Supplement of Earth Syst. Dynam., 15, 225–264, 2024
https://doi.org/10.5194/esd-15-225-2024-supplement

Earth System Dynamics
Open Access

Supplement of

Possible role of anthropogenic climate change in the record-breaking 2020 Lake Victoria levels and floods

Rosa Pietroiusti et al.

Correspondence to: Rosa Pietroiusti (rosa.pietroiusti@vub.be)

S1 Statistical methods

The extreme event attribution methodology applied is based on Philip et al. (2020), and summarized hereafter.

S1.1 Step 1: Analysis trigger

The first step in probabilistic extreme event attribution study is to select the event that will be studied, and to justify its choice based on scientific and societal interest, and the capacity to carry out the study (Philip et al., 2020; van Oldenborgh et al., 2021).

S1.2 Step 2: Event definition

The second step in EEA is to create a class-based univariate definition of the observed extreme event (Philip et al., 2020). Different possible event definitions were compared based on the observed lake levels. Namely, rates of change in lake levels, calculated over time-periods of different sizes are compared. The possible event definitions were assessed for their ability to reflect the hydro-meteorological rarity of the event, while not excessively violating the assumptions behind extreme value theory, in particular those related to the number, independence and identical (or linearly non-stationary, according to the shift fit method) distribution of observations in a sample. Based on the chosen event definition and the optimal time-window (Δt) selected, the climate variable of interest (X) is defined as:

$$X = \frac{\Delta L}{\Delta t} \tag{S1}$$

The observed 2020 flood event is defined in a class-based way as the rate of change in lake levels leading up to the period of peak lake levels:

$$x_{2020} = \left(\frac{\Delta L}{\Delta t}\right)_{2020} \tag{S2}$$

The probability of observing an event as extreme as that observed in 2020 can therefore be indicated as a probability of exceedance of a threshold:

$$P\left(X \geq x_{2020}\right) \tag{S3}$$

S1.3 Step 3: Observed probability and trend

The third step in a probabilistic EEA study is to estimate the return period of the observed event in the current climate and to test whether a trend in the likelihood of the observed event is detectable in observations. Here, observational lake levels are used to estimate the probability of the flood in the current climate. First, based on the event definition, the variable of interest (X = ΔL/Δt) is extracted from the lake level time series for the years 1949-2020. This is done with a rolling window methodology, so that the extracted time series has a daily temporal resolution, like the original lake level time series. This time series can be represented as the series of a random variable X_i with a cumulative distribution function $F(x)$. :

$$X_1, X_2, X_3, ...X_n \quad X_i \sim F(x)$$
$$P(X \leq x) = F(x) \tag{S4}$$

In order to be fitted to extreme value distributions, the variable (X) must be assumed to be independently and identically distributed (iid). If there is a known deviation from identical distribution, for example due to non-stationarity, this must be accounted for and modelled statistically (Slater et al., 2021). For example, non-stationarity is present nearly everywhere in variables related to temperatures, due to thermodynamic changes induced by climate change. In general, therefore, the probability distribution of temperatures has been changing and shifting, with higher temperatures becoming more likely through time. The parameters of the distributions that are fitted to the data can be modelled in such a way as they vary through time, for example by using a covariate, in order to account for changes in the location or variability of the distributions.

In a second step, a block maxima approach is applied to the time series. For each calendar year the maximum observation is extracted, creating an annual block maxima time series with one observation per year:

$$M_n = max\{X_1, X_2, X_3, ...X_n\} \quad \text{for } n = 365 \text{ days} \tag{S5}$$

$$M_{n,1}, M_{n,2}...M_{n,m} \quad \text{for } m \text{ years} \tag{S6}$$

The maxima M_n can be represented as a random variable, and given the iid assumption on its parent variable X, the cumulative distribution function of the maxima can be represented as:

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x \cap X_2 \leq x ... \cap X_n \leq x) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x) ... \cdot P(X_n \leq x) \\ &= F(x) \cdot F(x) ... \cdot F(x) \\ &= F^n(x) \end{aligned} \tag{S7}$$

In extreme value theory, it is assumed that the cumulative distribution function of a block maxima time series, such as $F^n(x)$, is described by a Generalized Extreme Value (GEV) distribution. This is based on the Fisher–Tippett–Gnedenko Extremal Types Theorem, which states that if a random variable X is independent and identically distributed and there is convergence at its tail (i.e. the distribution of maxima M_n converges to one shape as the number of observations n becomes very large), then the limit distribution (i.e. for n tending to infinity) of the maxima M_n must be in the GEV family, independently of the parent distribution $F(x)$ (Coles, 2001). This can be represented as:

$$F^n\left(\frac{x - a_n}{b_n}\right) \xrightarrow[n \to \infty]{} G(x) \tag{S8}$$

Eq. S8 states that if a series of numbers a_n and b_n can be found such that as n increases to infinity the cumulative distribution function, $F^n(x)$, can be shifted and rescaled to always result in the same distribution $G(x)$, then there is a convergence of the distribution, and the limit distribution of the maxima time series is $G(x)$. The Extremal Types Theorem states that $G(x)$ must be one of the three distributions in the GEV family, because these are the only max-stable functions, which can therefore meet the requirement in Eq. S8. The GEV family contains the Weibull, Fréchet and Gumbel distributions, which can be represented collectively by the following cumulative distribution function (Coles, 2001):

$$G(x; \mu, \sigma, \xi) = P(M_n \leq x) = exp\left[-\left(1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right)^{-1/\xi}\right] \tag{S9}$$

GEV distributions are described by the location (μ), scale (σ) and shape (ξ) parameters. These respectively represent the central location of the maxima, the spread of the maxima, and the shape of the distribution. The Gumbel distribution has ξ = 0 and has a non-zero probability density for all values of the variable x (unbounded), the Fréchet distribution has $\xi > 0$ and is heavy tailed (has a lower bound) and finally, the Weibull distribution has $\xi < 0$ and is thin tailed (has an upper bound).

In a stationary fit, the distribution parameters are assumed to remain constant, whereas in non-stationarity conditions, the distribution parameters can vary over time.

The GEV in Eq. S9 gives the probability that a maximum will not exceed a certain threshold (x). The probability of threshold exceedance can therefore be calculated as:

$$P(M_n \geq x) = 1 - G(x) \tag{S10}$$

If $M_{n,k}$ is an annual block maxima time series, the probability of threshold exceedance ($P(M_n \geq x)$ represents the probability of observing a maximum in any given year that will exceed the threshold value x. The return period τ of a threshold value x indicates the average number of years that is expected to elapse between two observations that exceed the threshold value, and can be calculated as the inverse of the probability of exceedance:

$$\tau(x) = \frac{1}{P(M_n \geq x)} = \frac{1}{1 - G(x)} \tag{S11}$$

The annual block maxima time series of the defined climate variable ($M_{n,k}$) is fitted to a GEV distribution to estimate the distribution parameters with a maximum likelihood method. Based on the fit, the best estimate of the return period in the current climate of the flooding event observed in 2020 is calculated, and the 95% confidence interval is estimated using a 1000-member non-parametric bootstrap, with all years assumed to be independent. The observed 2020 event is included in the fit. The probability of observing the 2020 event is calculated as:

$$p_{obs} = P\left(M_n \geq x_{2020}\right) = 1 - G(x_{2020}; \mu, \sigma, \xi) \tag{S12}$$

To account for possible non-stationarity due to climate change, and to test whether a trend in the likelihood of observing the event is detectable in observations, distribution parameters are allowed to vary through time. A non-stationary GEV model is fitted to the block maxima time series, using a global mean surface temperature (GMST) anomaly time series smoothed with a 4-year rolling window (T') as a covariate. Following an EEA methodology commonly applied to temperature variables and known as a shift fit (Philip et al., 2020), it is assumed that the distribution shape and spread (ξ and σ) remain unchanged, and that only the central position of the distribution shifts with GMST. The location parameter μ is modelled as a linear function of GMST, the parameters of which (μ_0 and μ_1) are estimated simultaneously with the time-constant parameters:

$$\mu = \mu_0 + \mu_1 T' \tag{S13}$$

Two GEV distributions can then be modelled: the first one representing a counterfactual climate, where the location parameter is determined as a function of a pre-industrial GMST (μ_{ref}), and the second one representing the current climate, with the location parameter determined as a function of present-day GMST (μ_{new}). Pre-industrial GMST is defined as the temperature in 1900 and present-day GMST as that in 2020, and the location parameters are calculated as follows:

$$\begin{aligned} \mu_{new} &= \mu_0 + \mu_1 T'_{2020} \\ \mu_{ref} &= \mu_0 + \mu_1 T'_{1900} \end{aligned} \tag{S14}$$

If a positive or negative linear trend is identified between the location parameter and smoothed GMST based on Eq. S13, there will by definition be a shift between the two GEV distributions. Whether this is statistically significant or not depends on the magnitude of the linear relationship and the shape and confidence intervals of the GEV distributions. The probability of exceeding the observed threshold (x_{2020}) is then estimated in a 1900-like climate and in 2020-like climate based on the two shifted GEV fits. These GEV models have the same scale and shape parameters, and only differ in their location parameters. The probabilities of threshold exceedance in 2020 and 1900 are then compared, and the change in likelihood of observing the event is expressed as a probability ratio (PR), which indicates how many times more or less likely the event has become through time. Confidence intervals are estimated using bootstrapping. This can be summarized as follows:

$$p_{new} = P^{2020}\left(M_n \geq x_{2020}\right) = 1 - G(x_{2020}; \mu_{new}, \sigma, \xi)$$
$$p_{ref} = P^{1900}\left(M_n \geq x_{2020}\right) = 1 - G(x_{2020}; \mu_{ref}, \sigma, \xi) \tag{S15}$$

$$PR_{obs} = \frac{p_{new,obs}}{p_{ref,obs}} \tag{S16}$$

The change in magnitude or, equivalently, intensity (ΔI) of the event is also estimated by comparing the magnitude of the observed event (x_{2020}) with the magnitude of an event that would have the same probability of occurrence in a 1900-like climate (x_{1900}). If the event has been made more intense by a trend in the observational data, the change in intensity will be positive:

$$\Delta I = x_{2020} - x_{1900} \quad \text{for} \quad P^{2020}\left(M_n \geq x_{2020}\right) = P^{1900}\left(M_n \geq x_{1900}\right) \tag{S17}$$

When a shift fit is carried out the change in intensity does not depend on the return level (that is, the magnitude of the observed 2020 event), but only on the modelled linear regression of the location parameter expressed in Eq. S13, and can be calculated as follows (Philip et al., 2020):

$$\Delta I = \mu_1 (T'_{new} - T'_{ref}) \tag{S18}$$

S1.4 Step 4: Model evaluation

The fourth step in EEA is to determine whether models are fit for purpose to study the extreme event in question. To this end, first, biases in lake levels simulated with the WBM when forced with PERSIANN-CDR observational precipitation data are investigated, to assess the skill of the WBM. Second, the precipitation output from factual GCM simulations is evaluated using the PERSIANN-CDR observational product to assess spatial and seasonal biases in precipitation amounts. Third, the WBM is forced with precipitation from factual and counterfactual GCM simulations. The modelled lake levels are evaluated to determine whether the water balance closes or whether a further bias correction of the water balance terms is required. Fourth, the climate variable of interest (X) is extracted from each modelled lake level time series. Annual block maxima of each time series are then fitted to GEV distributions to estimate the three model parameters μ, σ and ξ. The resulting distribution parameters are compared with those derived when fitting observational data to a GEV to assess the similarity of the parameters and the skill of the GCMs to model the observed tail behaviour of the variable. If the confidence intervals of the distribution parameters overlap the modelled distributions can be assumed to be sufficiently similar to the observations (Philip et al., 2020).

S1.5 Step 5: Multi-model extreme event attribution

The fifth step in EEA is to use GCM simulations to estimate the return period of the observed event under factual (present-day) and counterfactual (pre-industrial) climates. Results are expressed as probability ratios, which indicate how much more or less likely the event has become due to climate change. Since the anthropogenic forcing in historical GCM simulations changes through time, we apply non-stationary methods used for transient forcing simulations as described in Philip et al. (2020).

First, GEV distributions are fitted on the annual block maxima time series of the climate variable X modelled by the WBM using GCM precipitation simulations with all historical emissions (hist). For each GCM, a non-stationary GEV model is fitted to the time series, using the shift-fit methodology. The smoothed yearly time series of GMST as simulated by each model is used as a covariate. As for the observations, the modelled fit is used to estimate the probability of exceeding the observed threshold in a present-day climate ($p_{new,hist}$) and in a pre-industrial climate ($p_{ref,hist}$), respectively represented by the years 2020 and 1900. The result is summarized by calculating a probability ratio for each GCM:

$$PR_{GCM,hist} = \frac{p_{new,hist}}{p_{ref,hist}} \tag{S19}$$

To estimate uncertainty due to internal variability, two-sided 95% confidence intervals are calculated using bootstrapping. If the confidence interval of the probability ratio is above 1, a statistically significant increase in the likelihood of the event is detected.

Second, for each GCM, the same process is repeated using the time series obtained by running the WBM with precipitation from GCMs with natural-only forcings (hist-nat simulations). Since hist-nat simulations represent a hypothetical climate where anthropogenic greenhouse gas emissions have not occurred, this is equivalent to testing whether a trend in the likelihood of threshold exceedance is detectable due to natural forcings alone (Philip et al., 2020). Smoothed yearly GMST modelled by the hist-nat GCM simulations is used as a covariate. Probability ratios are calculated between 2020 and 1900:

$$PR_{GCM,hist\text{-}nat} = \frac{p_{new,hist\text{-}nat}}{p_{ref,hist\text{-}nat}} \tag{S20}$$

Two-sided 95% confidence intervals for the hist-nat estimates are then calculated. If the confidence interval of the probability ratio includes 1, a trend due to natural forcings is not detected. According to Philip et al. (2020) a trend detected based on historical simulations of the same GCM can then be attributed to climate change, following the methodology used, for example, in Philip et al. (2018).

Third, results from the historical and hist-nat fits for each GCM are combined as in Philip et al. (2018), computing two-sided 95% confidence intervals for the resulting probability ratio:

$$PR_{GCM,combined} = \frac{p_{new,hist}}{p_{new,hist\text{-}nat}} \tag{S21}$$

Since GCMs have biases regarding precipitation amounts, which are specific to each GCM, a simple bias correction is applied following one of the methods proposed in Philip et al. (2020). Namely, for each GCM, an adapted threshold is used to represent the observed extreme event (x_{GCM}) instead of directly using the magnitude derived from observatons (x_{2020}). The threshold is chosen by holding fixed the probability of exceedance in the present-day climate that was calculated based on observational data ($p_{new,obs}$ in Step 3). For each hist GCM fit, we then determine what magnitude of the climate variable has the same return period in the present-day climate as simulated by the GCM (x_{GCM}):

$$\begin{aligned}
p_{new,obs} &= P^{2020,obs}\left(M_n \geq x_{2020}\right) \\
p_{new,hist} &= P^{2020,hist}\left(M_n \geq x_{GCM}\right) \quad \text{for } x_{GCM} \text{ s.t. } p_{new,hist} = p_{new,obs}
\end{aligned}$$

For the natural forcings only simulations, the same GCM-specific threshold calculated based on the historical simulations is used (x_{GCM}), assuming model biases are the same in the historical and natural forcing only simulations.

GEV fits and the calculation of probabilities and probability ratios are carried out in R using the extRemes package (Gilleland and Katz, 2016).

S1.6 Step 6: Attribution synthesis

The sixth step in EEA is to synthesise the probability ratios calculated from observations and each GCM and to derive a single probability ratio. Here, probability ratios calculated between present-day and pre-industrial climates from observations and from combined historical and hist-nat GCM simulations, are synthesised:

$$PR_{synthesis} = f(PR_{obs}, PR_{GCM,1}, PR_{GCM,2}...PR_{GCM,n}) \tag{S22}$$

First, the model PR estimates are synthesised. If inter-model spread is greater than intra-model spread, this suggests uncertainty due to model differences is greater than uncertainty due to natural or internal variability. In this case, the Philip et al. (2020) protocol suggests using an unweighted synthesis methodology, which keeps the uncertainty on the final PR larger. If inter-model spread is small compared to inter-model spread, then Philip et al. (2020) suggests to use a weighted synthesis methodology, which makes the resulting uncertainty on the final PR smaller. The weight applied is the inverse square of the variability, so that more uncertain estimates are penalised (Harrington et al., 2021; Philip et al., 2021). Which source of uncertainty dominates is estimated by calculating a χ^2 statistic, which is the sum of the number of standard deviations squared that the synthesised mean is away from the various best estimates, and dividing this by the degrees of freedom (dof), which is one minus the number of estimates being compared. If χ^2/dof is lessgilleland than or close to unity, natural variability dominates, whereas if it is much larger than unity model disagreement dominates and uncertainty is larger than natural variability alone would suggest (Philip et al., 2020). The synthesis methodology assumes the logarithm of the PRs is normally distributed, and that each PR result is independent. Possible PR values from each modelled fit are then weighed by the probability density of that value. The unweighted synthesis methodology inflates the uncertainty of each model estimate by $\sqrt{\chi^2/dof}$ (when χ^2/dof is greater than 1) to keep the confidence intervals larger.

Second, the model PR is averaged with the PR from observations, either using a weighted synthesis methodology, where the weight is the inverse square of the variability, or using an unweighted synthesis methodology, which keeps the uncertainty interval larger and usually gives more importance to observations, to account for model errors as a source of uncertainty.

Although uncertainty due to intra-model spread is greater than uncertainty due to inter-model spread in our results, we present results only from an unweighted synthesis methodology to avoid artificially reducing uncertainties. The synthesis is carried out using the KNMI-WMO Climate Explorer.

References

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.

Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in *R*, Journal of Statistical Software, 72, https://doi.org/10.18637/jss.v072.i08, 2016.

Harrington, L. J., Wolski, P., Pinto, I., Ramarosandratana, A. M., Barimalala, R., Vautard, R., Philip, S., Kew, S., Singh, R., Heinrich, D., Arrighi, J., Raju, E., and Thalheimer, L.: Attribution of Severe Low Rainfall in Madagascar 2019-2021, Tech. rep., World Weather Attribution, 2021.

Philip, S., Kew, S. F., van Oldenborgh, G. J., Otto, F., O'Keefe, S., Haustein, K., King, A., Zegeye, A., Eshetu, Z., Hailemariam, K., Singh, R., Jjemba, E., Funk, C., and Cullen, H.: Attribution Analysis of the Ethiopian Drought of 2015, Journal of Climate, 31, 2465–2486, https://doi.org/10.1175/JCLI-D-17-0274.1, 2018.

Philip, S., Kew, S., van Oldenborgh, G. J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., and van Aalst, M.: A Protocol for Probabilistic Extreme Event Attribution Analyses, Advances in Statistical Climatology, Meteorology and Oceanography, 6, 177–203, https://doi.org/10.5194/ascmo-6-177-2020, 2020.

Philip, S. Y., Kew, S. F., Oldenborgh, G. J. V., Yang, W., Vecchi, G. A., Anslow, F. S., Li, S., Seneviratne, S. I., Luu, L. N., Arrighi, J., Singh, R., Aalst, V., Hauser, M., Schumacher, D. L., Marghidan, C. P., Ebi, K. L., Vautard, R., Tradowsky, J., Coumou, D., Lehner, F., Rodell, C., Stull, R., Howard, R., Gillett, N., and Otto, F. E. L.: Rapid Attribution Analysis of the Extraordinary Heatwave on the Pacific Coast of the US and Canada June 2021 ., World Weather Attribution, pp. 119–123, 2021.

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., and Wilby, R. L.: Nonstationary Weather and Water Extremes: A Review of Methods for Their Detection, Attribution, and Management, Hydrology and Earth System Sciences, 25, 3897–3935, https://doi.org/10.5194/hess-25-3897-2021, 2021.

van Oldenborgh, G. J., van der Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., King, A., Lott, F., Arrighi, J., Singh, R., and van Aalst, M.: Pathways and Pitfalls in Extreme Event Attribution, Climatic Change, 166, 1–27, https://doi.org/10.1007/s10584-021-03071-7, 2021.