

Comment on cp-2021-160

Anonymous Referee #1

Referee comment on "Influence of warming and atmospheric circulation changes on multidecadal European flood variability" by Stefan Brönnimann et al., Clim. Past Discuss., <https://doi.org/10.5194/cp-2021-160-RC1>, 2022

This study uses streamflow, daily weather data, reanalyses, and reconstructions to explore variations in European flood frequency during the past 2 centuries and reconcile indications of European flood frequency to be higher under warm as well as cold background climate conditions. The study illustrates how the relative role of atmospheric circulation and moisture content for moisture flux convergence, hence precipitation, changed historically. Based on the moisture content contribution becoming predominant in recent, warmer times, the manuscript discusses implications of such findings for projected floods.

I found the study overall well-conceived and the manuscript well organized and well written. I especially appreciated the efforts to combine different sources of information including observations, model output and reconstructions. I found the analysis overall sound and the conclusions well supported by the results. However, I have a few comments on the study that I ask the Authors to consider in a revised version of the manuscript.

On a general note, I would appreciate a stronger focus on the statistical analyses supporting the existence of linkages between the considered processes, for instance in terms of significance of co-variability between time series. I provide a few specific comments below to illustrate the occasions when I felt the interpretation of results requires further support. Similarly, the comparison between information from different sources appears occasionally to be only qualitative. This left me wondering about the purpose of some of the comparative analyses provided in the manuscript: central or just ancillary to show uncertainty? I think the manuscript would benefit from a bit more guidance by the authors about the purpose (and expected outcome) of some of the analyses. Again, I highlight the few occasions when this occurred in the specific comments below.

Concerning the adopted methodology, the only main question I have regards the

normalization: If I understand the method correctly, normalization is over the whole length of a time series. The relative representation of trends is affected by the fact that series of different length are compared (those including the trend period and those extending further back, during period of little or no trend). If this is true, maybe a word of caution on this approach when comparing trends in figures 2 and 3 or when different discharge time series are averaged out in Figure 1b can be appropriate. An alternative approach could be to normalize over periods as similar as possible across the different series (for instance using 1900-2000). A few more specific methodological questions follows in the specific comments.

Specific comments

Section 2.1: The authors use annual maximum streamflow as a reference for their analysis of floods. What if more floods occur within one year? In my understanding this possibility is not accounted for in the analysis, but might be relevant for the overall assessment of flooding statistics. In section 2.4, it comes clear that daily streamflow series are available for only two stations, so I guess this aspect is difficult to be assessed. Nonetheless I feel some discussion in section 2.1 would be worth it.

Line 89-90: "From the precipitation series we calculated Rx5d and Rx20d, *i.e.*, the annual maxima of precipitation sum over periods of 5 and 20 days, respectively." Is the temporal connection with the flood event checked? As far as I see it, especially for "flood intensities" that are about average this may not be reflective of a true connection between precipitation and discharge.

Line 52: typo in controversy

Figure 2 and 3: "All series are smoothed with a 30-yr moving average". It looks like a backward smoothing, not centered (the data reach well into the 2000s). Maybe it should be explicated. Is the type of moving window considered when identifying the three multidecadal periods of flood variability analysed later on?

Line 143-144: "selected from the $1 \times 1^\circ$ grid such as to best represent atmospheric processes relevant for the region)" is this based on some skill metric like correlation on some target? Some more words would help here, especially if in contrast with an alternative approach such as to spatially average the reanalysis data over several gridpoints.

Line 171: check typo "for only for"

Line 214: PMIP maybe worth to be explicated

Line 232-233: Maybe this statement requires a bit more support. In my understanding, non-climatic anthropogenic influences on river runoff processes (e.g., river network changes, dams, etc.) may enhance/dampen multidecadal runoff variability or at least affect the autocorrelation of the discharge time series hence the detectability of multidecadal fluctuations above the red-noise background.

Line 253-254: for me it was somehow difficult to check this statement by comparing the figures, especially given the premise provided in the preceding statements. I see that the documentary versus observational evidence is not central, but in its current form this aspect of the study appears to be missing some elaboration, either in the text or as additional analyses (for instance, I was just thinking that some bivariate wavelet analysis may work here).

Figure S4: the red line looks more like brown?

Line 269-270: correlation of 0.21 appears rather low to me in terms of shared normalized variance, especially for smoothed/temporally aggregated time series, as I understand is the case here, which may contain a significant trend. Significance levels appear to be missing and should be provided, possibly accounting for autocorrelation of the series.

Figure 5a: how is the regional mean calculated? As I understand the calculation, as we move further back in the past, less time series contribute to the average, so this could lead to an inhomogeneity that can explain the discrepancy in the early period between time series. Possibly some illustration of standard error can reveal this uncertainty...

Figure 6b: can regions with non-significant regressions be indicated (for instance through shading)? Both map show a clear imprint of topography, which might be related as well to different variances in seasonal temperatures, significance would help to illustrate such effect for the T-FPI connection.

Line 322: cyclonic weather type centered where?

Line 352-355: this may resemble a negative Eastern Atlantic rather than a negative NAO...

Line 376: where is significance shown? Hatching is mentioned in the caption but I do not see it in the figure (rather I see red contours, that may encompass regions of significance?)

Paragraph 3.4: This is another occasion when there is mostly a qualitative presentation of the comparison between different sources. Can this be improved?

Line 385: external forcing comes out a bit out of the blue here. Maybe some further elaboration would help.