The manuscript “An increase in the spatial extent of European floods over the last 70 years” by Fang et al., presents a large-scale analysis of the spatial extent of floods and its spatial and temporal variations in the last 7 decades. The analysis is based on model simulations from the mHM model driven by observational data. The study finds that increased on average over most parts Europe and attributes its changes to changes in the magnitude or the spatial dependence of its drivers. The manuscript is well written, and the analyses and results are presented in a convincing way. Please find my comments below:

Major comments:

Comparison with HANZE: at best the detection rate of the 100 largest flood events flood events compared to the HANZE database is 50%. This low detection rate raises questions on the threshold used for the definition of flood events. The authors define flood day using the 99th percentile. This translates in more than 3 events in a year on average in each pixel. Many of the selected ‘flood days’ are therefore ‘just high-flow days’ (not even all annual maxima cause inundations). This could be a cause of this big discrepancy with the HANZE database and could suggest the use of a higher threshold for the definition of flood days (e.g., the 5- or 10-year flood).

Response:

Thank you for your suggestion. In addition to utilizing the flood-day threshold of 99% as presented in the manuscript, we also investigated thresholds of 99.7% and 99.9%. The ratios of identified flood events do not change much (the detection rate increased to 54% and 52% for 99.7% and 99.9%, respectively). It is worth noting that our analysis focuses solely on runoff extremes, without accounting for human infrastructure or other exposure. Consequently, runoff extremes may not always result in measurable impacts, thus may not be documented in impact-based flood databases such as HANZE. Moreover, employing a much higher threshold (e.g., the 5- or 10-year flood) for defining flood days would substantially reduce the sample size for studying trends in flood extremes and consequently increase uncertainty. We will therefore keep using the 99th percentile as a compromise between sample size and accuracy of matching recorded flood impact.

Figure 9c and L241 “changes in soil moisture seem to play a minor role in the detected changes in flood characteristics. This is in contradiction with the findings of Blöschl et al. (2017, 2019) and Bertola et al. (2021), Tarasova et al. (2023) who find that antecedent soil moisture is relevant to explain negative trends in flood magnitudes (and shift in timing) and increase in flood-poor periods in the Mediterranean catchments. How do the results of this analysis compare to this literature? What are the reasons of this discrepancy? Are the changes in flood extents caused by different drivers than trends in flood magnitudes and temporal clustering of floods? How is soil moisture estimated in this analysis?

Response:

We admit that this sentence lacks precision, and replacing the word "characteristics" with "extent" would be more appropriate. Additionally, we would like to clarify that when referring to the "minor" role of soil moisture, we intend to emphasize its relative importance compared to the dominant roles of snowmelt in Northern Europe and rainfall in Central and Southern Europe in influencing the change of flood extent. However, this does not imply that the effect of soil moisture is negligible.

The reason why soil moisture is not thoroughly discussed in this study is because it exhibits a more localized and secondary effect compared to rainfall and snowmelt. For instance, in Germany and France, although soil moisture decreases, flood extent increases due to more extensive heavy rainfall, highlighting the dominance of rainfall over soil moisture. However, thanks to the reviewer's suggestion, upon closer examination of the soil moisture effect, we also find that it is significant in some regions. For example, in the Northern UK, a significant increase in soil moisture contributes to the expansion of flood extent; while in Northern Iberia, the decrease in soil moisture correlates with the reduction in flood extent. These findings align with those of Blöschl et al., (2019) and Bertola et al., (2021), which similarly revealed the role of soil moisture in influencing flood magnitude in these regions. Similarly, in the Mediterranean Sea region, a decrease in flood extent is partly attributed to the decrease in soil moisture, generally aligning with the findings of Tarasova et al., (2023). Therefore, our results are not contradictory to previous studies; rather, we prioritize the investigation of dominant drivers over secondary and more localized factors such as soil moisture. However, we will also incorporate this discussion into the manuscript.

The separation of the contribution from runoff magnitude and runoff spatial dependence not only aims to disentangle the contribution from these two sources but also serves as a bridge to enhance our understanding of the drivers (e.g., rainfall, snowmelt, soil moisture, as illustrated in section 3.4) of flood extent. In addition, soil moisture is an output from the mHM model, as stated in Line 79.

L346-348: only drivers occurring in the spatial area of flood events are considered. This has big implications in the attribution analysis especially for very large rivers, as contributing drivers occurring within the actual catchment area are not considered (often drivers occurring in one part of the catchment, e.g. snowmelt or rainfall, cause flooding downstream where these drivers do not necessarily occur). What are the implications on the attribution results?

Response:

We acknowledge the reviewer's concern regarding the potential implications of flood contributing area selection on the attribution results, and we also commented on this aspect in the Discussion (Section 3.5). For instance, in Figure A7, we illustrate the total number of flood days contingent upon days when an outlet of a specific catchment is flooding. Our analysis reveals that "if an outlet grid cell experiences flooding, there is a high likelihood that grid cells within the corresponding catchment also experience flooding." (Line 350). This suggests that although we do not encompass the entire contributing catchment for each grid cell, our event-based findings remain credible.

We do note that this approach performs more effectively for smaller catchments compared to much larger ones. Nonetheless, considering the drainage area for every grid within each event would be highly challenging. Despite this limitation, our attribution results, such as the spatial distribution of snow-driven and rainfall-driven floods, exhibit strong consistency with previous studies utilizing different attribution methods (Jiang et al., 2022).

Specific comments:

L39-46: it is true that “other studies rely on observations and may miss important information due to uneven spatial distribution of stations”. On the other hand, models (like the one used here – L65-72) are calibrated and validated using observations so have this intrinsic limitation too. Furthermore, models have other limitations, e.g. modelling uncertainty and resolution of simulations (in this case gridded runoff simulations that are quite coarse – 0.125° =~ 11km …). Furthermore, for the attribution analysis the study relies on EOBS (L79-80).

Response:

We agree that also hydrological models may be affected by the uneven distribution of runoff stations. However, it is important to highlight that validating grid-based simulations against station-based observations implies not only the model's feasibility and reliability for the validated locations but also the potential feasibility of other unvalidated locations due to the physical processes encoded into the model. This aspect can provide valuable insights into ungauged areas and improve the quantification of flood extent.

Additionally, despite the model's resolution limitation (e.g., 0.125° in our study), it still allows for a more refined estimation of flood extent compared to station-based research. For example, the flood synchrony scale, commonly used for quantifying flood extent from a station-based perspective, is defined as the largest radius of the circle within which half of the station flooding occurs near simultaneously (Berghuijs et al., 2019). However, its application in regions with scarce runoff stations may introduce considerable bias due to the distance between stations. While the grid-based simulation could help alleviate this limitation. Note that we also acknowledge other limitations of the model simulation besides resolution, as illustrated in Line 326-334.

L21 "spatially compounding river floods": the use of "compounding" seems inappropriate in this context as we are talking about spatially widespread events and there is no mention of simultaneous occurrence of flood drivers in this section (or elsewhere in the paper). I suggest substituting "compounding" with "widespread" or similar.

Response:

Thanks for your suggestion. We will change it.

L52: I suggest citing Lun et al. (2020) – the first study detecting flood-rich and flood-poor periods in Europe.

Response:

Thanks for your suggestion. We will add it in the manuscript.

L100-106: are flood days defined at each grid cell separately? Spatially connected flood days (i.e. pixels) and overlapping flood patches are 'further combined'. Does this mean that they are considered as one single event? Please clarify.

Response:

Yes, the flood days are defined at each grid cell separately and then the overlapping flood patches are further combined together to form a single event. We will clarify this part in the revised manuscript.

L125: E_past and E_pres denote the AVERAGE flood extent?

Response:

Yes, E_past and E_pres denote the temporally-averaged flood extent over the historical period (1951-1980) and present period (1991–2020) for each grid cell, respectively.

L158: 244 flash floods from the HANZE dataset are captured in the dataset. However, flash floods typically occur over small catchments (a few km2) while the catchment area that is captured is at least of the order of magnitude of 1000 km2 (due to the resolution of the gridded simulations). Similarly for the time dimension, i.e. flash floods typically last less than 24h, while the resolution of the simulations is daily. How can mHM model simulations capture such flash floods? What are the implications in terms of such identified events?

Response:

Yes, we acknowledge that flash floods often have relatively small spatial extents. However, note that in our comparison, we assess the identified flood extent against the affected NUTS3 regions documented in the HANZE database, rather than pinpointing the exact spatial extent of specific flash floods. This approach may potentially increase the detection ratio of flash floods.

Regarding temporal resolution, as outlined in the method section, our detection method employs a moving window of ±3 days. Additionally, the temporal resolution of the impact-based flood dataset (HANZE) is also daily, which helps in capturing short-duration flash floods. The implication of identifying such events is to provide confidence to the applicability of mHM simulations and the flood detection algorithm for studying changes in large-scale floods.

Figure 2: is this figure only representing the events in the HANZE dataset or does it contain results of this analysis? If it does not contain results of this analysis, it should be moved to another section (e.g. appendix).

Response:

Yes, thanks for your suggestion. The figure only represents the events in the HANZE dataset, and we will put it in the appendix section.

Figure 9: it is not clear if the maps show changes in snowmelt, rainfall and soil moisture over all days in the two periods or if they refer to changes for flood events only (i.e. only the rainfall causing flood events or changes in rainfall in general?)

Response:

Sorry for the confusion. The spatial maps in figure 9 show changes in snowmelt, rainfall and soil moisture over all days in the two periods.

L329: “aligns closely with an independent impact-based

Response:

Thanks, we will correct this in the revised manuscript.

Figure A1: labels and titles of the plots are not fully clear and not explained in the caption.

Response:

Thanks, we will revise this in the revised manuscript.

Reference

Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing Spatial Scales of Synchronous River Flooding in Europe. *Geophysical Research Letters*, *46*(3), 1423–1428. https://doi.org/10.1029/2018GL081883

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., & Blöschl, G. (2021). Do small and large floods have the same drivers of change? A regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, *25*(3), 1347–1364. https://doi.org/10.5194/hess-25-1347-2021

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., … Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*(7772), Article 7772. https://doi.org/10.1038/s41586-019-1495-6

Jiang, S., Bevacqua, E., & Zscheischler, J. (2022). River flooding mechanisms and their changes in Europe revealed by explainable machine learning. *Hydrology and Earth System Sciences*, *26*(24), 6339–6359. https://doi.org/10.5194/hess-26-6339-2022

Tarasova, L., Lun, D., Merz, R., Blöschl, G., Basso, S., Bertola, M., Miniussi, A., Rakovec, O., Samaniego, L., Thober, S., & Kumar, R. (2023). Shifts in flood generation processes exacerbate regional flood anomalies in Europe. *Communications Earth & Environment*, *4*(1), Article 1. https://doi.org/10.1038/s43247-023-00714-8