We would like to thank all reviewers for their constructive comments on our manuscript. A point-by-point response to all comments raised follows below. Throughout the response, the reviewers' comments are presented in black, and our responses are in blue. The line and figure numbers in red correspond to those in the clean version of the revised manuscript.

Response to Reviewer 2

I thank the authors for their reply. However, I am not fully satisfied by their answer. Please find my comments below:

1) Comparison with HANZE: based on the answer of the authors I argue that the choice of comparing the runoff extremes with HANZE events is reasonable/meaningful. This is because in the HANZE database only floods that had big impacts are reported and this reasonably correspond to a flood that is at least a flood associated with a 10-year return period which is much larger than your detection threshold (even the threshold 99.9% of daily runoff is not really a 'big flood'). From my point of view you are comparing two very different things and you should use a different observational database to 'validate' you event selection. Also, I do not agree with what stated at L358-360 ("Furthermore, the resulting flood event database aligns closely with an independent impact-based flood record database […]"). If this sentence refers to the HANZE (which I am not 100% sure) I would not define a 50% detection rate as "close alignment".

Response: We acknowledge the reviewer's concern regarding the differences between the two flood datasets. For a physically based evaluation of runoff, we provide Figure 2. In Figure 3, we assess whether the identified top events via the algorithm are impactful. We agree that flood events identified in this study using extreme runoff (i.e., 99th percentile) do not necessarily correspond to impactful floods as those documented in the HANZE database, not least because we might detect floods in areas with little exposure. We thus concur that the HANZE dataset may not be suitable for formal validation of our dataset due to significant differences in their severity levels.

However, we argue that impactful flood events recorded in the HANZE database should, in theory, be detectable by a robust flood detection algorithm, also when a relatively moderate detection threshold is used. Our results indicate that approximately 75% of river floods documented in HANZE are detected by our algorithm, thereby affirming that our approach captures impactful floods (besides many other floods). Moreover, we selected the HANZE dataset for comparison because it includes detailed information on the regions affected by floods, which facilitates a more direct comparison with the flood extent areas identified by our 3D detection algorithm.

In summary, rather than using the HANZE dataset for a formal validation of our database we use it to assess whether we are able to capture societally-relevant, impactful events through our flood detection algorithm. To avoid confusion, we now added to the manuscript "*Note that the HANZE dataset is not used for formal validation of our flood events due to the large differences in the severity levels of floods between these two datasets. Instead, it serves more as a comparison to assess our flood detection algorithm's adequacy, particularly in capturing impactful floods, even with a relatively moderate detection threshold*" in Lines 99-102.

Regarding L358-360 this statement primarily refers to our findings that our algorithm detects about 75% of the river floods recorded in HANZE. We acknowledge the need for clarity in this aspect and revised this point in the new manuscript by saying "*the resulting flood event database aligns closely with an independent impact-based flood record database by capturing 75% of recorded impactful river floods, ...*" in Lines 351-352.

2) Soil moisture effect: I do understand the general reasoning. However, why do the authors state in their reply that the effect of soil moisture is 'more localized'? I could not deduce it from the results/figures presented. By reasoning I would rather assume the opposite (i.e. soil moisture is usually changing gradually and more smoothly in space than e.g. extreme precipitation that can be very localized). Please justify your statement.

Response: We apologize for the confusion. In our previous response, the term "more localized" was used to emphasize an aspect of our findings, that is that the significance of soil moisture varies by region, while precipitation consistently plays a more dominant role in inducing floods across central and southern Europe.

3) only drivers occurring in the spatial area of flood events are considered. I accept the answers of the authors, but I think this is an important limitation that is not really clearly stated in the manuscript. L380-381 reads a bit vague and not accurate. E.g. what is "nearby precipitation"? It would be more correct to say that the precipitation and snow "falling within the corresponding hydrological catchment is not considered". Also, "spatial area of the flood events" is a bit vague. "Same pixel" would read more precise.

Response Thank you for your suggestion. We have clarified our statement in the revised manuscript by specifying that "*the precipitation and snowmelt falling outside the event but within the corresponding hydrological catchment are not considered*" in Lines 374-375. Additionally, we have modified the sentence to "*within the spatial area, that is, the same pixels of the flood events*" for greater precision in Lines 373-374.

Response to Reviewer 3

Overall, my previously raised points and concerns have been appropriately addressed in the revised version of the manuscript and it has improved a lot in clarity.

Some points that I had not noticed before should still be improved to improve clarify and further strengthen the document before I can recommend publication.

L 86: Please specify how snowmelt and soil moisture were derived from mHM. I.e. what quantities are used, what qualifies as such process. E.g. temperature thresholds used, how is it being checked if there is actual snow accumulation before the melting event, etc.

As this will be important for the interpretation of the rest of the study.

This will probably also help clarify the following point

Response: The snowmelt derived from mHM is based on the day-degree method, with a model-specific temperature threshold parameter (*snowTreshholdTemperature* = 1.32°C) used in the presented European setup. At an hourly model's internal time step, depending on this parameter, either snow accumulation or snowmelt occurs. Please refer to the model's code for the snow accumulations under this link:

https://git.ufz.de/mhm/mhm/-/blob/develop/src/mHM/mo_snow_accum_melt.f90?ref_type=heads#L133

As you can see from the model's equations, snow accumulation can happen before snowmelt, but we don't distinguish between these two types. As you can further see, the rate of snowmelt is driven by another parameter *degreeDayFactor*. We also further note that *degreeDayFactor* is separately defined for the three major land use classes (i.e., forest, impervious and pervious). The model's equation of mHM are described in earlier studies (Samaniego, et al. 2010; doi: 10.1029/2008WR007327 and Telteu et al. 2021; doi: 10.5194/gmd-14-3843-2021) and model's open source code is available at git repository (see https://git.ufz.de/mhm/mhm), where details on the soil moisture can be obtained. In short, soil moisture is represented by 6 layers/buckets, corresponding to the SoilGrids profiles (i.e. 5cm, 15cm, 30cm. 50cm, 100cm, 200cm), which is being filled by effective precipitation (rainfall and/or snowmelt), from where upward evapotranspiration flux happens depending on the degree of saturation of the soil and evaporative demand and consequently downward infiltration flux to unsaturated zone is estimated, which

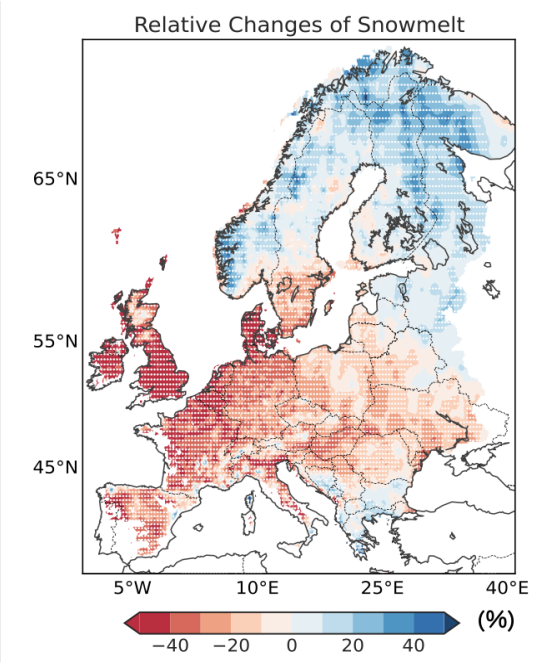
further drains the soil storage. More details are provided in Samaniego, et al. 2010; doi: 10.1029/2008WR007327 and Telteu et al. 2021; doi: 10.5194/gmd-14-3843-2021.

In our previous manuscript, we estimated snowmelt by calculating the difference in snowpack between two consecutive days, as the earlier version of the mHM model did not provide snowmelt data directly. However, in the latest version of mHM, snowmelt is now a direct output. We have updated all relevant results using this new snowmelt output, as shown in Figure 5, Figure 6 and Figure 9a. As anticipated, the updated results are highly consistent with the previous ones, demonstrating improved accuracy.

L 271: I find Figure 9 a confusing. However, based on the outcome of this map I´m not sure if the method used is appropriate… Please carefully check the results…

a) The map is showing significant reductions of snowmelt in areas where there is basically never snow… e.g. in southern Europe

Response: We appreciate the reviewer's observation regarding Figure 9a. In the current depiction, the map includes all grid cells, regardless of their frequency of snowfall. If we adjust the criteria to only include areas with over 30 days of snowmelt larger than 2 mm in both the earlier (1951-1980) and present (1991-2000) periods, the resulting map indeed omits some regions of southern Europe, predominantly retaining the mountainous areas (as shown in the figure below). We therefore have updated the manuscript using this figure.



b) around the Alqueva dam, the largest artificial lake in Europe, the map is showing an increase…

In that region the winter temperatures are so high that there is practically no snow accumulation and with the lake the temperatures are even higher…

This is further raising the question of how/on what basis an increase/decrease in snowmelt being calculated.

Response: Similar to the previous response, the observed increase in snowmelt at the Alqueva dam is statistically insignificant. As illustrated in the figure above, if we apply a criterion of over 30 days of snowmelt larger than 2 mm, this region does not meet the threshold and is therefore excluded from our analysis. We have included this revised figure in the manuscript to help clarify and resolve the confusion.

Maybe, to help with the interpretation, apart of a detailed explanation of how snowmelt and soil moisture were obtained/derived, it would be good to have on Fig 4 some sort of delineation that would indicate in which areas there is more than a zero fraction of snowmelt/rainfall driven floods as shown in Figure 5.

Response: We think that such a delineation would create more confusion than clarity as it would overload the figures and double some of the information in Figure 5. We hope the explanations above and the updated figures are clear enough on this matter.