Frequency Bias Causes Overestimation of Climate Change Impacts on Global Flood Occurrence

Fang Zhao[1,2,3,4] 📷, **Stefan Lange**[3] 📷, **Bedartha Goswami**[5] 📷, and **Katja Frieler**[3] 📷

[1]Key Laboratory of Geographic Information Science of the Ministry of Education, School of Geographic Sciences, East China Normal University, Shanghai, China, [2]Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, Shanghai, China, [3]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [4]Water Security Research Group, Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, [5]Machine Learning in Climate Science, University of Tübingen, Tübingen, Germany

Abstract The frequency change of 100-year flood events is often determined by fitting extreme value distributions to annual maximum discharge from a historical base period. This study demonstrates that this approach may significantly bias the computed flood frequency change. An idealized experiment shows frequency bias exceeding 100% for a 50-year base period. Further analyses using Monte Carlo simulations, mathematical derivations, and hydrological model outputs reveal that bias magnitude inversely relates to base period length and is weakly influenced by the generalized extreme value distribution's shape parameter. The bias, persisting across different estimation methods, implies floods may exceed local defenses designed based on short historical records more often than expected, even without climate change. We introduce a frequency bias adjustment method, which significantly reduces the projected rise in global flood occurrence. This suggests a substantial part of the earlier projected increase in flood occurrence and impacts is not attributable to climate change.

Plain Language Summary We report a previously overlooked bias in a common method for projecting the frequency of extreme floods, like 100-year events. This method typically uses a short period of historical data for two primary purposes: first, to define local thresholds for extreme floods, such as 100-year events, and second, to serve as a base for projecting future changes in flood occurrences. Our study, utilizing random data, demonstrates that this method can lead to inaccurate projections, increasing the number of expected extreme floods in future years and decreasing them in historical baseline years, even when no change is indicated by the data. Additionally, we find that the longer the historical period used, the smaller the bias. By simulating the characteristics of real hydrological model projections with artificial data, we are able to estimate a bias similar to that observed in these projections, and thus propose a simple method for its mitigation. Our findings indicate that previous studies might have overstated the impact of climate change on the projected increase in flood frequency and affected people. Additionally, using short historical records to design flood defenses can lead to more frequent flooding than anticipated, even without climate change.

1. Introduction

Understanding how extreme event frequencies change due to anthropogenic global warming is pivotal for climate change research and informed decision-making (Harrington et al., 2019; Sippel, Walton, & Otto, 2015). Extreme event thresholds are often based on relative terms, such as percentiles or standard deviations (IPCC, 2014; IPCC, 2021). However, using a fixed base period to estimate these thresholds can severely bias extreme event frequency estimates (Sippel, Zscheischler, et al., 2015; Sippel et al., 2017; Zhang et al., 2005). Zhang et al. (2005) noted artificial jumps in percentile-based temperature extremes' exceedance rates at base period boundaries due to quantile estimates biases and sampling variability. Sippel, Zscheischler, et al. (2015) found that fixed-period normalization could exaggerate increases in temperature or precipitation extremes. These findings have been widely recognized by the climate research community, improving climate change detection, attribution and monitoring (Abatzoglou & Williams, 2016; De Luca & Donat, 2023; Mahony & Cannon, 2018; Ombadi et al., 2023; Pokhrel et al., 2021; Ringard et al., 2019; Simolo & Corti, 2022; Stevenson et al., 2022; Williams & O'Gorman, 2022).

Changes in river floods, often proxied by annual maximum streamflow, are projected to increase across more global land areas (IPCC, 2021). The frequency of rare, devastating river floods, like 100-year floods that often surpass local flood protection levels, is of particular concern. These thresholds are typically determined by fitting an extreme value function to annual maximum discharge data—non-gaussian in nature—at each grid cell. More grid cells exceeding these thresholds indicate a rise in flood events. Previous studies vary in their choice of base period to estimate 100-year flood, from a short historical period (about 30 years) (Alfieri et al., 2015; Hirabayashi et al., 2013) to long pre-industrial simulations (over 400 years) (Lange et al., 2020), which could lead to substantial uncertainties. These uncertainties complicate effective adaptation strategies, such as optimizing flood defenses (Ward et al., 2017) and managing flood risk (Willner et al., 2018). However, the impact of using a fixed base period for extreme value function fitting on flood frequency estimates remains poorly understood.

In the next section, we first use Monte Carlo simulations to demonstrate that using a fixed base period for extreme value function fitting significantly biases exceedance frequencies, and present our mathematical insights. We then apply the insights to emulate biases in a real hydrological simulation. We further propose a practical bias-adjustment method based on Monte Carlo simulations and apply it to projected flood frequency changes. An overall flowchart of this section is shown in Figure S1 in Supporting Information S1. Section 3 concludes with an outlook.

2. Methodology and Results

2.1. Understanding Frequency Bias in Changes of Extremes

We conducted Monte Carlo simulations with independent and identically distributed (i.i.d.) random variables from a Generalized Extreme Value (GEV) distribution (mean $\mu = 0$, standard deviation $\sigma = 1$, shape parameter $\xi = 0.2$). Following Sippel, Zscheischler, et al. (2015), this experiment was set up in analogy to investigating flood frequency changes across 10,000 separate annual maximum discharge data series, each with 100 random, i. i.d. numbers representing 100 years. Using the first 50 values of each series, we estimated GEV distribution parameters with the L-moments method to set the threshold for local 100-year events. We then counted exceedances above the local thresholds in each year. For comparison, we also counted exceedances above a fixed theoretical 100-year return level from the original GEV distribution ($\mu = 0$, $\sigma = 1$, $\xi = 0.2$), which has an expected exceedance probability of 1%.

The artificial data, despite its stable time-invariant statistical properties, showed a notable increase in the number of extreme events outside the base period and a decrease within it, similar to Sippel, Zscheischler, et al. (2015) (Figure 1a). The 100-year flood exceedance probability in this data rose by about 100% out-of-base compared to the base period (red line in Figure 1a). Additionally, there was a 60% overestimation in 100-year flood events against the fixed theoretical threshold during the out-of-base period (black line in Figure 1a, 'true' line in Figure 1b), termed out-of-base relative frequency bias (ORB). This bias increased the exceedance probability from 1% to 1.6% (out-of-base frequency bias, or OFB of 0.6%). In contrast, the base period showed a 20% underestimation (in-base relative frequency bias, IRB), or an in-base frequency bias (IFB) of −0.2%.

To unravel the cause of the apparent increase in extremes during the out-of-base period, we demonstrate why an unbiased estimator of return level does not preclude bias in the exceedance probability. Although the L-moment method produces an unbiased estimator for GEV distribution parameters, individual sample estimates can still significantly differ from the true value (green histogram, Figure 1b). This sampling variability causes the average exceedance frequency of the estimated 100-year return levels to be over 1%, as briefly described by Zhang et al. (2005). For an error of Δq above and below the true return level, the positive bias in the exceedance probability (shaded area right of the return level under the probability density function curve) is shown by the yellow area being larger than the orange area (Figure 1b inset). This occurs due to the concave upward shape of the upper tail of the exceedance probability function G (Figure 1c, Text S1 in Supporting Information S1). This bias persists, across different population distributions (Normal, Gumbel, GEV) and quantile estimation methods (nonparametric, moments, L-moments), as seen in Figures S2–S5 and Text S1 in Supporting Information S1. Unlike previous studies, we provide a closed-form expression for the OFB in the large-sample limit for the Normal distribution and numerically expand it for Gumbel and GEV cases (Text S1 in Supporting Information S1). Key mathematical results are summarized below.
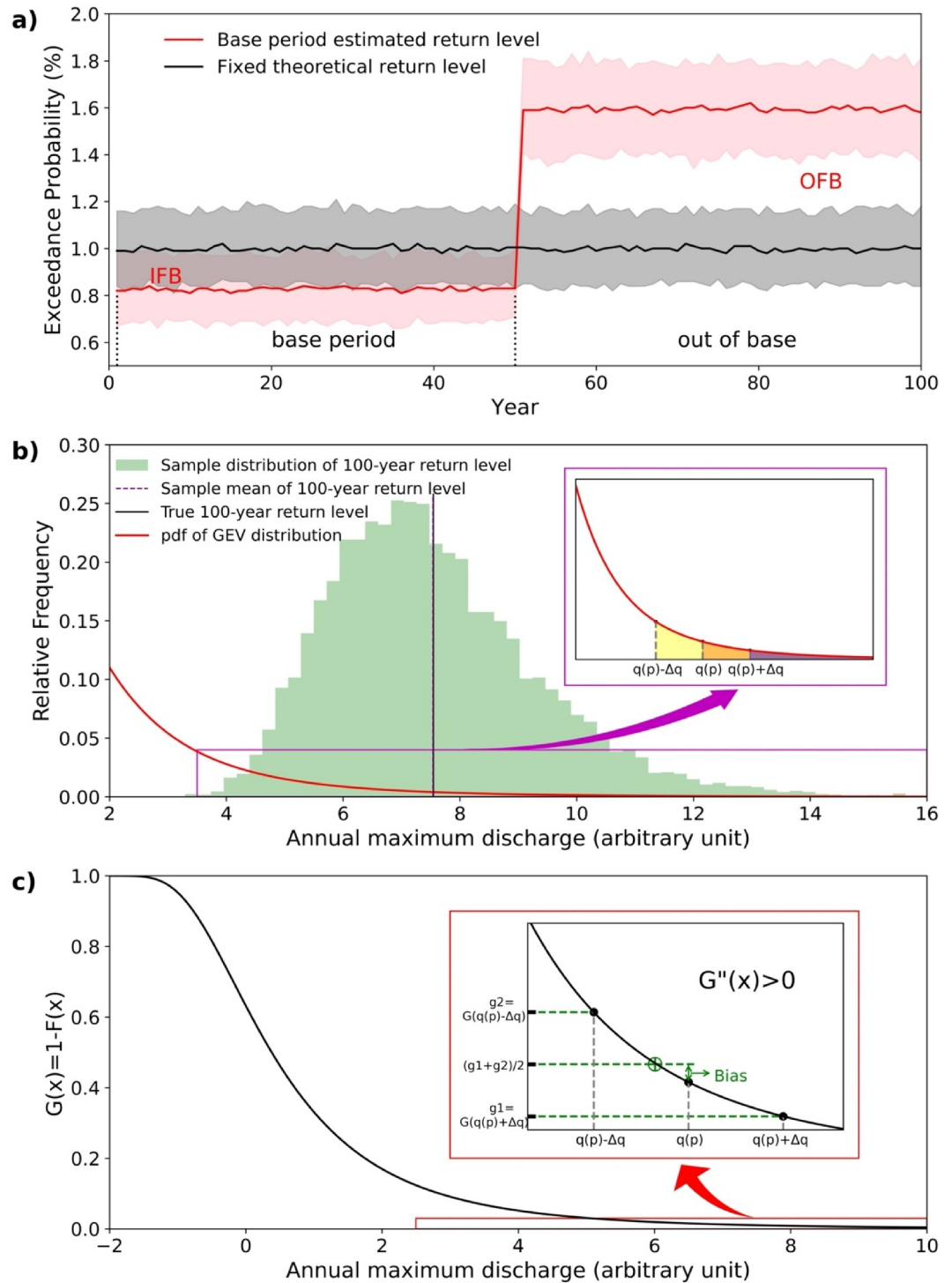
Figure 1. Biases in detecting extreme events using local 100-year thresholds. (a) Exceedance probability of the 100-year return level in an artificial GEV series from 10,000 samples (100 points each, first 50 as base sample). Red and black lines show frequencies exceeding the estimated and true 100-year levels, respectively. Shaded areas indicate the 5th to 95th percentile range in 200 simulations. (b) Histogram of 100-year return levels estimated by GEV distribution fits to the 10,000 base samples, its sample mean, and the true 100-year return level from the population GEV distribution. The inset visualizes the bias. (c) Illustration of how the shape of the tail of $G(x) = 1-F(x)$ leads to OFB; $F(x)$ is the GEV distribution's cumulative distribution function, $q(p)$ is the quantile for cumulative probability p. The inset shows the bias, with g1 corresponding to the purple shaded area in inset (b).

Define F as the cumulative distribution function (CDF) of the random variable, p as the cumulative probability ($p = 0.99$ for a 100-year event), $q(p) = F^{-1}(p)$ as the true quantile value for p (q (0.99) for the 100-year level), $f = F'$ as the probability density function (PDF), and f' as the derivative of f. Using Taylor expansion and omitting higher order terms, we demonstrate in Text S1 in Supporting Information S1 that the OFB can approximately be expressed for data from any distribution as

$$OFB(p) \cong - f(q(p)) E[\delta_q(p)] - \frac{1}{2} f'(q(p)) E[\delta_q^2(p)] \tag{1}$$

where $\delta_q(p)$ represents the estimation error of the quantile function at p, E denotes its expectation, and higher order terms in $\delta_q(p)$ are ignored. For the Normal distribution, using the definition:

$$x(p) \equiv erf^{-1}(2p - 1) \tag{2}$$

where erf denotes the error function, we derive (Text S2 in Supporting Information S1) a close-form expression for OFB in the limit of infinite base sample size n:

$$n\, OFB(p) \cong \frac{1}{2\sqrt{\pi}} x(p) \left(\frac{3}{2} + x(p)^2\right) exp\left(-x(p)^2\right) \tag{3}$$

Equation 3 shows that the relative bias is independent of the sample's mean or standard deviation. Monte Carlo simulations for various return periods and sample sizes show that the simulations converge to the analytical solution as sample size increases, validating Equation 3's accuracy (Figure 2a). For Gumbel and GEV distributions, with quantiles estimated using the L-moments method, Monte Carlo simulations reveal that the ORB is not affected by the base sample's mean or standard deviation, and only weakly depends on the GEV shape parameter (Figures 2b–2d). The bias inversely scales with sample size, and smaller sample introduces additional bias, especially for the Normal and Gumbel distributions. The GEV distribution shows higher ORB, due to the shape parameter adding a degree of freedom and causing a wider range of estimated return levels (Figures S2–S5, Text S1 in Supporting Information S1).

The IRB fundamentally differs from ORB. In IRB, the sample for estimating the return level is the same as the one used for determining the frequency of exceedance, leading to statistical dependence and typically, but not always, resulting in frequency underestimation. For i. i.d. samples from a Normal distribution, we derive an analytical expression for IFB (see Text S3, Equation (44) in Supporting Information S1). We further determine (Text S4 in Supporting Information S1) a simple closed-form expression of the IFB (Figures S6–S7 in Supporting Information S1) in the limit of infinite base sample size n:

$$n\, IFB(p) \cong \frac{1}{2\sqrt{\pi}} x(p) \left(\frac{1}{2} - x(p)^2\right) exp\left(-x(p)^2\right) \tag{4}$$

Our study builds on previous research by Zhang et al. (2005) and Sippel, Zscheischler, et al. (2015) on biases in extreme event frequency estimates, with our contributions including: (a) simple closed-form expressions for IRB and ORB for Normal distribution data in the limit of infinite sample size, (b) a novel mathematical depiction of ORB using Taylor expansion applicable to data from any distribution, (c) Monte Carlo simulations of IRB and ORB for Normal, Gumbel, and GEV distributions, and (d) analysis of factors affecting IRB and ORB like sample size, distribution parameters, and quantile estimation methods (nonparametric, moments, L-moments). Detailed derivations of Equations 1, 3 and 4 (the latter two pertaining to Normal distribution), additional numerical results, and Python code for IFB and OFB calculations are provided in Texts S1 to S4 in Supporting Information S1 and the Data Availability section.

2.2. Emulating Frequency Bias Using a Hydrological Model Simulation

Utilizing random data, we demonstrated significant bias in assessing flood frequency changes relative to a short base period when the flood threshold is estimated within that period. Fitting extreme value distribution to annual maximum discharge timeseries simulated by hydrological models results in varied parameters at the grid cells,
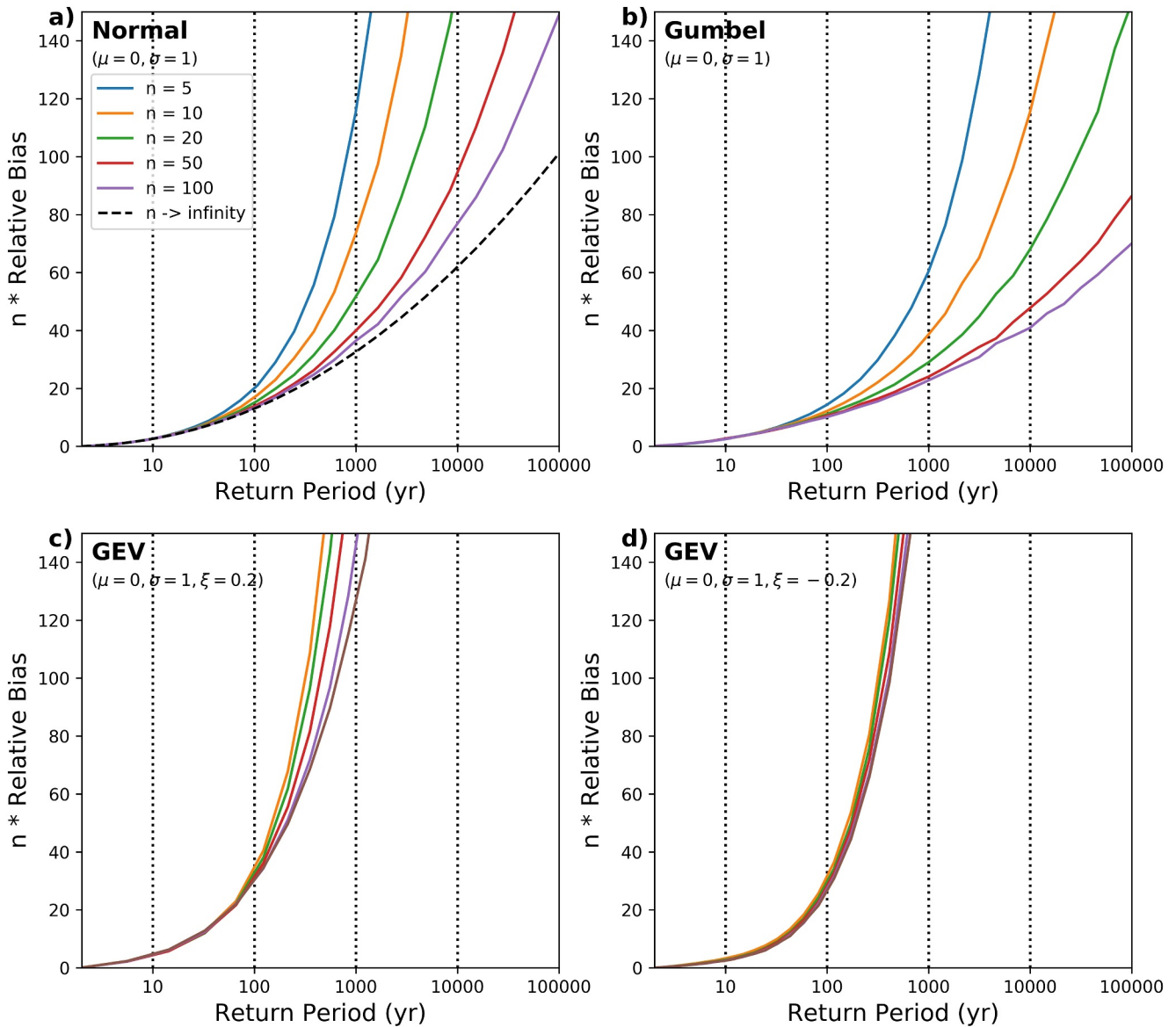
Figure 2. Relative out-of-base bias in exceedance probability for different return periods for the (a) Normal distribution, (b) Gumbel distribution, (c) GEV distribution with $\xi = 0.2$, and (d) GEV distribution with $\xi = -0.2$, based on Monte Carlo simulations using 100,000 samples of various sizes (n). The black dashed line in (a) represents the analytical solution for large n.

and spatial and temporal autocorrelations exist. We reveal that this bias persists in actual hydrological simulations and is minimally affected by differences from i. i. d. random data. We specifically examine experiment results using data from the ISIMIP2b project (Frieler et al., 2017), which provides pre-industrial control (piControl), historical and future simulation results. The piControl simulation spans from 1660 to 2299, with climate forcing set at pre-industrial level and direct human influences (land and water management activities and socioeconomic conditions) fixed at 2005 level for the 2006–2299 period (see Figure 1 in Frieler et al., 2017). The future scenarios (e.g., RCP2.6 and RCP6.0) also have direct human influences fixed at 2005 level, so differences between piControl and RCP2.6/RCP6.0 experiments result from pure climate change.

Global daily discharge at 0.25-degree spatial resolution were simulated using CaMa-Flood v3.4.4 (Yamazaki et al., 2011, 2013), with runoff input from the ISIMIP2b experiments. Specifically, we used data from one land surface model, LPJmL (von Bloh et al., 2010; Rost et al., 2008), driven by bias-corrected IPSL-CM5A-LR climate data. CaMa-Flood's default settings and a standard model spin-up procedure were employed. In this

subsection, we focus on the piControl experiment only, while simulations from other experiments are also used in the next subsection. Given the stable climate and socioeconomic conditions during the piControl period, we expect no change in flood frequency.

We divided the 2006–2299 piControl simulation (294-year period) into 11 25-year epochs (e.g., 2006–2030, 2031–2055, etc.) to analyze flood frequency. We combined these epochs into samples of various lengths (25, 50, 100, 200 years) to maintain data's auto-correlation, then randomly drew 11 samples for uncertainty estimation. We then fitted a GEV distribution to the annual maximum discharge at each grid cell for different sample sizes, estimating the 100-year return levels locally with the L-moments method. The percentage of grid cells that exceeded the local 100-year return level, defined as the exceedance probability of 100-year floods, was calculated yearly. We used the probability plot correlation coefficient test (Vogel, 1986) to ensure the quality of the GEV fit, keeping only those cells with a score above 0.99. This resulted in 101,877 grid cells being selected for further analysis (Figure S8 in Supporting Information S1).

We emulated the statistical properties of the piControl simulation by generating random samples ($\mu = 0$, $\sigma = 1$) with the same length and shape parameters as the real data but without autocorrelation, and applied the same epoch combination method. Both real and random data were analyzed to compare the exceedance probability for in-base (base samples created by combining the epochs) and out-of-base (remainder of the 294-year timeseries) periods. This comparison, detailed in Figure 3 and Table S1 in Supporting Information S1, revealed a consistent bias in estimating flood frequency in both real and emulated data. The ORB is inversely related to base period length, suggesting the potential for substantial biases with shorter base periods for return level estimation. The bias of the ratio of out-of-base to in-base frequencies of 100-year floods averages 329% (91%, 39%, 18%) for 25 (50, 100, 200)-year base samples. These findings closely align with our earlier idealized experiment and allow us to accurately estimate the magnitude of bias in actual hydrological simulations.

To analyze the sensitivity of ORB, IRB, and the total bias to the GEV shape parameter, we divided the above samples into three categories: shape parameters close to 0 (−0.05 to 0.05), above 0.1, and below −0.1. The results, shown in Figure 3, reveal that the increase in the occurrence of extremes is more pronounced with shorter base periods due to larger sampling variability. For the 25-year base period, the greatest increase in the occurrence of extremes is observed in the group with shape parameters greater than 0.1, mainly due to a large negative IRB (−70%). In contrast, the group with shape parameters less than −0.1 shows the smallest increase in the occurrence of extremes and a less negative IRB, even exhibiting a positive IRB for base periods of 100 or 200 years. Interestingly, random data with identical shape parameters but without autocorrelation yielded similar results to the actual hydrological simulation, indicating that location and scale parameters, as well as the degree of auto-correlation, have a minor impact on the observed increase in 100-year flood frequency.

The bias estimation uncertainties, as depicted by error bars in Figure 3, are smaller for Monte Carlo estimates compared to real data. This difference is attributed to the independent and random nature of Monte Carlo samples, in contrast to real data that encompasses natural climate variability. Real data's 25-year epochs reflect different phases of the climate system's natural variability, resulting in diverse statistical properties across epochs. Meanwhile, Monte Carlo simulations exhibit consistent statistical properties, leading to reduced variability and lower estimation uncertainties.

For grid cells with GEV distribution shape parameters near zero, using Gumbel distribution fits to estimate the 100-year return levels effectively reduces biases (Figure S9 in Supporting Information S1). However, for other cells, the Gumbel distribution performs poorly, resulting in significant IRB and ORB (Figure S9c–S9d in Supporting Information S1).

2.3. Adjusting Frequency Bias in Historical and Future Flood Simulations

We have established that defining extreme occurrence using extreme value theory over a fixed base period introduces significant bias. This bias cannot be corrected analytically with existing methods because the extreme distribution does not satisfy normality assumptions (Sippel, Zscheischler, et al., 2015). Moreover, the bootstrap procedure (Zhang et al., 2005), which we recommend for estimating accurate changes at the grid cell level, is computationally intensive. Although a long base period can mitigate these biases, it's often impractical due to the lack of extended piControl experiments in many impact studies. Here, we introduce a first-order adjustment
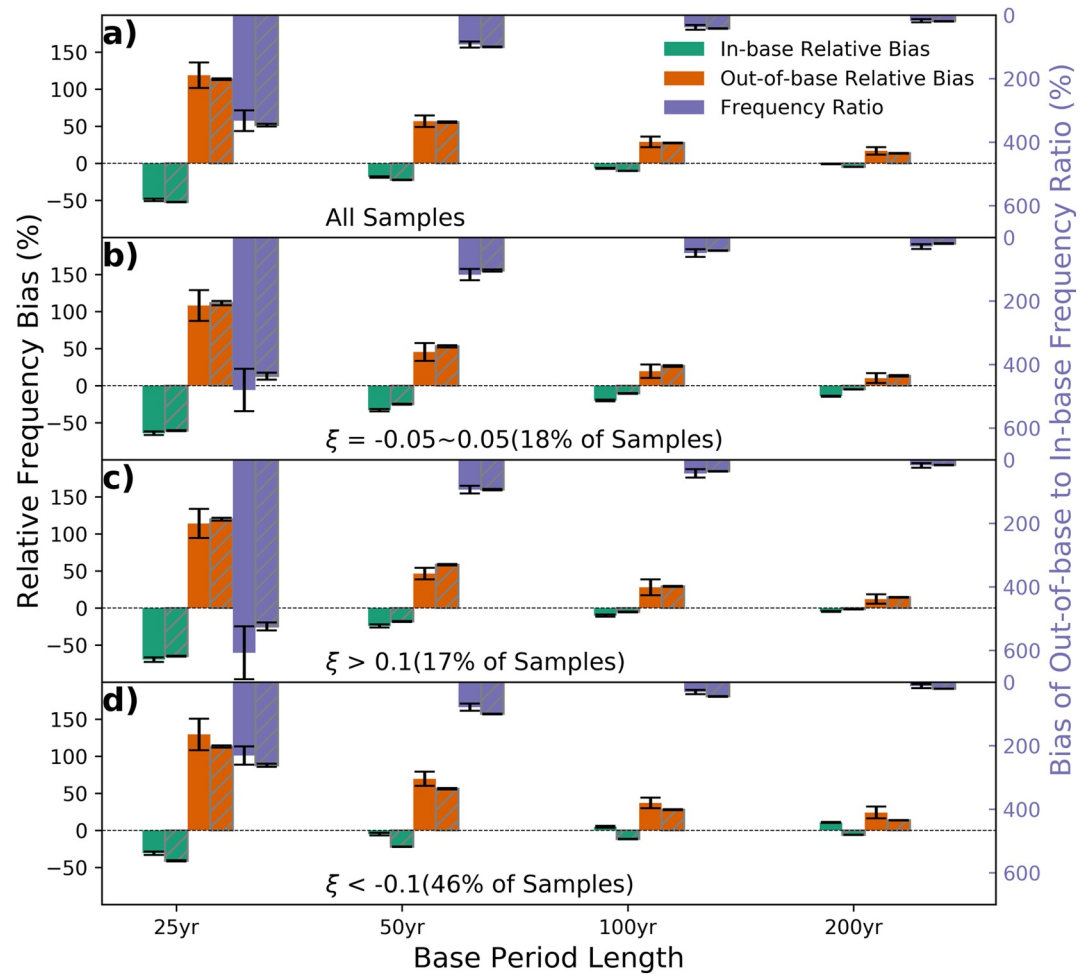
Figure 3. Relative biases in detecting 100-year extreme events, introduced by setting local 100-year extreme thresholds from a base period, across various base period length (25, 50, 100, 200 years), for the full sample (a) and groups with selected shape parameter ranges (percentages of the full sample in brackets) (b)–(d). Data are from GEV fits to 101,877 timeseries of 294-year annual maximum discharge, simulated by LPJmL-IPSL-CM5A-LR under piControl conditions (solid), and emulated by random samples with matched shape parameters but no autocorrelation and parameters $\mu = 0$, $\sigma = 1$ (hatched bars). Metrics include in-base (green) and out-of-base (orange) relative frequency biases, and the bias of the ratio of out-of-base to in-base frequencies (purple, right axis), all as percentages. Error bars show one-standard-deviation ranges estimated from 11 random epoch combinations within the 294-year timeseries.

method aimed at substantially reducing the frequency biases in aggregated statistics. This practical bias adjustment method can also be applied as a post-processing procedure for existing studies.

Assuming that projected global (or in a large region) flood frequency changes are a linear combination of climate change, natural variability, and frequency bias effects (ignoring their interactive effects), we can remove the frequency bias from these projections. We can approximate the frequency bias in hydrological model outputs using Monte Carlo simulations that mimic the model's statistical properties. These approximations can then be used to adjust the estimated flood frequencies.

For the base period, we adjust the aggregated annual frequency of extreme events exceeding the N-year threshold to 1/N %. For example, if 100-year flood frequency averages x% (typically <1% due to IFB) during the base period, we modify it to its theoretical 1% occurrence by multiplying 1/x for all grid cells for each year of the base period. This accounts for natural variability's significant influence on short base periods, which can complicate interpretations in impact assessments. Thus, our adjustment removes both IFB and natural climate variability from the base period.
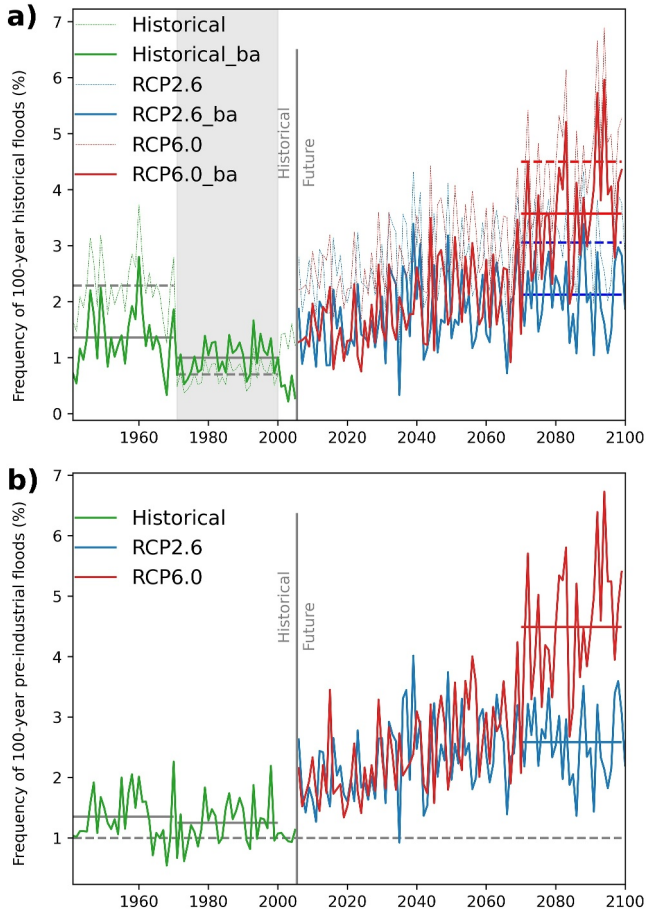
Figure 4. Increase in annual flood occurrence based on LPJmL, driven by IPSL-CM5A-LR across historical and future (RCP2.6 and RCP6.0) periods, using (a) historical 1971–2000, and (b) piControl as the base period. GEV function fits were applied to estimate the return level of the 100-year flood in all grid cells. Dashed (solid) lines in (a) depict original (bias-adjusted) annual time series of spatially averaged 100-year return level exceedances. Horizontal dashed (original) and solid (bias-adjusted) lines denote 30-year mean values for the respective periods.

For out-of-base years, we use Monte Carlo experiments to determine the OFB for the selected base period length, threshold estimation method, and estimated shape parameters. Let x_{ij} represent the projected flood occurrence (1 for above, 0 for below threshold) at grid cell i in year j, and x_j ($x_j = \frac{1}{m}\sum_i x_{ij}$) denote the averaged flood occurrence in year j, where m is the number of grid cells. The adjusted x_{ij} would then be

$$x_{ij}^{adj} = x_{ij}\frac{x_j - OFB(p)}{x_j} \tag{5}$$

This adjusts for the OFB since $\frac{1}{m}\sum_i x_{ij}^{adj} = x_j - OFB(p)$. We prefer this multiplicative OFB adjustment method over the alternative, additive version, $x_{ij}^{adj} = x_{ij} - OFB(p)$, since the multiplicative method avoids negative values in the adjusted flood occurrence.

We illustrate the impact of frequency bias on projected future occurrence of 100-year floods, and the effectiveness of our bias adjustment using experiments described in the previous subsection. First, the 100-year thresholds were estimated by fitting GEV functions to annual maximum discharge timeseries at all non-dry grid cells for the historical (1971–2000) and the 294-year piControl base periods, respectively. Then, the unadjusted annual flood frequencies were computed using these thresholds (Figure 4a, dashed lines, and Figure 4b). Note that the stationarity assumption for the historical base period is a prerequisite for GEV fitting, and nonstationary extreme function fitting methods and their bias adjustment are out of scope for this study. Our bias adjustment procedure (Data Availability) for the historical base period case includes the following steps.

1. Retrieve the shape parameters from the GEV function fit at all non-dry grid cells.
2. Run Monte Carlo simulations to estimate OFB using the retrieved shape parameters, base period length, and distribution name (gev).
3. Compute the mean in-base 100-year flood frequency as x%, then adjust in-base flood frequency by multiplying $1/x$.
4. Compute region-averaged flood occurrence in each year, then adjust out-of-base flood frequency using Equation 5.

For the historical (1971–2000) base period, the unadjusted 100-year flood frequency increased from 0.7% (1971–2000) to 3.05% (4.5%) in 2071–2100 under RCP2.6 (RCP6.0), representing a 336% (543%) increase. Assuming Gumbel distribution data, flood frequency is observed to increase from 0.84% in 1971–2000 to 2.37% (3.91%) in 2071–2100 under RCP2.6 (RCP6.0), a 182% (365%) uptick (Figure S10a in Supporting Information S1).

Applying our frequency bias adjustment method significantly reduces projected increases in flood occurrence. Assuming GEV-distributed data (Figure 4a), the increase in flood frequency diminishes to 112% (257%) under RCP2.6 (RCP6.0). Assuming Gumbel-distributed data (Figure S10 in Supporting Information S1), the revised increases are 101% (255%) under RCP2.6 (RCP6.0). This demonstrates that under low-end global warming scenarios like RCP2.6, the increase due to frequency bias is comparable to the climate change signal for the Gumbel fit and substantially exceeds it with the GEV fit.

Using separate data samples for estimating return levels and flood frequencies, as in Lange et al. (2020), reduces distortions in flood frequency changes due to frequency bias (Figure 4b). This approach employs 294-year pre-industrial control data to estimate 100-year return levels, which are then applied to historical and future flood frequency assessment. Assuming GEV-distributed data, pre-industrial 100-year flood frequency increases from 1.25% to 2.59% (4.49%) from 1971–2000 to 2071–2100 under RCP2.6 (RCP6.0), a +107% (+260%) increase. It

offers a more accurate representation of climate change impacts on flood frequency, as it avoids IRB and significantly reduces ORB with a longer base period. Consequently, the effect of frequency bias adjustment is small, slightly raising flood frequency increases to 116% (281%) for RCP2.6 (RCP6.0).

3. Outlook and Conclusion

Through a combination of analytical derivation, Monte Carlo simulations, and hydrological simulation analyses, we demonstrated that using base period-derived thresholds for evaluating frequency changes in extreme events like floods leads to frequency underestimation within the base period and overestimation beyond it. The degree of these biases inversely scales with the base period's length and is comparable to the climate change signal for low-end warming scenarios like RCP2.6. This is vital for researchers attributing changes in frequencies of floods or other extreme events and their socioeconomic effects. When using the historical period for both extreme function fitting and future change baseline, it should be noted that projected changes are only partly attributable to climate/socioeconomic shifts. For focused studies on climate/socioeconomic impacts, using separate periods for fitting extreme functions and baseline assessment is preferable. We recommend piControl or extended historical simulations are used for improved detection/attribution studies. If the base period must double as a future projection baseline, applying a frequency bias adjustment method, such as outlined in this study, is advised.

Our study's implications go beyond simulating changes in extremes; they are also relevant to flood defense design. Engineers often use local historical data fitted to extreme value distributions to design flood defenses. When averaged over many locations, this could potentially lead to more frequent floods than expected, even without climatic or socioeconomic changes. The shorter the historical base period, the higher the risk of floods exceeding local defenses more often than anticipated. Our findings highlight that this risk depends on the base period's length and the chosen extreme value distribution. This knowledge is crucial for local stakeholders, raising awareness about the potential for more frequent floods than previously thought, independent of climate and socioeconomic changes.

In conclusion, extreme value distribution fitting for detecting changes in extremes must be carefully conducted to avoid overestimation due to threshold exceedance frequency biases. Our study offers methods to separate the impact of climate change and avoid these biases. Nonetheless, this does not contradict previous findings (e.g., Alfieri et al., 2015; Hirabayashi et al., 2013) that flood impact will likely increase under higher levels of future warming, posing substantial adaptation challenges globally.

Data Availability Statement

The runoff data from LPJmL is publicly available through the ISIMIP archive (Gosling et al., 2023, downloaded from https://data.isimip.org/, search "lpjml ipsl isimip2b water qtot", then download "2005soc" files for the picontrol, rcp26 and rcp60 scenarios, and "histsoc" files for historical scenario). Data archiving for the CaMa-Flood simulated annual maximum discharge, the bias adjustment example, and python code for analytical and numerical frequency bias estimation, is publicly available on Zenodo (Zhao & Lange, 2024). Figures were made with Matplotlib version 3.3.3 (Caswell et al., 2020; Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, *113*(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, *19*(5), 2247–2260. https://doi.org/10.5194/hess-19-2247-2015

Caswell, T. A., Droettboom, M., Lee, A., Hunter, J., Andrade, E. S., Firing, E., et al. (2020). matplotlib/matplotlib: Rel: v3.3.3 (Version v3.3.3). [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.4268928

De Luca, P., & Donat, M. G. (2023). Projected changes in hot, dry, and compound hot-dry extremes over global land regions. *Geophysical Research Letters*, *50*(13), e2022GL102493. https://doi.org/10.1029/2022GL102493

Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., et al. (2017). Assessing the impacts of 1.5°C global warming - simulation protocol of the inter-sectoral impact model intercomparison project (ISIMIP2b). *Geoscientific Model Development*, *10*(12), 4321–4345. https://doi.org/10.5194/gmd-10-4321-2017

Gosling, S. N., Schmied, H. M., Burek, P., Chang, J., Ciais, P., Döll, P., et al. (2023). ISIMIP2b simulation data from the global water sector (v1.0). [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.626689

Harrington, L. J., Otto, F. E. L., Cowan, T., & Hegerl, G. C. (2019). Circulation analogues and uncertainty in the time-evolution of extreme event probabilities: Evidence from the 1947 central European heatwave. *Climate Dynamics*, *53*(3–4), 2229–2247. https://doi.org/10.1007/s00382-019-04820-2

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

IPCC. (2014). Summary for policymakers. In C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, et al. (Eds.), *Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change* (pp. 1–32). Cambridge University Press.

IPCC. (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. In press. https://doi.org/10.1017/9781009157896

Lange, S., Volkholz, J., Geiger, T., Zhao, F., Vega, I., Veldkamp, T., et al. (2020). Projecting exposure to extreme climate impact events across six event categories and three spatial scales. *Earth's Future*, *8*(12). https://doi.org/10.1029/2020EF001616

Mahony, C. R., & Cannon, A. J. (2018). Wetter summers can intensify departures from natural variability in a warming climate. *Nature Communications*, *9*(1), 783. https://doi.org/10.1038/s41467-018-03132-z

Ombadi, M., Risser, M. D., Rhoades, A. M., & Varadharajan, C. (2023). A warming-induced reduction in snow fraction amplifies rainfall extremes. *Nature*, *619*(7969), 305–310. https://doi.org/10.1038/s41586-023-06092-7

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., et al. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, *11*(3), 226–233. https://doi.org/10.1038/s41558-020-00972-w

Ringard, J., Chiriaco, M., Bastin, S., & Habets, F. (2019). Recent trends in climate variability at the local scale using 40 years of observations: The case of the Paris region of France. *Atmospheric Chemistry and Physics*, *19*(20), 13129–13155. https://doi.org/10.5194/acp-19-13129-2019

Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J., & Schaphoff, S. (2008). Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, *44*(9), 1–17. https://doi.org/10.1029/2007WR006331

Simolo, C., & Corti, S. (2022). Quantifying the role of variability in future intensification of heat extremes. *Nature Communications*, *13*(1), 7930. https://doi.org/10.1038/s41467-022-35571-0

Sippel, S., Walton, P., & Otto, F. E. L. (2015). Stakeholder perspectives on the attribution of extreme weather events: An explorative enquiry. *Weather, Climate, and Society*, *7*(3), 224–237. https://doi.org/10.1175/WCAS-D-14-00045.1

Sippel, S., Zscheischler, J., Heimann, M., Lange, H., Mahecha, M. D., Jan Van Oldenborgh, G., et al. (2017). Have precipitation extremes and annual totals been increasing in the world's dry regions over the last 60 years? *Hydrology and Earth System Sciences*, *21*(1), 441–458. https://doi.org/10.5194/hess-21-441-2017

Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., & Mahecha, M. D. (2015). Quantifying changes in climate variability and extremes: Pitfalls and their overcoming. *Geophysical Research Letters*, *42*(22), 9990–9998. https://doi.org/10.1002/2015GL066307

Stevenson, S., Coats, S., Touma, D., Cole, J., Lehner, F., Fasullo, J., & Otto-Bliesner, B. (2022). Twenty-first century hydroclimate: A continually changing baseline, with more frequent extremes. *Proceedings of the National Academy of Sciences*, *119*(12), e2108124119. https://doi.org/10.1073/pnas.2108124119

Vogel, R. M. (1986). The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses. *Water Resources Research*, *22*(4), 587–590. https://doi.org/10.1029/WR022i004p00587

von Bloh, W., Rost, S., Gerten, D., & Lucht, W. (2010). Efficient parallelization of a dynamic global vegetation model with river routing. *Environmental Modelling and Software*, *25*(6), 685–690. https://doi.org/10.1016/j.envsoft.2009.11.012

Ward, P. J., Jongman, B., Aerts, J. C. J. H., Bates, P. D., Botzen, W. J. W., Diaz Loaiza, A., et al. (2017). A global framework for future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*, *7*(9), 642–646. https://doi.org/10.1038/nclimate3350

Williams, A. I. L., & O'Gorman, P. A. (2022). Summer-winter contrast in the response of precipitation extremes to climate change over northern Hemisphere land. *Geophysical Research Letters*, *49*(10), e2021GL096531. https://doi.org/10.1029/2021GL096531

Willner, S. N., Levermann, A., Zhao, F., & Frieler, K. (2018). Adaptation required to preserve future high-end river flood risk at present levels. *Science Advances*, *4*(1), eaao1914. https://doi.org/10.1126/sciadv.aao1914

Yamazaki, D., De Almeida, G. A. M., & Bates, P. D. (2013). Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resources Research*, *49*(11), 7221–7235. https://doi.org/10.1002/wrcr.20552

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4), 1–21. https://doi.org/10.1029/2010WR009726

Zhang, X., Hegerl, G. C., Zwiers, F. W., & Kenyon, J. (2005). Avoiding inhomogeneity in precentile-based indices of temperature extremes. *Journal of Climate*, *18*(1999), 1641–1651. https://doi.org/10.1175/JCLI3366.1

Zhao, F., & Lange, S. (2024). Frequency bias estimation of extremes and a practical adjustment method. [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.11207706