

Reply to comments by Reviewer #2

We thank Reviewer #2 for the helpful comments, and for the positive view of our manuscript.

We address each point in detail below.

General comments

The manuscript “ATTRIBUTION OF CHANGES IN SMALL AND LARGE FLOODS ACROSS BRAZIL” by Anzolin et al. analyses trends in floods of different magnitudes in Brazil and the relative contribution of two drivers (peak rainfall and antecedent wetness state) to these changes. The authors find that in low water storage capacity regions, different drivers explain trends in floods of different magnitude, while, in high water storage capacity regions, antecedent wetness explains both. The analyses and findings are clearly presented and well supported by figures. Please find my comments below.

R: We thank Reviewer #2 for the detailed read of our manuscript and constructive comments.

We address each point in detail below.

Major comments

Major#1. L11: “corresponding to 80% of substantial flood increases “. Unclear, please rephrase.

R: We have now clarified this point in the abstract of the revised manuscript (please refer to L11-L12), and throughout the manuscript (see Major#6 and Major#11). We define substantial flood increases as trends exceeding 13.5% per decade (one standard deviation of the flood-change distribution). Within this subset of substantial increases, 80% occurs in large flood events, indicating that strong positive changes are more frequently associated with large floods.

Major#2. I have some doubts about the event selection criteria adopted. Authors sample flood events based on a peak-over-threshold approach based on rainfall magnitudes and further refine the selection by excluding small flood peaks. Their rationale is not to bias the selection towards events with wet antecedent conditions. However, peak rainfall is also a driver used later in the attribution analysis and I believe that the event selection should be independent of the drivers. With the current criteria, events with very wet initial conditions and very small event rainfall are excluded from the analysis. Isn't the selection then biased towards high rainfall events? Also the fact that further refinement (too small flood peaks are later excluded) is, to me, an indication that this is more an "event selection" rather than a "flood event selection". I suggest repeating the analysis with different event selection criteria (e.g. pot or annual maxima based on floods only) that are independent on the drivers at least to prove that the selection is not biasing the results.

R: We agree that the choice about the event sampling is very important to flood change assessment. We have now clarified these points in the revised manuscript (please, refer to L74-L78).

Regarding the event sampling, the first step consists of excluding discharge peaks smaller than the median daily streamflow. This initial filtering ensures that the analysis focuses on hydrologically meaningful flood responses rather than minor runoff fluctuations. Only after this step are rainfall peaks used within a peak-over-threshold framework to identify the events associated with the remaining discharge peaks.

We agree that sampling events based on rainfall peaks may reduce the number of events with very wet antecedent conditions combined with small rainfall. Event selection based solely on flood peaks tend to preferentially sample events occurring under wetter antecedent conditions, which may hide the role of rainfall variability. For instance, drier conditions at the onset of the rainfall event would cause a decrease flood magnitude even with increasing rainfall. Our rainfall-based event sampling allows us to better capture events occurring across a wider range of antecedent wetness conditions and, therefore, better assess the role of both rainfall peaks and antecedent wetness in explaining flood changes.

We have now addressed the concern that using rainfall peaks for event selection could bias the attribution analysis (since rainfall is also a driver in the model). We repeated the attribution analysis using a purely streamflow-based peak-over-threshold approach. In

this alternative analysis, only events with streamflow peaks above a threshold ensuring, on average, three events per year are retained in the sample (Fig. R1).

The results show that the main attribution patterns are broadly consistent with those obtained using the original sampling strategy. The overall findings remain quite consistent, but the differences between the relative contributions are less pronounced in most hotspots. This is expected, as streamflow-based selection naturally favors events driven by higher antecedent soil moisture, which attenuates the role of rainfall. This consistency across sampling strategies suggests that the main conclusions of the attribution analysis are not strongly sensitive to the event selection procedure.

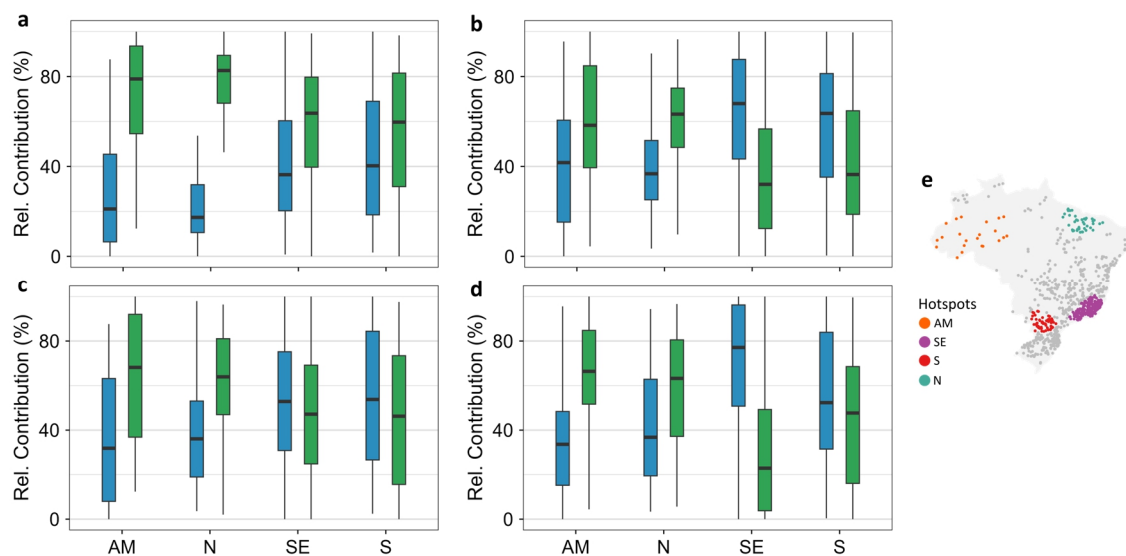


Figure R1. Estimated relative contributions of drivers to small and large flood events. Relative contribution of rainfall peak (blue) and antecedent wetness (green) to flood changes for: (a) small events; and (b) large events using rainfall-based peak-over-threshold sampling. (c-d) as in (a-b), but for streamflow-based peak-over-threshold sampling. (e) Brazilian hotspots.

Major#3. The attribution model is based on local regressions between flood quantiles and the drivers. I am wondering how robust the regressions and results over large floods are with limited flood series. I suggest that the authors could add something about the uncertainty of the results. Perhaps a regional approach (e.g. over the hotspots) could increase the robustness of the attribution results.

R: We agree that regional approaches enhance the robustness of attribution results, a relevant issue that was also highlighted by Reviewer #1. To assess the robustness of our results with respect to our modeling choice, we conducted an additional analysis using panel quantile regression. We run a panel regression model with fixed effects by polling catchments across Brazil and across our four hotspots (Fig. R2). We find that the regional median elasticities derived from the catchment-specific regressions are broadly consistent with those obtained from the panel quantile regression model. In our interpretation, the consistency between the estimates indicates that the results are unlikely to be methodological artifacts but rather reflect an intrinsic characteristic of the data. We have added a brief discussion of these methodological choices and their implications in the Material and Methods section of the revised manuscript (please, refer to L154-L158).

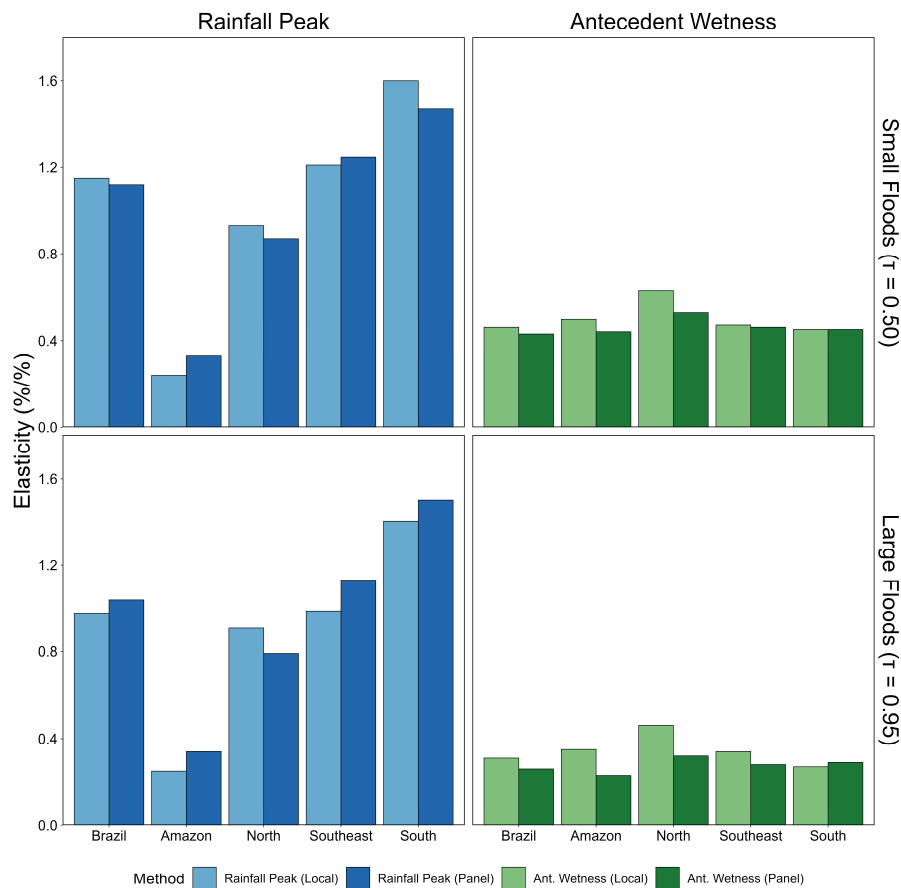


Figure R2: Comparison between elasticities estimation between local and panel quantile regression for small (0.50 quantile) and large floods (0.95 quantile).

We have decided to retain the local quantile regression approach, as adopting a panel regression approach would require assuming that, within a given region, all catchments share a common average sensitivity (i.e., elasticity) of flood magnitudes to rainfall peak

and antecedent wetness, potentially hiding important catchment-to-catchment heterogeneity in hydrological response. Our current framework is based on catchment-specific (local) quantile regressions, which allow us to explicitly account for spatial variability in elasticities and assess how sensitivities differ across individual catchments.

Regarding model uncertainties, we provide a map with standard errors (SE) of our model parameters in the Support Information of the revised manuscript (Fig. S4). As expected, SE increases with event magnitude (Fig. R3). However, for most catchments, the estimated SE is smaller than the parameter itself, indicating that uncertainty is less than the value being estimated. This occurs in 93% of catchments for small events (both for rainfall peak and antecedent wetness). For larger events, 75% of catchments show the same pattern for rainfall peak, compared to 67% for antecedent wetness. These findings indicate that the modeling framework yields robust and reliable estimates across varying event magnitudes, with uncertainty remaining well-controlled in most catchments.

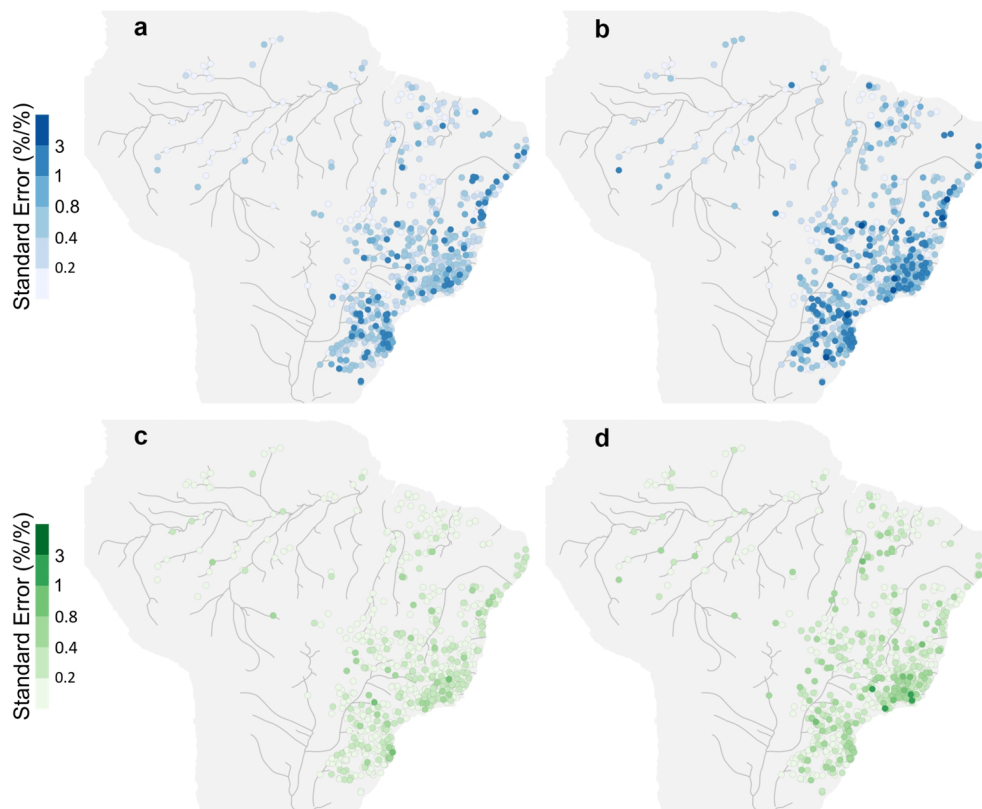


Fig. R3. Estimated standard errors (SE). Rainfall peak elasticities SE for (a) small events, and (b) large events. (c-d) As in (a-b), but for antecedent wetness. Note that the SE associated with rainfall peak are larger because the SE scales with the magnitude of the parameter. Since rainfall peaks elasticities are larger, their SE are also larger; however, this does not necessarily imply greater relative uncertainty.

Major#4. Small and large floods are defined by 0.5 and 0.95 flood quantiles. It would be useful to add a statement about equivalent return periods of such flood events in order to make findings comparable with previous studies and also more easily interpretable by practitioners.

R: We agree that providing an equivalent return period is helpful to make our results more comparable with previous studies and easier to interpret by practitioners. Following Stedinger et al. (1993), the non-exceedance quantile (τ) associated with an annual recurrence interval T is given by:

$$\tau = 1 + \frac{1}{\lambda} \log_e \left(1 - \frac{1}{T} \right) \quad (\text{Eq. R1})$$

where λ is the average sampling frequency. Here, we sample our events using a frequency-based rainfall peak-over-threshold approach, ensuring, on average, three events/year/catchment ($\lambda = 3$). Therefore, small events (0.50 quantile) correspond to an equivalent return period of approximately 1.3 years, while large events (0.95 quantile) correspond to an equivalent return period of approximately 7.2 years. These values are consistent with the relatively high sampling frequency of our peak-over-threshold approach, which allows us to characterize both frequent and relatively rare flood responses within each catchment. We have now included the equivalent return periods of our selected quantiles in the revised manuscript (please, refer to L132-L133).

REFERENCES

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency analysis of extreme events, in: *Handbook of Hydrology*, edited by: Maidment, D. R., McGraw-Hill, New York, 18.1–18.66, 1993.

Major#5. What about catchment area effect? Your catchments vary across 5 or 6 orders of magnitude. One would expect that different processes may be relevant.

We agree that both flood changes and their underlying mechanisms are likely scale dependent. We have now included a new discussion topic about scale dependence of flood changes (please, refer to L332-L350).

“Catchment size is expected to influence temporal changes in flood magnitudes and their causing mechanisms (Sharma et al., 2018). In smaller catchments, flood responses are

typically faster and more directly linked to intense rainfall, whereas in larger basins, processes such as storage and river routing may attenuate and delay flood peaks. Globally, increases in annual maximum flood are more often observed in small catchments (Do et al., 2017). Similarly, using a regional nonstationary modeling framework, Bertola et al. (2020) found that small catchments are more likely to exhibit increasing flood magnitudes. In this study, however, no such pattern was observed, as the Spearman correlation between trends in small and large floods and catchment area was close to zero (on the order of 10^{-2}). This is consistent with findings for annual maximum series in Brazil (Chagas et al., 2022b).

Regarding the scale dependence of flood change mechanisms, the elasticities of flood magnitudes to rainfall peaks and antecedent wetness are largely independent of catchment size. The Spearman correlation between elasticity to rainfall peaks and catchment size is -0.18 for small events and -0.006 for large events, meaning that the sensitivity of flood magnitudes to rainfall peaks decreases weakly with increasing catchment size for small events, and no clear relationship for large events. In contrast, elasticities to antecedent wetness show correlations of about 0.14 for both small and large events, suggesting a slight increase in its influence with catchment size. Although the scaling of flood processes with catchment size is broadly consistent with theoretical expectations, the weak relationships we found suggest that catchment size is not a dominant control on either flood trends or their driving mechanisms in our dataset. This indicates that the processes driving flood changes are largely consistent across a wide range of catchment sizes, with no clear or systematic scale dependence emerging at the continental scale.”

Major#6. L173-175: unclear sentence. Do you mean heavier distribution tails?

R: This sentence has the same interpretation as the sentence presented in the abstract. We have now clarified this point in the revised manuscript (please refer to L194-L197). For more details, please see Major Comment #1.

Major#7. Section 3.2: please check units of elasticity. Is it % or %/% (as it appeared to be defined in section 2.3)?

R: We have checked and clarified the units of elasticity in the revised manuscript. Elasticities are expressed as %/% (percentage change in flood magnitude per percentage change in the driver), consistent with the definition provided in Section 2.3.

Major#7. L194: is the median calculated over the hotspots?

R: That is correct; those values are the median over the hotspots. We have now clarified this information in the revised manuscript (please refer to L215).

Major#9. Figure 7: it would be useful to see results about the absolute contributions as well. Those are useful because show whether drivers contribute to positive or negative flood changes. Also, relative contributions alone may return a biased picture in cases where absolute changes are small.

R: We have now provided a map with the absolute values of the contribution (Fig. R4) in the Support Information of the revised manuscript, please see Fig. S2.

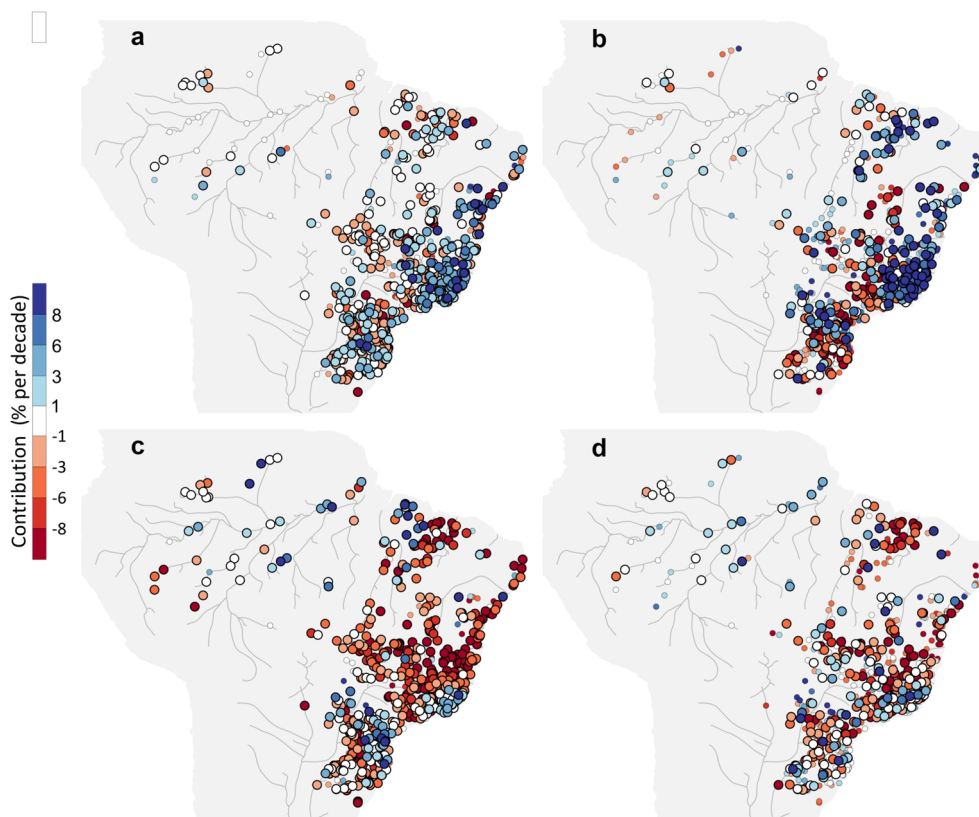


Figure R4. Absolute contribution of drivers to flood change. (a-b) Absolute contributions of rainfall peak to flood change for (a) small and (b) large events. (c-d) As in (a-b), but

for absolute contributions of antecedent wetness to flood change. Large circles with thick borders in (a-d) indicate significance at an FDR-adjusted significance level ($\alpha_{\text{FDR}} = 10\%$).

Major#10. L259-270: I find that the comparison of the findings with the literature should be a bit clarified. No direct comparison is possible because of the different small/large flood definitions (here I think large floods correspond to ~10year return period, which is not really comparable to 100yr or event with significant initiation. Also the (flood) event selection used here and, in some cases, the local vs regional nature of the findings, does not allow a direct comparison.

R: We agree that a direct comparison with previous studies is limited due to differences in the definition of small and large floods. In our case, the empirical return periods associated with the events range from 1.3 to 7.2 years, which differ from studies focusing on more extreme floods (e.g., ~100-year events) or floods defined based on damage or impact thresholds. We have clarified this point in the manuscript and now frame the comparison with previous studies as a discussion of general patterns and mechanisms rather than a strict comparison of flood magnitudes or return periods (please, refer to L265-267)

Major#11. L303: unclear sentence.

R: This sentence has the same interpretation as the sentence presented in the abstract. We have now clarified this point in the revised manuscript (please refer to L354-L357). For more details, please see Major Comment #1.