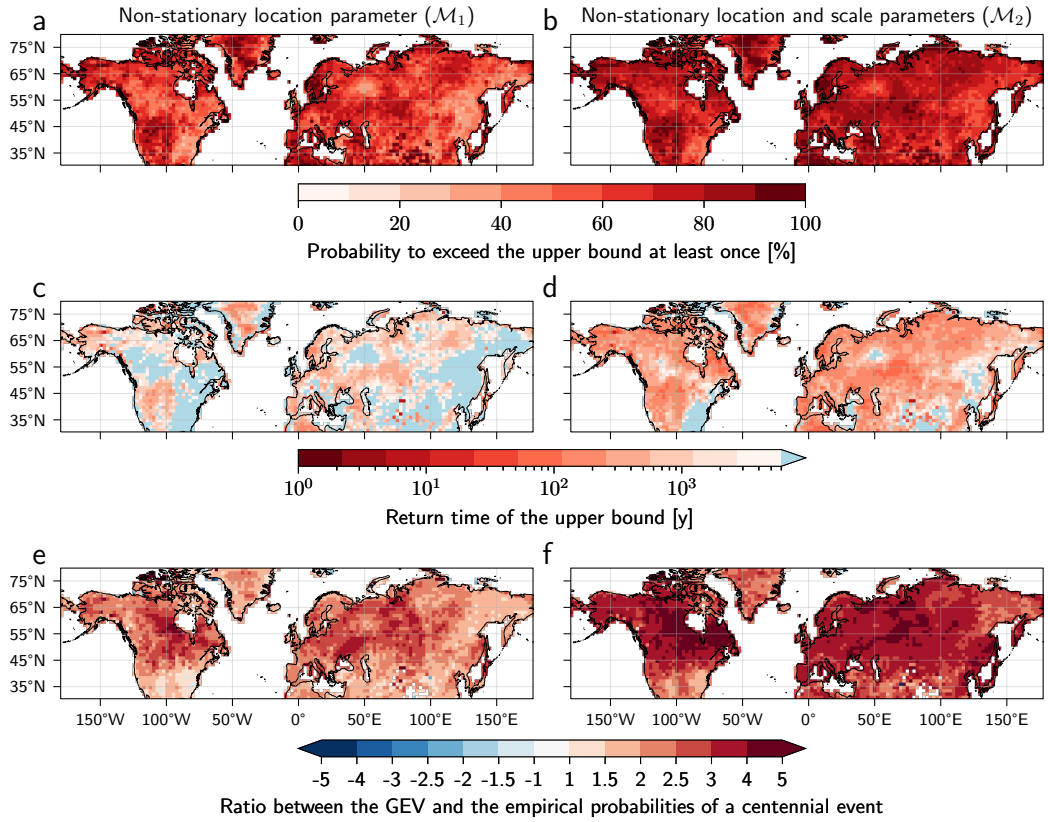

Figure 1. Bias of tail risks of extreme temperatures with the traditional GEV fitting approach. The first row (panels a and b) shows the empirical probability that the statistical upper bound estimated using 70 points resampled on the period 1945-2014 over all 31 members is exceeded at least once on the full dataset (1850-2059). The second row (panels c and d) shows the estimated return time of this upper bound on the full dataset. Grid points where only one TXx exceeds the upper bound (return time above 6000 years) are colored in light blue. The third row (panels e and f) shows the ratio between the GEV and the empirical probabilities of a centennial event. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate. The first column shows the results for the statistical model with a linear dependence on GMST of the location parameter only ($\mathcal{M}_1$) and the second column for the statistical model with a linear dependence of both the location and scale parameters ($\mathcal{M}_2$).

eters μ_0 and μ_1 — with a slight overestimation of the first one and underestimation of the second one — but they almost all underestimate the scale parameter (Fig. B8c). Models $\mathcal{M}_1$, $\mathcal{M}_1^\mu$ and $\mathcal{M}_1^\sigma$ underestimate the shape parameter and there is a large spread around the true value (Fig. B8d). On the contrary, when we impose the upper bound the estimation of the shape parameter is remarkably precise. As a consequence, the upper bound is systematically underestimated for models $\mathcal{M}_1$, $\mathcal{M}_1^\mu$ and $\mathcal{M}_1^\sigma$. It is also the case for the model $\mathcal{M}_1^\xi$ but the estimation is much more precise. We note here the paramount importance of correctly estimating the shape parameter for correctly estimating the upper bound and vice versa.

The return level at 1% (Fig. B8f) is underestimated and the corresponding return time (Fig. B8g) is overestimated for all models but the model with the upper bound imposed perform best both in bias and in spread. When comparing models in terms of the norms L_∞ and L_2 with respect to the true Cumulative Distribution Function (CDF), the model $\mathcal{M}_\mu$ performs best but the other models have similar performance. Increasing the sample size improves the estimation for all models, but for the estimation of tail probabilities the model $\mathcal{M}_1^B$ remains the best (e.g. Fig. B8g).

The SI provides similar estimations for the models $\mathcal{M}_0$ (Fig. B9) and $\mathcal{M}_2$ (Fig. B8) for which we have similar results as those presented here. Our results demonstrate that once one knows the upper bound of a bounded GEV distribution, imposing the value of the bound during the fit gives the best fitting results on the metrics displayed here. Nonetheless, when it comes to non-synthetic data, it is likely that we will make an error on the imposed value of the upper bound — which is unknown. To investigate this issue, we additionally fitted the model $\mathcal{M}_1^B$ with an error B_{err} on the bound B. Results are presented in Appendix B for an error of $B_{err} = +2.5$ (Fig. B11) and $B_{err} = +5$ (Fig. B12). In these cases, the scale parameter (Fig. B11c and S12c) tend to be underestimated and the shape parameter (Fig. B11d and B12d) to be overestimated. Accordingly, the return level at 1% (Fig. B11fc and B12f) is overestimated and the return time for the true return level at 1% is underestimated (Fig. B11g and B12g). We nevertheless note that the error made is small when compared to the spread around the estimation for the other models and that for $B_{err} = 2.5$ the overestimation of the return time is of the same order of magnitude as the underestimation of the full model $\mathcal{M}_1$. On the rest of the distribution, the performance of the biased model $\mathcal{M}_1^{B+B_{err}}$ is comparable to the other models (Fig. B11hi and B12hi). For the quality of the fit it is therefore preferable to impose an upper bound as close as possible to the correct upper bound but errors up to +5°C can be tolerated even though they tend to overestimate events at the upper tail of the distribution. We come back to this question in the discussion section.

3.3 Imposing the upper bound on a climate model outputs

We apply our approach to temperatures simulated by a climate model and determine the distribution of yearly maxima of daily maximum temperature at one grid point in Western Europe (lon=0°, lat=49.44°N) for the 31 members of the IPSL model over the period 1850-2059. Figure 2a shows the evolution of these maxima with time. We note in particular that there is an extreme outlier on the year 2019 which reaches the value of 38.6°C (shown by a black triangle), shattering the previous record of all members by a margin of 1.9°C. We fit the models $\mathcal{M}_1$ and $\mathcal{M}_2$ with the GMST as covariate by resampling 1000 times over all members during the period 1945–2014 (70 years), as previously. Figure 2a shows the median, 5th and 95th percentiles of the estimated upper bounds B_1 and B_2 for both statistical models along with the estimated physical upper bound B_ϕ. The median estimates of upper bounds B_1 and B_2 are *below* 10 data points of the temperature time series. On the contrary, the physical upper bound is never reached and is above the 95-th percentile of the distribution of the upper bounds B_1 and B_2.

Figure 2b illustrates the issue of the underestimation of the upper bound with the case of the 2019 extreme event. In black we show the empirical distribution of maxima for years 2016-2021, for which the value of the covariate GMST is similar to the one of 2019 (14.32°C vs 14.24 to 14.38°C). As an illustration we show the PDF fitted on member 14 during the period 1945-2014 for models without the upper bound imposed ($\mathcal{M}_1$ and $\mathcal{M}_2$) and with the upper bound imposed ($\mathcal{M}_1^B$ and $\mathcal{M}_2^B$). The PDF of the former are very close and are not distinguishable on the graph. Remarkably, they both estimate an upper bound which is around 1°C *smaller* than the 2019 extreme outlier. To show that we did not choose member 14 for its underestimation of the bound, the two blue boxplots shows the distribution of 1000 estimations of the 2019 upper bound done by resampling as previously. More than 75% of these estimations are below the maximum value of 2019.

Fig. 2c shows a comparison between the estimated return time in the full data set (1850-2059) of the 1% return level using the fits of the different models. Models $\mathcal{M}_1$ and $\mathcal{M}_2$ vastly overestimate this return time: the median return time empirical estimate for the return level at 1% according to their distribution is around 40 years. On the contrary, models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$ underestimate this return time, by around 250 years for the latter and 100 years for the former. This translates into an overestimation of tail risks of extreme temperatures for the two models for which the physical upper bound is imposed. Figure 2de show the performance of the different models on the rest of the distribution using average norms L_∞ and L_2 between the fitted CDF and the empirical CDF (see SI for the detail of the computation). Models have similar performance for both metrics, with a slightly better one for models $\mathcal{M}_1$ and $\mathcal{M}_1^B$. Models where the upper bound is prescribed perform slightly worse compared to traditional models. We provide in Appendix B the same figure using the RMST as covariate (Fig. B13), which gives similar results.

Finally, Figure 3 shows the same results as Fig. 1 for models where the upper bound is prescribed ($\mathcal{M}_1^B$ and $\mathcal{M}_2^B$). Remarkably, Fig. 3ab show that a for a majority of grid points the probability to exceed the upper bound is null after 1000 resamplings (grid points colored in blue). For grid points where the probability is not null, it is close to zero (e.g. in the Iberian peninsula or in North West America). There is however two regions over Western Russia and Central Eurasia where the probability to exceed the upper bound at least once is still above 60 %. Nevertheless, Fig. 3cd show that for the vast majority of these grid points, the return time of the estimated upper bound is very high: above 6000 years which implies that only a single TXx on the dataset is above this value. In Fig. B14 and B15 we provide the fits for example grid points in these two regions. They show that only one TXx in the 19th century is actually above the median estimate of the physical upper bound (respectively in 1850 and 1860). We also note that for those grid points the statistical and physical estimates of the upper bounds are much closer than for the grid point in Western Europe — although the physical upper bound is more precisely estimated.

Additionally, Fig. 3ef show that imposing the upper bound partially alleviates the issue of underestimation of centennial events. Contrary to Fig. 1ef, there is no uniform underestimation of the return level of the centennial events. In particular, coastal regions even tend to overestimate the risk of these events. This effect is probably due to the choice we made to take the lowest value of the surface specific humidity in the resampled data to compute an estimate of the physical upper bound. On the other hand, the overestimation bias over Europe for example is largely reduced and some regions even slightly overestimate the risk. One should note that there are still large regions where our procedure underestimate the risk of centennial events, for example in Canada, Western Russia and Central Eurasia. This effect may be due to historical aerosols forcing which are used by the model and the choice we made to use data between 1945 and 2014 for the fit. This period indeed corresponds to higher anthropogenic aerosols levels in these re-

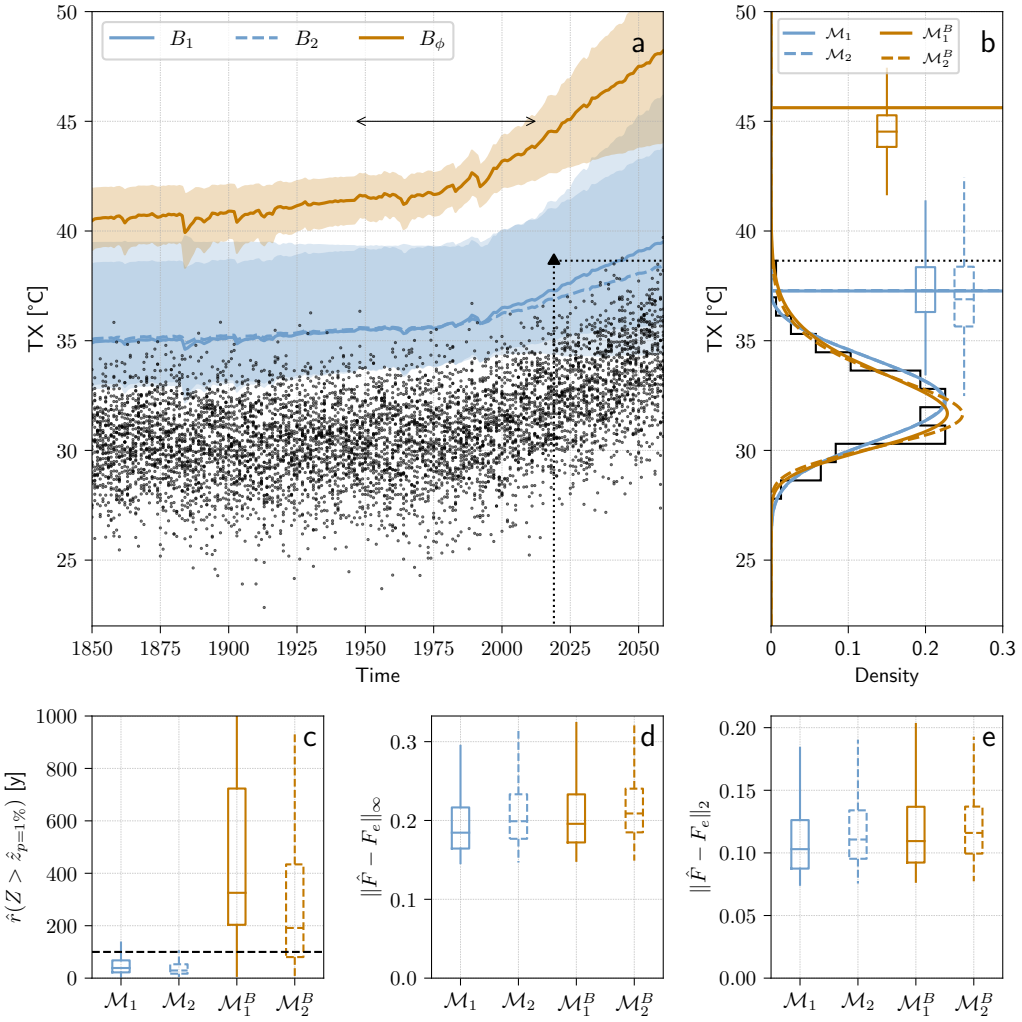

Figure 2. Comparison between the statistical models with and without imposing the upper bound for the IPSL-CM6A-LR climate model outputs. (a) Yearly maxima of daily maximum temperature (TXx) at a grid point in Western Europe (lon=0°, lat=49.44°N) for the 31 members of the IPSL model (black), median, 5 and 95-th percentiles of the distribution of upper bounds B_1 and B_2 (blue plain and dashed) for statistical models $\mathcal{M}_1$ and $\mathcal{M}_2$ for 1000 fits with resampled data points on the period 1945-2014 (black arrow), and median, 5 and 95-th percentiles of the distribution of physical upper bounds B_ϕ (orange). The black triangle shows the value of the 2019 extreme outlier (38.6°C). (b) Histogram of TXx over the period 2016-2021 (black) and PDFs (colors) using the fit on member 14 for statistical models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$. The black dotted line shows the extreme 2019 event and the colored lines shows the physical upper bound and the estimated statistical upper bounds for the fit on member 14 (the blue plain and dashed lines are almost confounded). The boxplots shows the distribution of upper bounds for the different models for 1000 fits on resampled data. (c) Distribution of return time for the return level at 1% for all statistical models. Average norms (d) L_∞ and (e) L_2 between the estimated and the empirical CDFs.

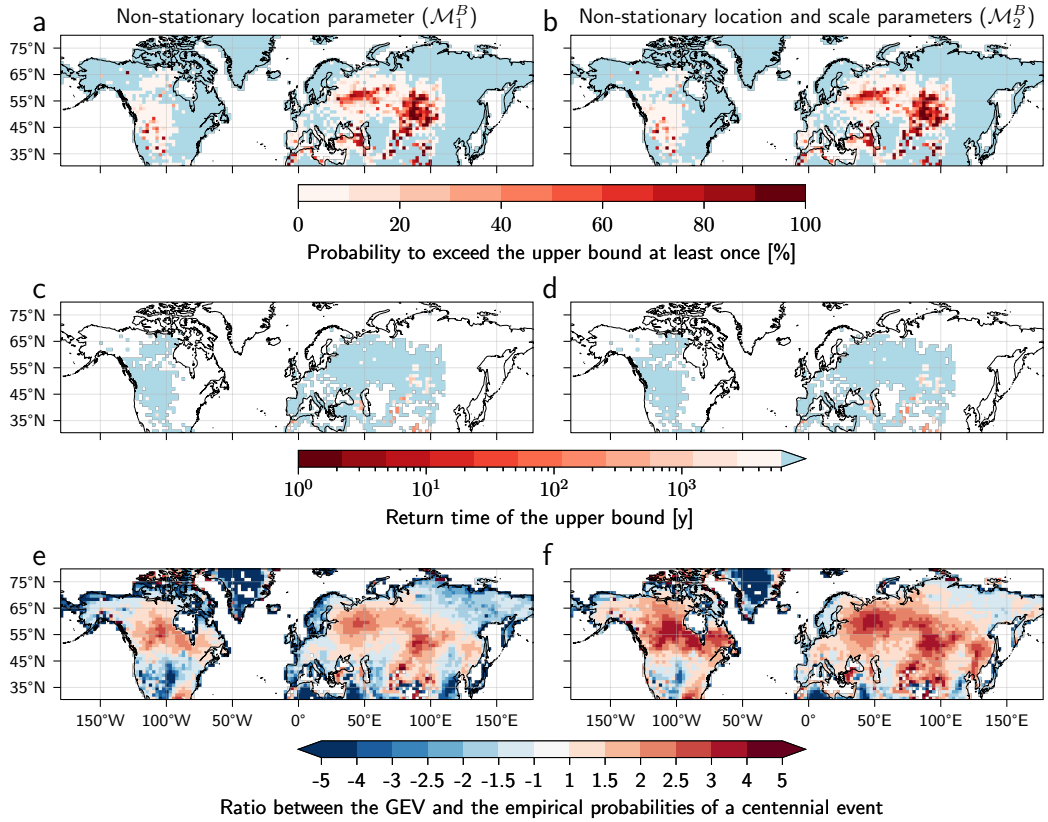

Figure 3. Prescribing the upper bound leads to better estimation of tail risks of extreme temperature. This figure is the same as Figure 1 with models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$. The first row (panels a and b) shows the empirical probability that the statistical upper bound estimated using 70 points resampled on the period 1945-2014 over all 31 members is exceeded at least once on the full dataset (1850-2059). The grid points where the probability is null are colored in light blue. The second row (panels c and d) shows the estimated return time of this upper bound on the full dataset. Grid points where only one TXx exceeds the upper bound (return time above 6000 years) are colored in light blue. Grid points where no TXx exceeds the upper bound are colored in white. The third row (panels e and f) shows the ratio between the GEV and the empirical probabilities of a centennial event. A positive (negative) value gives how much more (less) likely a centennial event occurs in the dataset with respect to the GEV estimate. The first column shows the results for the statistical model with a linear dependence on GMST of the location parameter only ($\mathcal{M}_1^B$) and the second column for the statistical model with a linear dependence of both the location and scale parameters ($\mathcal{M}_2^B$).

gions, which are increasingly recognized to strongly influence the magnitude of extreme temperatures (Sillmann et al., 2013; Dong et al., 2017; Westervelt et al., 2020; Luo et al., 2020; Seong et al., 2021). Nevertheless, the overestimation is reduced to around 2.5 more events in the dataset than according to the GEV estimate (Fig. 3e). Finally, as previously, model $\mathcal{M}_2^B$ performs worse according to this metric than model $\mathcal{M}_1^B$.

4 Discussion and conclusions

We presented a new method to bridge the gap between physical knowledge and statistical estimates of tail probabilities of extreme mid-latitude surface temperatures. We proposed to incorporate a physical estimate of the upper bound in the fitting of GEV distributions for yearly maximum of daily maximum 2-m air surface temperature. We showed on synthetic GEV data and on climate model outputs that doing so leads to more precisely fitted parameters — especially the crucial shape parameter, reduces the underestimation of tail probabilities and does not deteriorate the performance of the fit on the rest of the distribution. The underestimation of the upper bound is largely avoided and we additionally showed that our method improves the estimation of probabilities of centennial events, although this was not intended by the method. Our method also has the advantage of rendering the statistical fit more stable with respect to its tail properties. Figure B7 indeed shows that the standard deviation on the estimated upper bounds is much smaller for models $\mathcal{M}_1^B$ and $\mathcal{M}_2^B$ than for classical models $\mathcal{M}_1$ and $\mathcal{M}_2$. It especially ensures that the upper bound always exists — which even though is clear physically, is not often found in practice. We additionally showed that overestimating the upper bound of the distribution would tend to also overestimate tail probabilities and therefore risks of very high temperatures. This overestimation is however of the same order of magnitude as the underestimation associated with classical fits of GEV distribution.

Other metrics could have been chosen to quantify the quality our fits with respect to data. We nevertheless emphasize that all classical metrics based on the log-likelihood, such as the Akaike information criterion (Akaike, 1998), the Bayesian information criterion (Schwarz, 1978) or the Likelihood ratio test (Coles et al., 2001), are not adapted here because they give an infinitely bad score to classical models. The latter indeed underestimate the upper bound: a model which gives a zero probability to an event which still occurs has an infinite log-likelihood. To the best of our knowledge there is also no consensus on the metrics to evaluate the quality of the fit of a model in a non-stationary context.

Choosing the value of the upper bound to be imposed during the fit is the crucial element for the success of our procedure. The estimation procedure of the physical upper bound B_ϕ that we proposed here likely tends to overestimate this quantity because it combines extremized value of temperature and geopotential at 500hPa and surface specific humidity which are unlikely to happen at the same time in practice. As shown on synthetic GEV data and on climate model outputs, it is therefore possible that we overestimate tail risks at some places. However, given the fact that GEV estimates of tail probabilities may be used by practitioners for adaptation purposes, it is not only a scientific but also an ethical question to choose to over- or underestimate risks of very high temperature events. Because of the large consequences that an underestimation of tail probabilities could have on societies, we here advocate for a scientific choice which avoids this underestimation even though it comes at the risk of overestimation of tail probabilities.

Appendix A Supplementary information

A1 Computation details for Figure 1

A11 Probability of exceeding the upper bound

We here detail the procedure to compute the probability of exceeding the upper bound that is presented in Figure 1. This procedure is applied for each grid point over the Northern Hemisphere mid-latitudes (30N-80N). We first draw randomly 70 TXx from the pooled data of the 31 members between the period 1945-2014 (2170 data points) without replacement. We then fit the statistical model into consideration. If the shape parameter of the fit is negative (i.e. if there is indeed an upper bound), we consider the estimated statistical upper bound $\hat{B}_i$ and we count the number n_i of TXx in the full dataset (1850-2059, 6510 data points) that are above $\hat{B}_i$. Note that for models $\mathcal{M}_1$ and $\mathcal{M}_2$ the estimated upper bound depends on the covariate GMST (or RMST, see below) and therefore on time. We do this procedure $N = 1000$ times and the probability of exceeding the upper bound at least once is then computed as:

$$\hat{P}(Z > \hat{B}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(n_i > 0) \tag{A1}$$

where $\mathbf{1}(n_i > 0)$ is equal to 1 if $n_i > 0$ and equal to 0 otherwise.

A12 Median return time of the upper bound

The median return time of the estimated upper bound is then computed as:

$$\hat{r}(Z > \hat{B}) = \left(\text{Med} \left(\frac{n_i}{N_d} \right) \right)^{-1} \tag{A2}$$

where $N_d = 6510$ is the size of the full dataset and Med is the median operation. $\frac{n_i}{N_d}$ is an estimation of the probability that the yearly maxima is above the bound $\hat{B}_i$ for the fit i. We then compute the median of this probability over the $N = 1000$ resampling and inverse it to give an estimate of the return time of the estimated upper bound.

A13 Ratio between the GEV and the empirical probabilities of a centennial event

We proceed similarly to estimate the return time of a centennial event. For each fit i, we first find the return level $\hat{z}_{p=1\%,i}$ corresponding to the 99th quantile of the fitted GEV PDF. Note that for models $\mathcal{M}_1$ and $\mathcal{M}_2$ the estimated return level depends on the covariate GMST (or RMST, see below) and therefore on time. We then count the number $\tilde{n}_i$ of TXx in the full dataset (1850-2059, 6510 data points) that are above $\hat{z}_{p=1\%,i}$. The median empirical return time of the centennial return level is thus:

$$\hat{r}(Z > \hat{z}_{p=1\%}) = \left(\text{Med} \left(\frac{\tilde{n}_i}{N_d} \right) \right)^{-1} \tag{A3}$$

where as previously $N_d = 6510$ is the size of the full dataset and Med is the median operation. $\frac{\tilde{n}_i}{N_d}$ is an estimation of the probability that the yearly maxima is above the GEV centennial return level $\hat{z}_{p=1\%,i}$ for the fit i. We then compute the median of this probability over the $N = 1000$ resampling and inverse it to give an estimate of the empirical return time of the estimated centennial return level. If the fit corresponds to the true underlying distribution, we should find $\hat{r}(Z > \hat{z}_{p=1\%}) = 100$ by definition of the

centennial return level. However this is not the case in practice and we quantify this difference by dividing by the true return time. The ratio presented in Figure 1 is then computed as:

$$\text{Ratio} = \begin{cases} -\hat{r}(Z > \hat{z}_{p=1\%})/100, & \text{if } \hat{r}(Z > \hat{z}_{p=1\%}) > 100, \\ 100/\hat{r}(Z > \hat{z}_{p=1\%}), & \text{if } \hat{r}(Z > \hat{z}_{p=1\%}) \leq 100. \end{cases} \tag{A4}$$

If for example $\hat{r}(Z > \hat{z}_{p=1\%}) = 50$, the events that are said to be centennial by the statistical model actually occur every 50 years in the dataset. Therefore the ratio is equal to 2: these events are twice *more* likely in the dataset than according to the GEV estimate. On the other hand, if $\hat{r}(Z > \hat{z}_{p=1\%}) = 200$, the events that are said to be centennial by the statistical model actually occur every 200 years in the dataset. Therefore the ratio is equal to -2: these events are twice *less* likely in the dataset that according to the GEV estimate.

A2 Computation of RMST

To define the RMST covariates, we split the Northern Hemisphere mid-latitude regions (30N-80N) into three parts: North America (180W-30W), Europe (30W-50E) and Asia (50E-180E). For each grid points in these three regions, we first compute the area weighted 2-m air temperature average. The RMST is then the ensemble mean of these quantities. The results are presented in Fig. B1bcd.

A3 Computation details for synthetic GEV data

A31 Simulations

For the simulation of synthetic GEV data we use the following parameters:

$$\mathcal{M}_0 \begin{cases} \mu = 23 \\ \sigma = 1.35 \\ \xi = -0.15 \end{cases} \tag{A5}$$

$$\mathcal{M}_1 \begin{cases} \mu_0 = 23 \\ \mu_1 = 1.6 \\ \sigma = 1.35 \\ \xi = -0.15 \end{cases} \tag{A6}$$

$$\mathcal{M}_2 \begin{cases} \mu_0 = 23 \\ \mu_1 = 1.6 \\ \sigma_0 = 1.35 \\ \sigma_1 = 0.1 \\ \xi = -0.15. \end{cases} \tag{A7}$$

A32 Norm metrics

The norm metrics L_∞ and L_2 are computed between the CDFs of the fitted distribution $\hat{F}$ and the true distribution F as such:

$$\|\hat{F} - F\|_\infty = \sup_{t \in R} |\hat{F}(t) - F(t)| \tag{A8}$$

and:

$$\|\hat{F} - F\|_2 = \sqrt{\int_{-\infty}^{+\infty} (\hat{F}(t) - F(t))^2 dt}. \tag{A9}$$

Numerically these two quantities are computed using a sampling at $dt = 10^{-2}$ for t between -200 and 200.

A4 Norm metrics for climate model outputs

The distribution of return times presented in Figure 2 panel c and Figure B13 panel c are computed the same way as for the return time presented above for Figure 1. The norm metrics L_∞ and L_2 are computed between the empirical CDF F_e of data and the fitted CDF $\hat{F}$ for each year (using the 31 members) and then averaged over the full period (1850-2059).

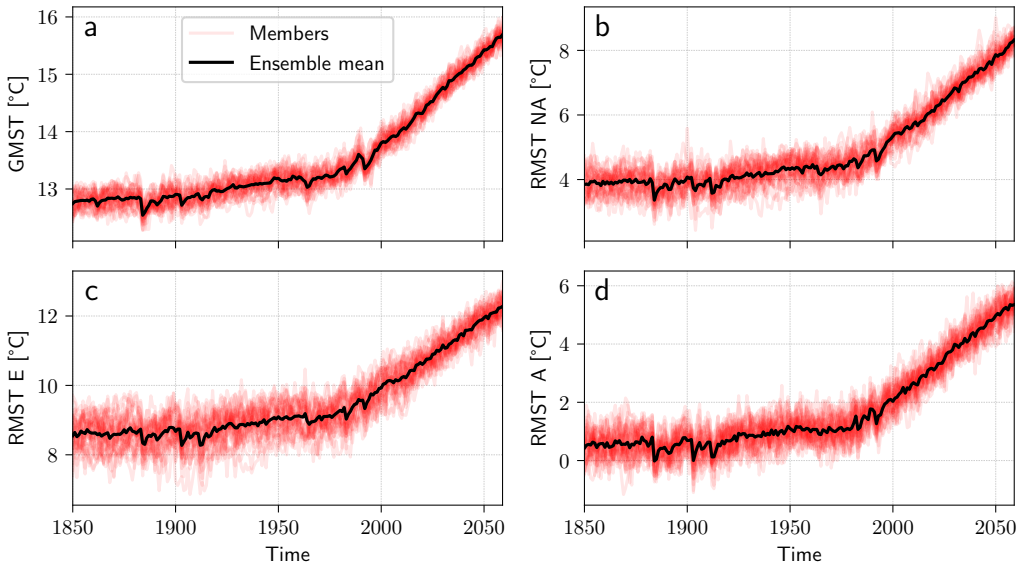

Figure B1. Evolution of covariate for the GEV fit. (a) Ensemble mean (black) and members (red) GMST. (b) Same for RMST over North America (180W-30W, 30N-80N). (c) Same for RMST over Europe (30W-50E, 30N-80N). (d) Same for RMST over Asia (50E-180E, 30N-80N).

Appendix B Additional figures

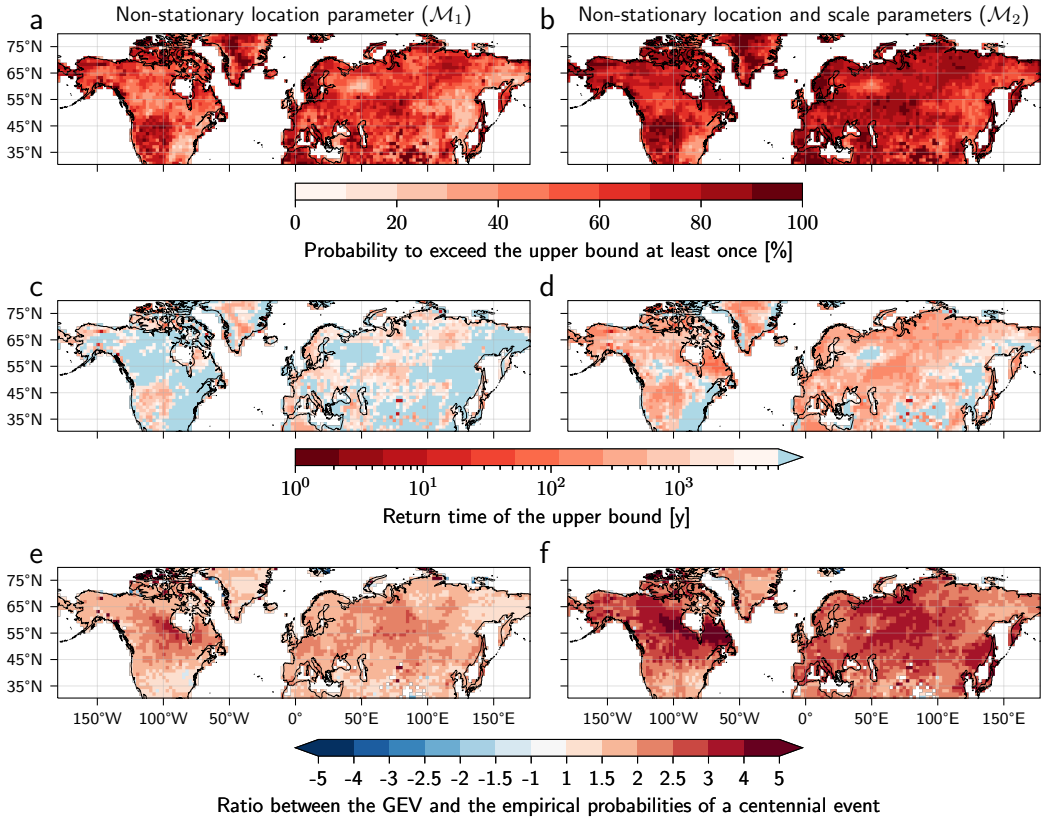

Figure B2. Same as Fig. 1. with 100 points resampled over the period 1915-2014.

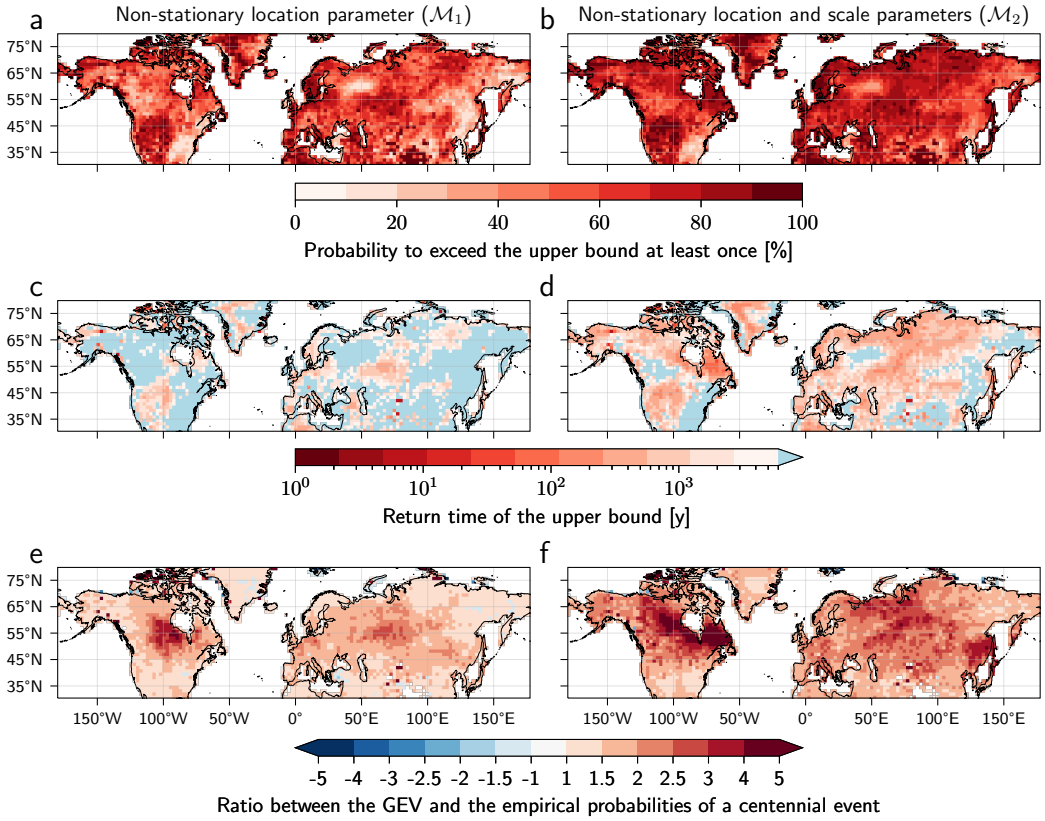

Figure B3. Same as Fig. 1. with 150 points resampled over the period 1865-2014.

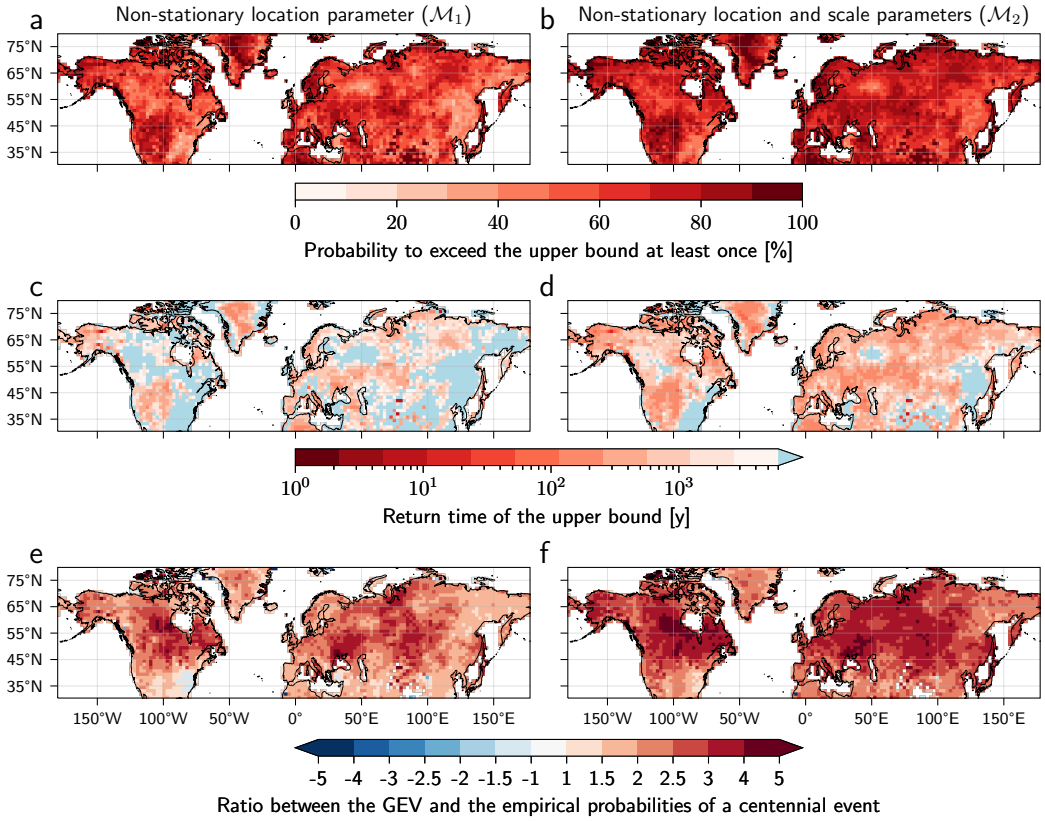

Figure B4. Same as Fig. 1. with 70 points resampled over the period 1956-2025.

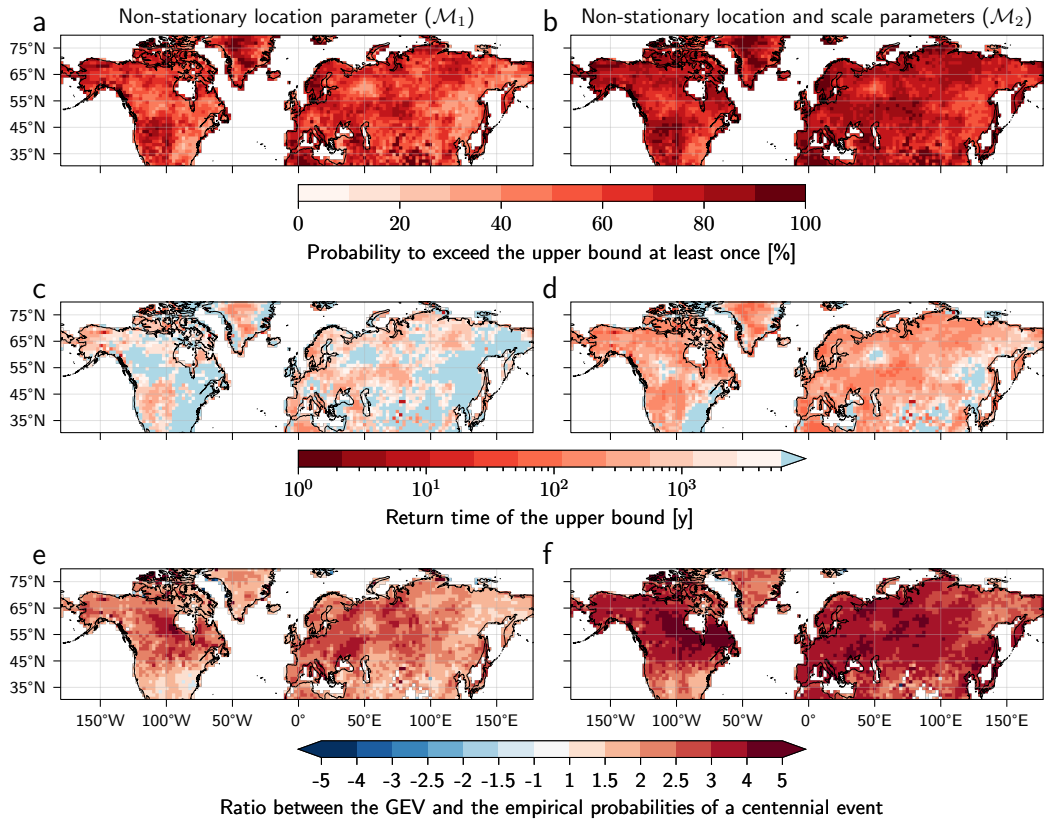

Figure B5. Same as Fig. 1. with 70 points resampled over the period 1945-2014 with a regional covariate.

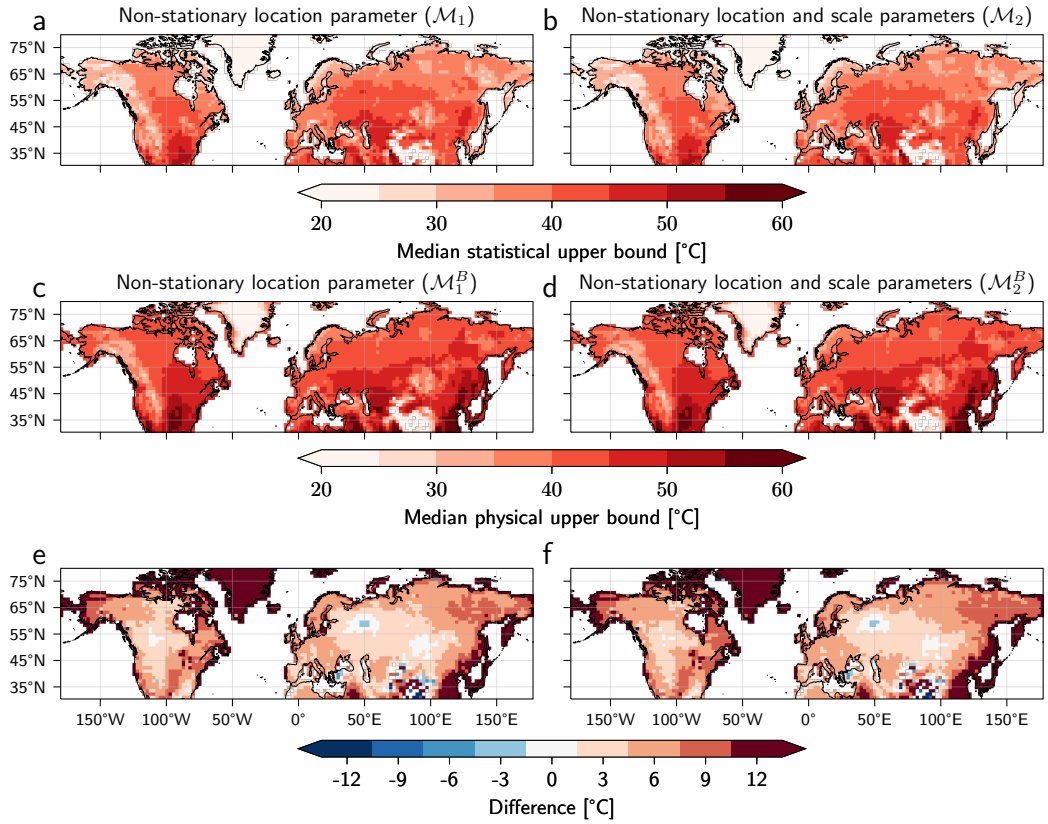

Figure B6. Absolute values of the upper bounds. Panels a and b show the median statistical upper bound estimated with models $\mathcal{M}_1$ and $\mathcal{M}_2$ obtained after 1000 fits with data resampled over the period 1945-2014. Panels c and d show the corresponding median physical upper bound. Panels e and f show the difference between the two. For all the plots we chose the covariate to be +0.5K above the 1951-1980 average (corresponding roughly to the year 2000).

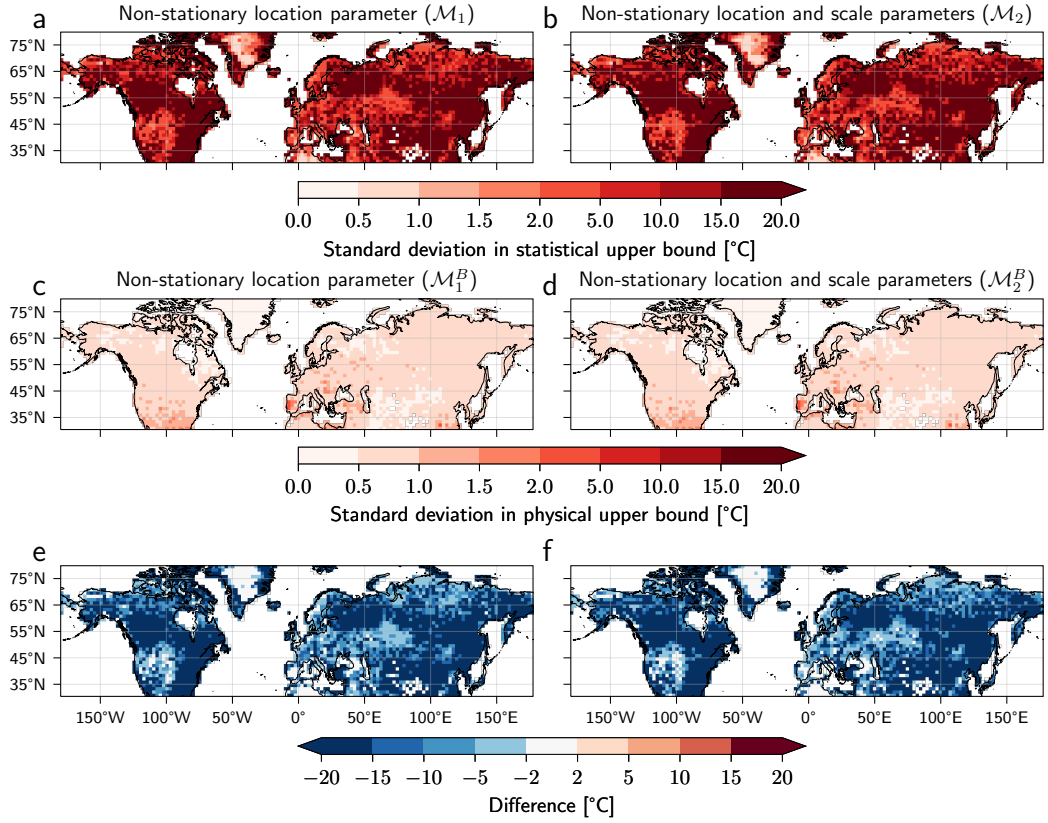

Figure B7. Precision of the estimation of the upper bounds. Panels a and b show the standard deviation in the estimated statistical upper bounds found for models $\mathcal{M}_1$ and $\mathcal{M}_2$ obtained after 1000 fits with data resampled over the period 1945-2014. Panels c and d show the corresponding standard deviation in the estimated physical upper bounds. Panels e and f show the difference between the two. For all the plots we chose the covariate to be +0.5K above the 1951-1980 average (corresponding roughly to the year 2000).

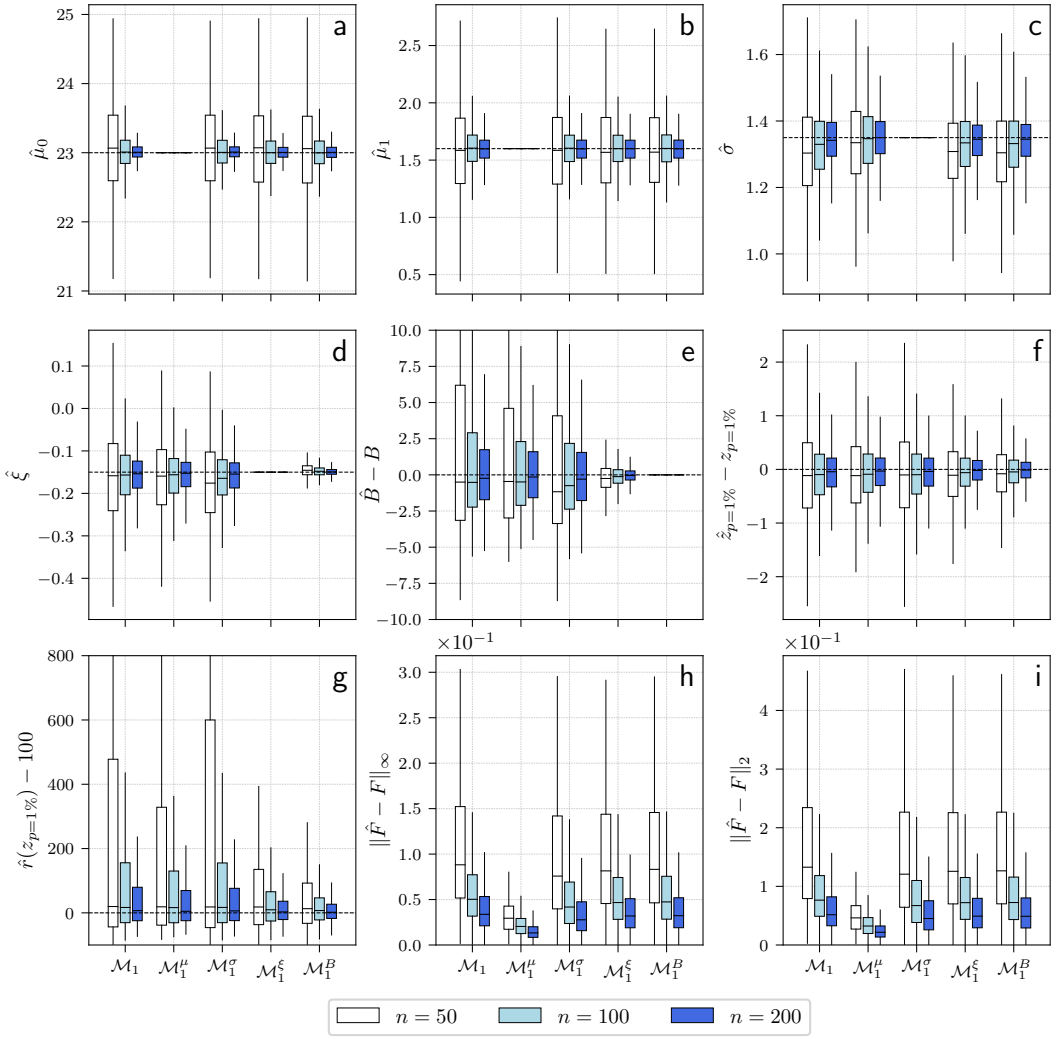

Figure B8. Fit of the statistical models derived from $\mathcal{M}_1$ on synthetic GEV data. Estimation of the parameters of the model: (a) and (b) location parameters, (c) scale parameter and (d) shape parameter. Metrics to measure the quality of the fit: (e) bias in the upper bound, (f) bias in the return level at 1%, (g) bias in the return time for the true return level at 1%, norm (h) L_∞ and (i) norm L_2 between the estimated and the true CDFs. Each boxplot is derived from 1000 fits of randomly sampled GEV data from a $\mathcal{M}_1$ model with $n = 50, 100$ and 200 samples.

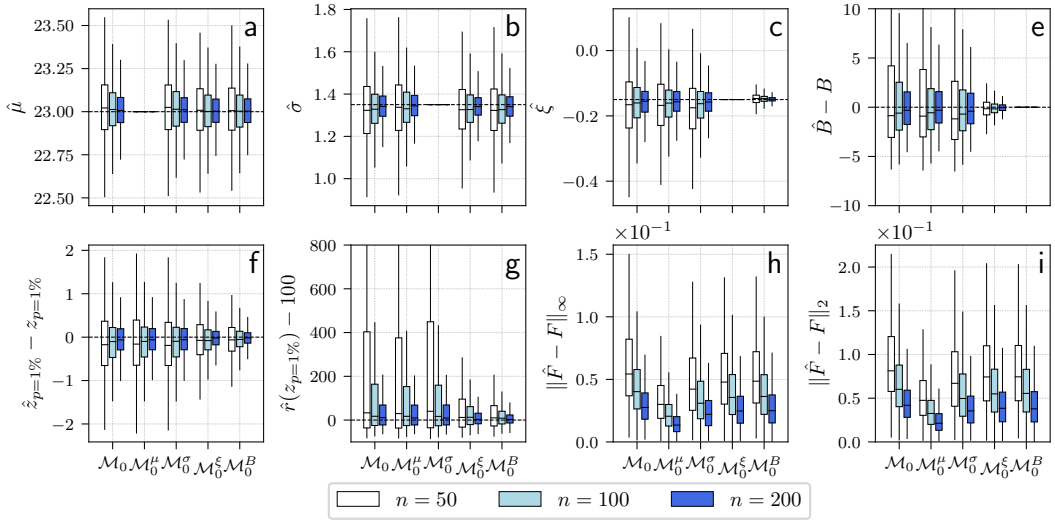

Figure B9. Same as Figure B8 with synthetic data simulated according to model $\mathcal{M}_0$.

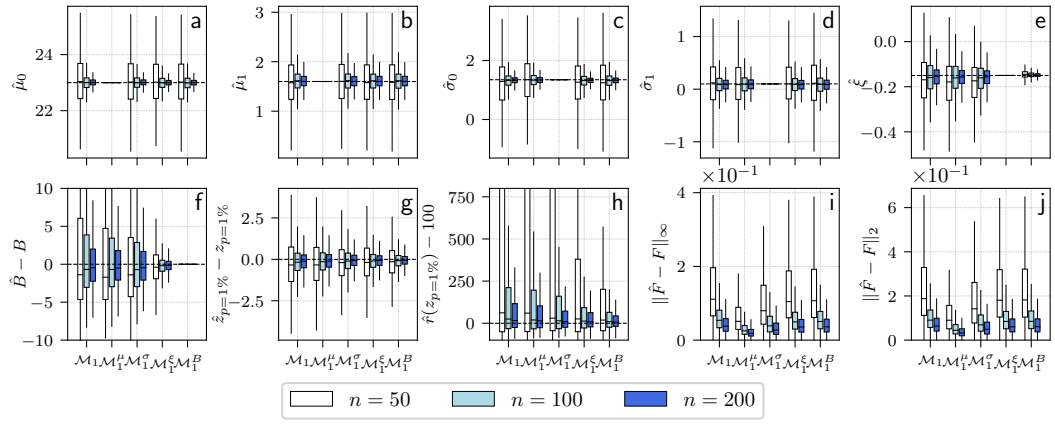

Figure B10. Same as Figure B8 with synthetic data simulated according to model $\mathcal{M}_2$.

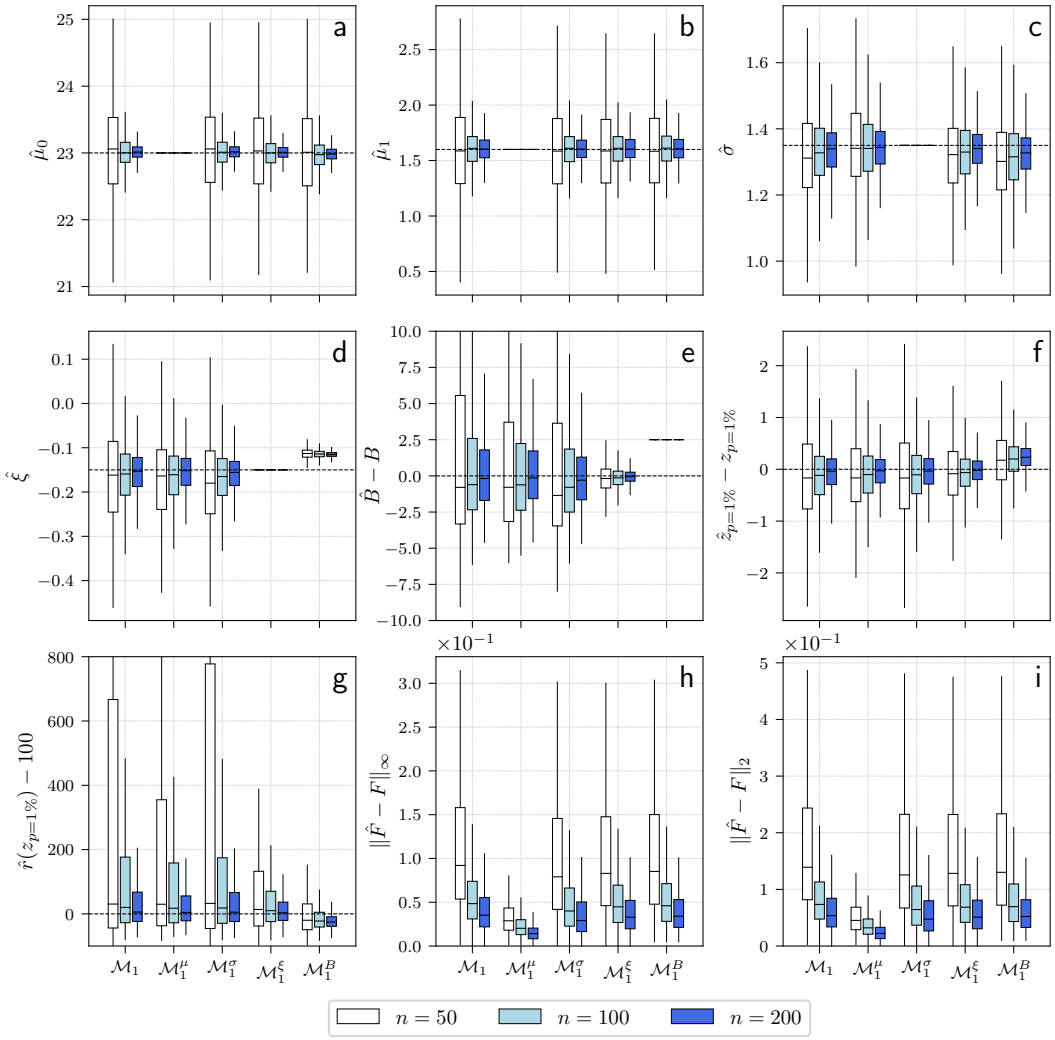

Figure B11. Same as Figure B8 with an error $B_{err} = 2.5$ imposed for model $\mathcal{M}_1^B$.

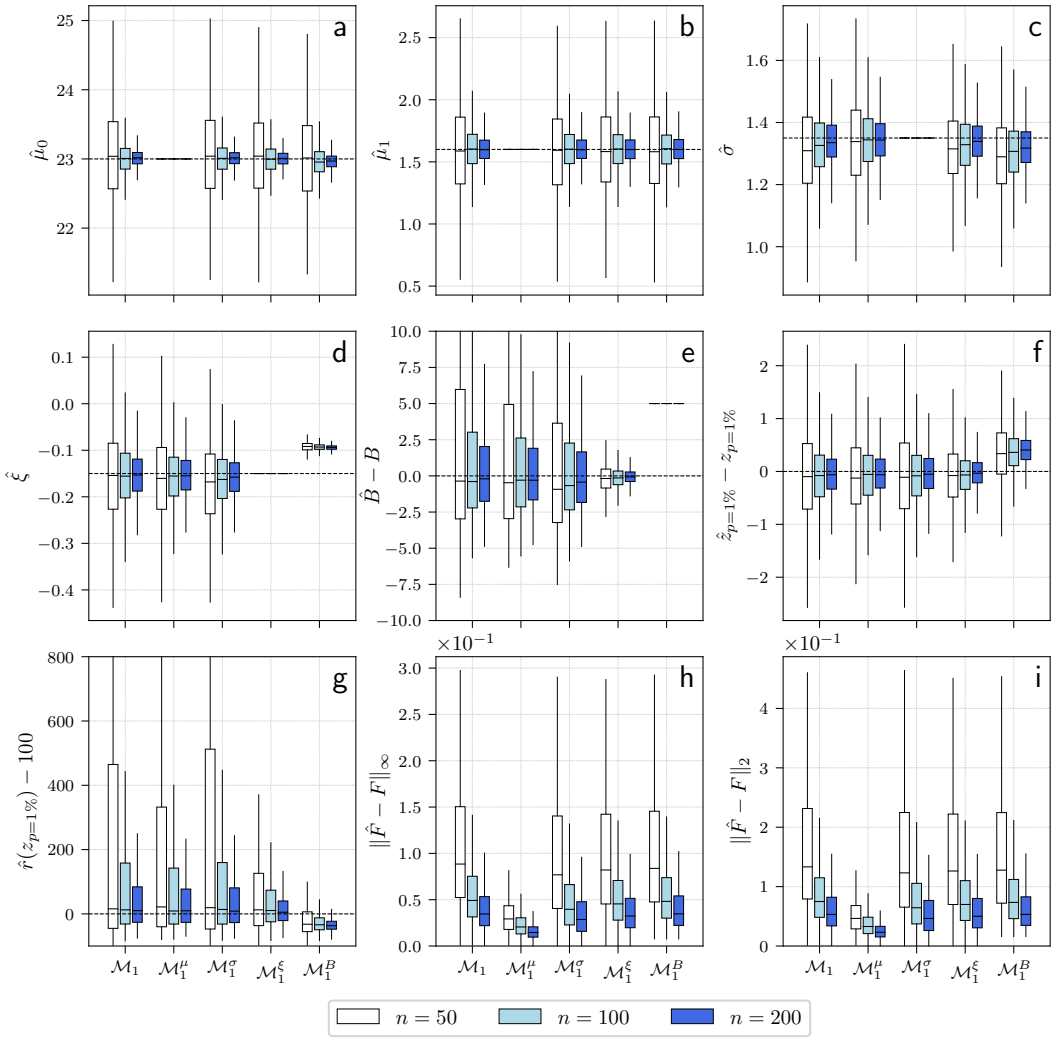

Figure B12. Same as Figure B8 with an error $B_{err} = 5$ imposed for model $\mathcal{M}_1^B$.

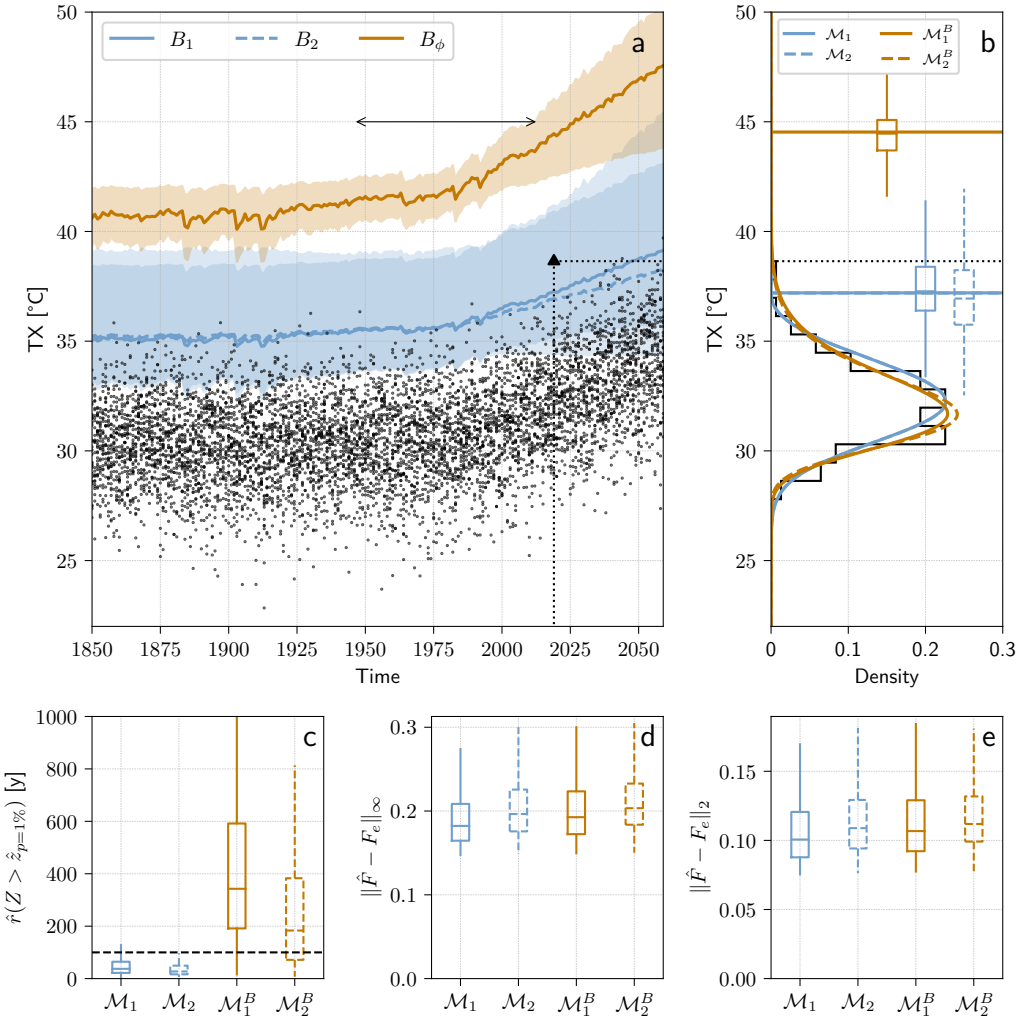

Figure B13. Same as Fig. 2 with a regional covariate.

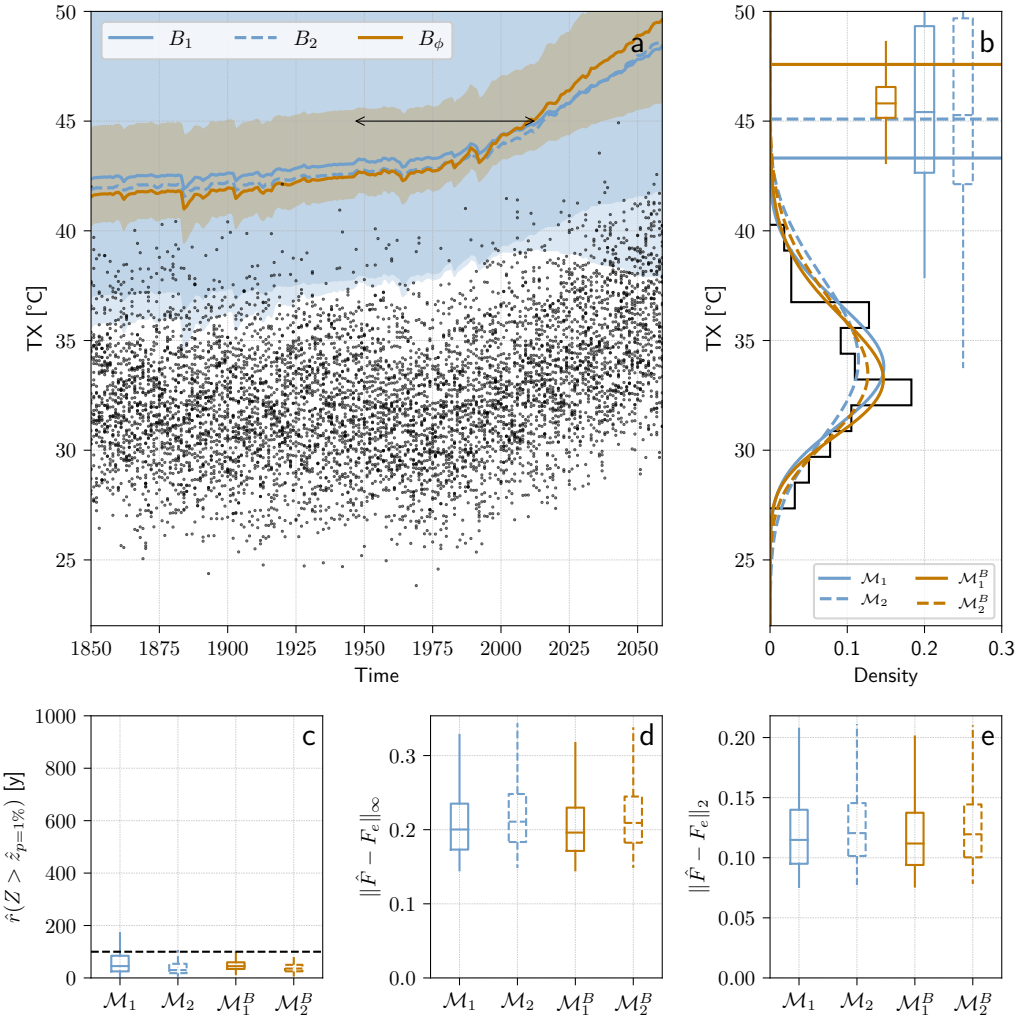

Figure B14. Same as Fig. 2 with a grid point in Western Russia (lon=47.5°E, lat=57.04°N).

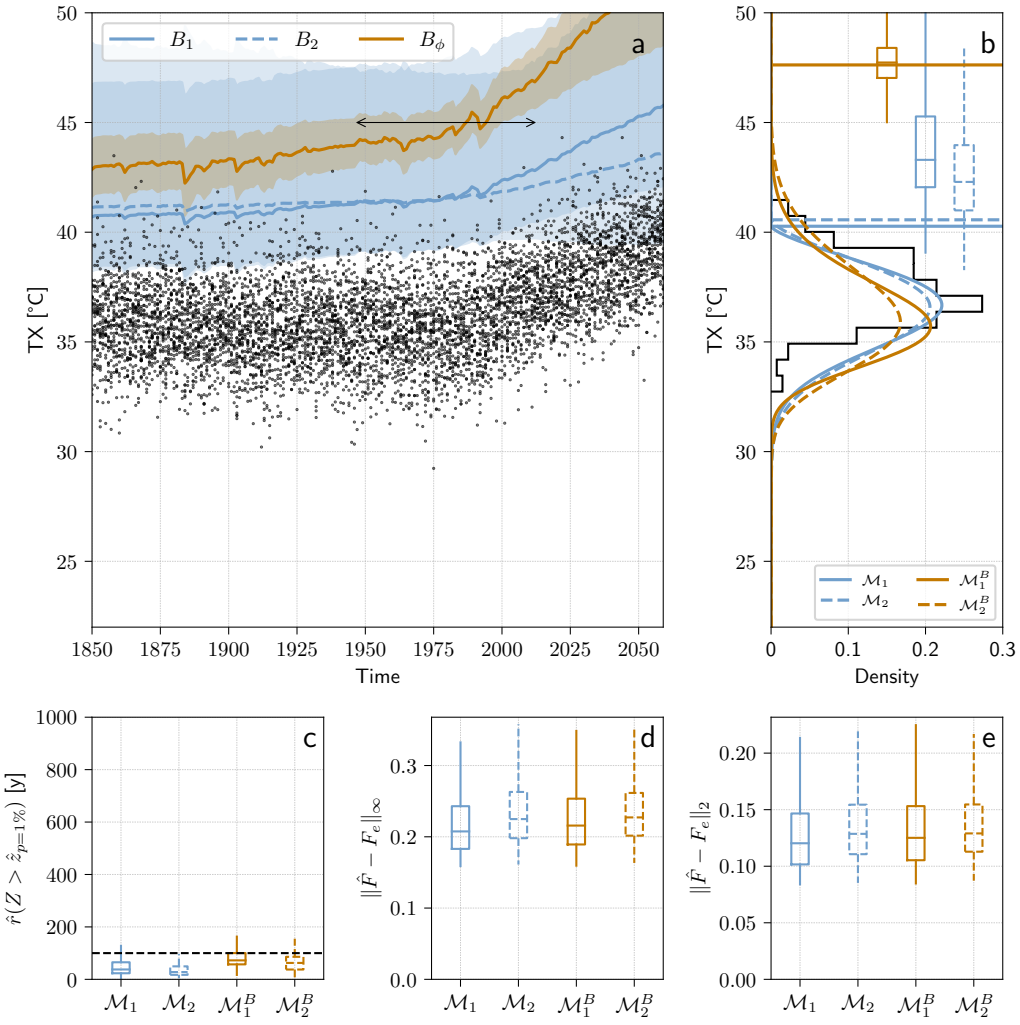

Figure B15. Same as Fig. 2 with a grid point in Eastern Kazakhstan (lon=80°E, lat=49.44°N).

Appendix C Open Research

Pre-processed climate model outputs and GEV data used in this paper are available at `https://zenodo.org/doi/10.5281/zenodo.10715007`. The analysis presented here is done using Python. The Python code is available at `https://doi.org/10.5281/zenodo.10679570`. It makes use of the SDFC package developed by Yoann Robin and available at `https://zenodo.org/doi/10.5281/zenodo.4263885`.

Acknowledgments

The authors would like to thank Yi Zhang for discussions which led to the work carried out in this paper. This work has received support from the European Union's Horizon 2020 research and innovation programme under Grant agreement no. 101003469 (XAIDA), by the Integrated Training Network (ITN) under the Marie Sklodowska-Curie Grant agreement no. 956396 (EDIPI), and the French ANR grant ANR-20-CE01-0008-01 (SAMPRACE).

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Auld, G., Hegerl, G., & Papastathopoulos, I. (2021). Changes in the distribution of observed annual maximum temperatures in europe. *arXiv preprint arXiv:2112.15117*.

Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D., & Salcedo-Sanz, S. (2023). Heat waves: Physical understanding and scientific challenges. *Reviews of Geophysics*, e2022RG000780.

Bastos, A., Orth, R., Reichstein, M., Ciais, P., Viovy, N., Zaehle, S., ... others (2021). Vulnerability of european ecosystems to two compound dry and hot summers in 2018 and 2019. *Earth system dynamics*, *12*(4), 1015–1035.

Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2005). *Statistics of Extremes: Theory and Applications*. doi: 10.1002/0470012382

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... others (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010.

Brás, T. A., Seixas, J., Carvalhais, N., & Jägermeyr, J. (2021). Severity of drought and heatwave crop losses tripled over the last five decades in europe. *Environmental Research Letters*, *16*(6), 065012.

Breshears, D. D., Fontaine, J. B., Ruthrof, K. X., Field, J. P., Feng, X., Burger, J. R., ... Hardy, G. E. S. J. (2021). Underappreciated plant vulnerabilities to heat waves. *New Phytologist*, *231*(1), 32–39.

Campbell, S., Remenyi, T. A., White, C. J., & Johnston, F. H. (2018). Heatwave and health impact research: A global review. *Health & place*, *53*, 210–218.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer. Retrieved 2022-11-09, from `http://link.springer.com/10.1007/978-1-4471-3675-0` doi: 10.1007/978-1-4471-3675-0

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.

Coumou, D., Petoukhov, V., Rahmstorf, S., Petri, S., & Schellnhuber, H. J. (2014). Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proceedings of the National Academy of Sciences*, *111*(34), 12331–12336.

Di Capua, G., Sparrow, S., Kornhuber, K., Rousi, E., Osprey, S., Wallom, D., ... Coumou, D. (2021). Drivers behind the summer 2010 wave train leading to russian heatwave and pakistan flooding. *npj Climate and Atmospheric Science*, *4*(1), 1–14.

Diffenbaugh, N. S. (2020). Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events. *Science Advances*, *6*(12), eaay2368.

Dirmeyer, P. A., Balsamo, G., Blyth, E. M., Morrison, R., & Cooper, H. M. (2021). Land-atmosphere interactions exacerbated the drought and heatwave over northern europe during summer 2018. *AGU Advances*, *2*(2), e2020AV000283.

Domeisen, D. I., Eltahir, E. A., Fischer, E. M., Knutti, R., Perkins-Kirkpatrick, S. E., Schär, C., ... Wernli, H. (2022). Prediction and projection of heatwaves. *Nature Reviews Earth & Environment*, 1–15.

Dong, B., Sutton, R. T., & Shaffrey, L. (2017). Understanding the rapid summer warming and changes in temperature extremes since the mid-1990s over western europe. *Climate Dynamics*, *48*, 1537–1554.

Fischbacher-Smith, D. (2010). Beyond the worst case scenario:'managing'the risks of extreme events. *Risk Management*, *12*, 1–8.

Fischer, E., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, *11*(8), 689–695.

Fischer, E. M., Beyerle, U., Bloin-Wibe, L., Gessner, C., Humphrey, V., Lehner, F., ... Knutti, R. (2023). Storylines for unprecedented heatwaves based on ensemble boosting. *Nature Communications*, *14*(1), 4643.

Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C. (2007). Contribution of land-atmosphere coupling to recent european summer heat waves. *Geophysical Research Letters*, *34*(6).

Gessner, C., Fischer, E. M., Beyerle, U., & Knutti, R. (2021). Very rare heat extremes: quantifying and understanding using ensemble reinitialization. *Journal of Climate*, *34*(16), 6619–6634.

Ghil, M., Yiou, P., Hallegatte, S., Malamud, B., Naveau, P., Soloviev, A., ... others (2011). Extreme events: dynamics, statistics and prediction. *Nonlinear Processes in Geophysics*, *18*(3), 295–350.

Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Annals of the Institute of Statistical Mathematics*, *36*(1), 71–85.

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

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and climate extremes*, *10*, 4–10.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., ... Stepanek, P. (2011). Observational evidence for soil-moisture impact on hot extremes in southeastern europe. *Nature Geoscience*, *4*(1), 17–21.

Horton, R. M., Mankin, J. S., Lesk, C., Coffel, E., & Raymond, C. (2016). A review of recent advances in research on extreme heat events. *Current Climate Change Reports*, *2*(4), 242–259.

Huang, W. T. K., Braithwaite, I., Charlton-Perez, A., Sarran, C., & Sun, T. (2022). Non-linear response of temperature-related mortality risk to global warming in england and wales. *Environmental Research Letters*, *17*(3), 034017.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8-12), 1287. Retrieved 2022-11-09, from `https://www.academia.edu/30839592/Statistics_of_extremes_in_hydrology`

Kiriliouk, A., & Naveau, P. (2020). Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Annals of Applied Statistics*, *14*(3), 1342-1358.

Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J. F., Lehmann, J., & Horton, R. M. (2020). Amplified rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nature Climate Change*, *10*(1), 48–53.

Luo, F., Wilcox, L., Dong, B., Su, Q., Chen, W., Dunstone, N., ... Gao, Y. (2020).

Projected near-term changes of temperature extremes in europe and china under different aerosol emissions. *Environmental Research Letters*, *15*(3), 034013.

Masselot, P., Mistry, M., Vanoli, J., Schneider, R., Iungman, T., Garcia-Leon, D., ... others (2023). Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in europe. *The Lancet Planetary Health*, *7*(4), e271–e281.

Miralles, D. G., Teuling, A. J., Van Heerwaarden, C. C., & Vilà-Guerau de Arellano, J. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature geoscience*, *7*(5), 345–349.

Miralles, D. G., Van Den Berg, M., Teuling, A., & De Jeu, R. (2012). Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters*, *39*(21).

Mitchell, D., Kornhuber, K., Huntingford, C., & Uhe, P. (2019). The day the 2003 european heatwave record was broken. *The Lancet Planetary Health*, *3*(7), e290–e292.

National Academies of Sciences Engineering and Medicine. (2016). *Attribution of extreme weather events in the context of climate change.* The National Academies Press, Washington, DC. doi: 10.17226/21852

Naveau, P., Hannart, A., & Ribes, A. (2020). Statistical Methods for Extreme Event Attribution in Climate Science. *Annual Reviews of Statistics and its Application*. Retrieved from `https://hal.archives-ouvertes.fr/hal-02408516` doi: 10.1146/annurev-statistics-031219-

Naveau, P., Ribes, A., Zwiers, F., Hannart, A., Tuel, A., & Yiou, P. (2018, May). Revising Return Periods for Record Events in a Climate Event Attribution Context. *Journal of Climate*, *31*(9), 3411–3422. Retrieved 2022-11-08, from `https://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0752.1` doi: 10.1175/JCLI-D-16-0752.1

Noyelle, R., Zhang, Y., Yiou, P., & Faranda, D. (2023). Maximal reachable temperatures for western europe in current climate. *Environmental Research Letters*, *18*(9), 094061.

Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *128*(581), 747–774.

Patterson, M. (2023). North-west europe hottest days are warming twice as fast as mean summer days. *Geophysical Research Letters*, *50*(10), e2023GL102757.

Perkins, S. E. (2015). A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164*, 242–267.

Petoukhov, V., Rahmstorf, S., Petri, S., & Schellnhuber, H. J. (2013). Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. *Proceedings of the National Academy of Sciences*, *110*(14), 5336–5341.

Philip, S., Kew, S., van Oldenborgh, G., Otto, F., Vautard, R., van der Wiel, K., ... others (2020). *A protocol for probabilistic extreme event attribution analyses. adv stat climatol meteorol oceanogr 6: 177–203.*

Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vautard, R., ... others (2021). Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the us and canada june 2021. *Earth System Dynamics Discussions*, 1–34.

Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., ... others (2022). Climate change 2022: Impacts, adaptation and vulnerability. *IPCC Sixth Assessment Report*.

Ragone, F., & Bouchet, F. (2021). Rare event algorithm study of extreme warm summers and heatwaves over europe. *Geophysical Research Letters*, *48*(12),

e2020GL091197.

Ragone, F., Wouters, J., & Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, *115*(1), 24–29.

Rasmijn, L., Van der Schrier, G., Bintanja, R., Barkmeijer, J., Sterl, A., & Hazeleger, W. (2018). Future equivalent of 2010 russian heatwave intensified by weakening soil moisture constraints. *Nature Climate Change*, *8*(5), 381–385.

Robin, Y., & Ribes, A. (2020). Nonstationary extreme value analysis for event attribution combining climate models and observations. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 205–221.

Scher, S., Jewson, S., & Messori, G. (2021). Robust worst-case scenarios from ensemble forecasts. *Weather and forecasting*, *36*(4), 1357–1373.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., ... Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3-4), 125–161.

Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land–atmosphere coupling and climate change in europe. *Nature*, *443*(7108), 205–209.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., ... Zhou, B. (2021). *11 chapter 11: Weather and climate extreme events in a changing climate.*

Seong, M.-G., Min, S.-K., Kim, Y.-H., Zhang, X., & Sun, Y. (2021). Anthropogenic greenhouse gas and aerosol contributions to extreme temperature changes during 1951–2015. *Journal of Climate*, *34*(3), 857–870.

Sillmann, J., Pozzoli, L., Vignati, E., Kloster, S., & Feichter, J. (2013). Aerosol effect on climate extremes in europe under different future scenarios. *Geophysical Research Letters*, *40*(10), 2290–2295.

Thompson, V., Kennedy-Asser, A. T., Vosper, E., Lo, Y. E., Huntingford, C., Andrews, O., ... Mitchell, D. (2022). The 2021 western north america heat wave among the most extreme events ever recorded globally. *Science advances*, *8*(18), eabm6860.

Van Oldenborgh, G. J., Van Der Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., ... others (2021). Pathways and pitfalls in extreme event attribution. *Climatic Change*, *166*(1-2), 13.

Van Oldenborgh, G. J., Wehner, M. F., Vautard, R., Otto, F. E., Seneviratne, S. I., Stott, P. A., ... Kew, S. F. (2022). Attributing and projecting heatwaves is hard: We can do better. *Earth's Future*, *10*(6), e2021EF002271.

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., ... others (2023). Heat extremes in western europe are increasing faster than simulated due to missed atmospheric circulation trends.

Vautard, R., van Aalst, M., Boucher, O., Drouin, A., Haustein, K., Kreienkamp, F., ... others (2020). Human contribution to the record-breaking june and july 2019 heatwaves in western europe. *Environmental Research Letters*, *15*(9), 094077.

Westervelt, D. M., Mascioli, N. R., Fiore, A. M., Conley, A. J., Lamarque, J.-F., Shindell, D. T., ... Horowitz, L. W. (2020). Local and remote mean and extreme temperature response to regional aerosol emissions reductions. *Atmospheric Chemistry and Physics*, *20*(5), 3009–3027.

Worms, J., & Naveau, P. (2022). Record events attribution in climate studies. *Environmetrics*. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2777` doi: 10.1002/env.2777

Wouters, J., & Bouchet, F. (2016). Rare event computation in deterministic chaotic systems using genealogical particle analysis. *Journal of Physics A: Mathemati-*

cal and Theoretical, *49*(37), 374002.

Yiou, P., & Jézéquel, A. (2020). Simulation of extreme heat waves with empirical¡? xmltex\break?¿ importance sampling. *Geoscientific Model Development*, *13*(2), 763–781.

Zeder, J., Sippel, S., Pasche, O. C., Engelke, S., & Fischer, E. M. (2023). The effect of a short observational record on the statistics of temperature extremes. *Geophysical Research Letters*, *50*(16), e2023GL104090.

Zhang, Y., & Boos, W. R. (2023). An upper bound for extreme temperatures over midlatitude land. *Proceedings of the National Academy of Sciences*, *120*(12), e2215278120.

Zuo, J., Pullen, S., Palmer, J., Bennetts, H., Chileshe, N., & Ma, T. (2015). Impacts of heat waves and corresponding measures: a review. *Journal of Cleaner Production*, *92*, 1–12.