Authors' responses to the comments of Reviewer #2

We appreciate your review and comments on our manuscript, "**The Pluvial Flood Index (PFI): a new instrument for evaluating flash flood hazards and facilitating real-time warning**". Your feedback is valuable to us, and we will make the recommended revisions accordingly. We provide detailed responses to each of your comments below.

This article introduces a novel approach to assessing and mapping pluvial (flash) flood risk, referred to as the Pluvial Flood Index (PFI). The proposed method is demonstrated through two case studies: first, a hindcast of a flash flood that occurred on June 2, 2024, in the Wieslauf catchment in Germany; and second, the generation of flood hazard maps for a region of approximately 1,000 km² in the southwest of Baden-Württemberg. In the latter case, the reliability of the PFI maps is evaluated using a database of flash flood incidents that caused damage to infrastructure and roads. This database, originally compiled by the state authorities of Baden-Württemberg and extended by the authors, covers events from 1995 to 2024.

The proposed approach, along with its underlying motivation and evaluation, may initially appear appealing; however, upon closer examination, it raises several concerns.

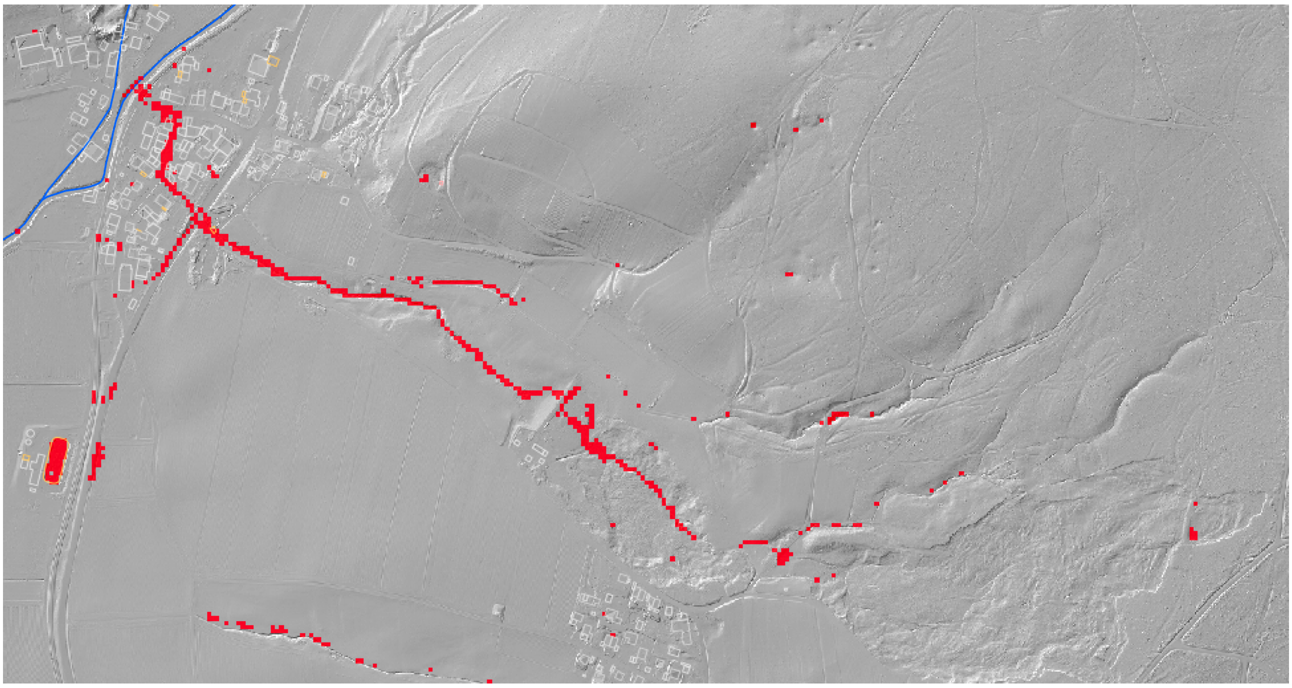
First, the method does not actually generate new information but rather involves a post-processing of standard runoff maps produced using a combination of a rainfall-runoff model and a 2D hydraulic model. This post-processing is carried out in two steps. In the first step, a Boolean variable called PFHA (Pluvial Flood Hazard Area) is assigned to each pixel based on predefined flood depth and velocity thresholds, categorizing them as either low or high hazard. In the second step, a spatial smoothing technique is applied to calculate, for each pixel, a Pluvial Flood Index (PFI), which reflects the proportion of high-hazard pixels within a circular buffer surrounding that pixel. Four proportion ranges are defined, corresponding to four levels of PFI. However, this spatial smoothing—central to the PFI concept—is not adequately justified, and the choice of the circular buffer size (2 km²) remains unexplained.

It is understandable that the initial motivation behind the development of the PFI method was to enhance the readability of flood risk maps at a large spatial scale. In the two case studies presented, the "Pluvial Flood Hazard Areas" (PFHA) are mostly confined to riverbeds and broader floodplains, making them difficult to discern on large-scale maps. To address this, the proposed PFI method uses spatial smoothing to create broader, more visually prominent zones around clusters of high-hazard PFHA pixels. While this improves map legibility, it is crucial to recognize that spatial smoothing is a digital artifice. The resulting PFI levels are not solely the product of hydrological and hydraulic modeling—they are shaped by additional processing choices and assumptions, and should therefore not be overinterpreted.

Unfortunately, the authors appear to let their enthusiasm override critical assessment. One must ask: is there any sound reason to believe that hillslopes located in the surroundings of inundated floodplains are more exposed to pluvial flooding than hillslopes located elsewhere? A careless reading of the PFI map might suggest so, but such a conclusion lacks physical basis. Proximity to a floodplain does not inherently increase a hillslope's vulnerability to pluvial flooding.

In short, the PFI method should be presented for what it truly is: a straightforward graphical tool designed to highlight clusters of high flood hazard at the regional scale, not a physically grounded index of flood exposure.

Thank you for this critical assessment of the general PFHA and PFI methodology. We agree that the PFHA and PFI could be considered as a simple post processing of the output of a hydrological and 2D-hydraulic model chain and could be applied to any rasterized, large-scale data of maximum water depth, maximum flow velocity and maximum specific discharge – if these variables are available.
First, we will improve to show the location of the PFHA. As we did not include the location of rivers in Fig 5a, it looks like that PFHA are only confided to riverbeds and broader floodplains. But this is not the case for many regions. I have included an example below, and we will provide a similar map and information in a revised paper to showcase the different locations of PFHA



One point we will make clearer in the revised version of the manuscript is the spatial scale the PFI aims at. As we stated in 177ff, the PFI is intended for a broad assessment of pluvial floods at a regional (e.g. state-level ~ 200x200km) scale. It should allow decision makers to get a quick information if a certain community or region is generally prone to pluvial flooding (in the case of flood hazard maps) or if a certain rainfall event has the potential to trigger a hazardous flood event in a certain community/region (in the case of forecasting). Hence the PFI is not designed for distinguishing the flood vulnerability of single hillslopes or local features or buildings. This information shall be taken from the detailed PFHA maps or already existing local pluvial flood maps, but the PFI can provide the information which pluvial flood map is the most relevant in the current situation.
We will restructure section 2 in a way that the general purpose of the PFHA and the PFI and the link to already existing pluvial flood maps becomes more clear. We will also discuss the pro and cons of the different methods for spatial aggregation (e.g. could also be on political/organizational units) more in detail and detail more why we think that the chosen approach (with radial weighting) has some benefits compared to the others.
Regarding the underlying hydrological and hydraulic processes, we still think that this is relevant to be included in the manuscript, since a meaningful PFHA and related PFI (for both, operational forecasting and identifying general pluvial flood susceptibility) requires the sound modelling of hydrological and hydraulic process. As already discussed, these processes are often not represented in current developments (e.g. 'Hinweiskarte Starkregen') implemented at state or even federal level. To make it clearer that this is not a direct part of the PFI but a prerequisite, we would suggest to move the chapter 2.1 outside the definition of the PFI.

Second, the evaluation procedure is insufficiently described and critically discussed. In Figure 6, past damaging flood events are depicted as dots. However, flood events have spatial extents—especially those occurring in catchments with upstream drainage areas of 10 to 20 km². How was the representative location of each flood event determined for the evaluation? It appears likely that the dots indicate the locations where the most significant damages occurred, typically along riverbeds and, in many cases, within flood-prone plains. If this is the case, then the apparent skill of the flood risk assessment may not stem from the PFI method itself, but rather from the underlying rainfall-runoff and hydraulic simulation models that generate the base maps.

We agree that it would be beneficial to have better damage data to verify the PFI. However, there is hardly any information available in the case of pluvial flood events at the scale of interest. For a few villages, we have some more explicit spatial information, but this is typically related only to the location of emergency calls during the flooding. The database which is available for the state of BW and included in Fig. 6 is more or less a binary data set, indicating if in a certain community there was a damage following a heavy rainfall event or not. Hence the dots simply represent communities where there was a damage. One other possibility could be to mark the whole area of the village, but depending on these boundaries, this may also be misleading. But we will make this clearer in the text.

Since the quality of the PFHA/PFI depends on the quality of the models used to estimate the parameters (maximum water depth, maximum flow velocity and maximum specific discharge) to define hazard areas, it is simply impossible to disentangle the quality of the PFI from the quality of the underlying model data. We will also make this clearer in the revised version.

For the Wieslauf example, we collected some data on maximum inundation – but of course only for a few locations and a few days after the catastrophic event. Although this data does not at all allow the identify all spots affected by the pluvial flood, it could be included to verify that PFHA/PFI structures somehow are reflecting inundated areas.

Moreover, although the article is ostensibly focused on pluvial flooding, the evaluation is confined to areas with catchment sizes larger than 10 to 20 km², which more closely align with riverine or flash flood events rather than true pluvial floods. This choice contradicts the stated objective of the manuscript. What proportion of the DWD-CatRaRe database was excluded due to this threshold? It would be helpful to visualize all flood events from the database in Figure 6, using a distinct color to differentiate those included in the evaluation. Additionally, the relationship between the recorded heavy precipitation events (Figure 6a) and the damaging flood events (Figure 6b) is unclear—some commentary on their correspondence is needed.

It seems that this is a misunderstanding which probably arises from the fact that we did not describe the accumulation threshold and its motivation in sufficient detail. The PFI aims at large scale, however, in order to reflect the spatial scale of convective precipitation events and the typical catchment areas related to pluvial floods, we limit the accumulation area of generated flows to 10km² and 20km², respectively.

So we do not consider any catchment areas larger than 10 to 20km², depending on the chosen threshold. Hence we also did not exclude any convective, short and small scale events from CatRaRe, but the opposite – we excluded all events with a duration of more than 4h, which can be assumed to be rather of frontal nature. We also did not exclude events from the ‘damage’ database, but used all data that was reported – definitely knowing that it is incomplete.

The comparison between heavy precipitation events (Figure 6a) and the damaging flood events (Figure 6b) was included to show that there were many more heavy rain events (actually more or less all over the study region as can be expected) compared to the damage events (which of course are somehow linked to the valleys or hillslopes – however not to all valleys) – so basically this shows the added value of a PFI forecast with respect to a solely precipitation forecast. We will reformulate the section so that this becomes clearer.

Further concerns arise in Section 2.1, "Relevant Processes." The section implies that pluvial runoff and flash floods are primarily caused by direct overland flow due to either saturation or infiltration excess. This is an outdated and overly simplistic view that has long been challenged in hydrological literature. Decades of research into hillslope processes and flash flood generation have demonstrated the complexity and spatial-temporal variability of runoff mechanisms, including the often-dominant role of subsurface flows (interflows). Such simplification is not appropriate for publication in an international journal and should be reconsidered. Since detailed discussion of rainfall-runoff processes is not central to the manuscript's main contribution, it may be best to remove or significantly revise this section.

We agree that interflow can play an important role but mainly for events with longer duration. For short events under typically dries soil moisture conditions, interflow is quite unlikely in our regions, which we can show with a larger body of observational data. Therefore, we wanted to make the case of the relevance of saturation or infiltration excess overland flow for the generation of pluvial floods. Numerous sprinkling experiments revealed the importance of excess infiltration in the case of short-term high-intensity heavy rainfall events representing the main cause for pluvial flooding (e.g. Ries, F., Kirn, L., & Weiler, M. (2020). Runoff reaction from extreme rainfall events on natural hillslopes: a data set from 132 large-scale sprinkling experiments in south-western Germany. *Earth System Science Data*, *12*(1), 245-255).

As mentioned previously we think that the correct representation of the dominant runoff generation processes is key for a meaningful estimation of the PFI. Hence, we suggest to leave the section in, streamline it, but also restructure the method part of the paper.

In conclusion, the article presents a potentially useful approach. However, the PFI method should not be overstated. It should be honestly described for what it is: a simple graphical tool aimed at highlighting clusters of flood hazard on regional maps. If the primary focus is on pluvial flooding, then there is no justification for excluding events in small upstream catchments—especially when past records of such events exist. While including these areas may lead to less favorable evaluation outcomes, such results would be highly relevant and valuable for the community working on pluvial flood risk mapping. If the current threshold is retained, then the manuscript should clearly state that its focus lies with riverine or flash floods rather than pluvial floods.

As mentioned above, we will restructure the manuscript to make it clearer what the PFI method is based on and for what it designed, but also what the prerequisites are to have a sound estimate of pluvial floods. We will reformulate section 3.2 so it becomes more clear that we did not exclude any small upstream catchments (the focus was on explicitly including them!) from the analysis and focus more on the need of defining accumulation thresholds.

Finally, the discussion and analysis need to be further developed. It appears that the PFI and PFHA indices primarily identify flood-prone river plains. If this is indeed the case, it should be explicitly acknowledged and critically considered.

We disagree that PFHA are primarily linked to river plains (see response above), but some valley (with and without rivers!) are certainly included, which of course makes sense, since water usually accumulates not at hill tops. However, if one examines the spatial PFI patterns in Fig 6 closely, one can see that not all river-plains are characterized per see as high hazard areas but that regional differences due to runoff generation processes as well as the 2d hydrodynamic flow accumulation are present. If a simple topographical index would be used, the regional differences would be much less pronounced than in the case of the PFI. The same would be true, if we would not consider the runoff generation processes but simply use rainfall estimates as input for the 2d-hydrodynamic flow accumulation – which as already discussed, are still represented in many current developments (e.g. 'Hinweiskarte Starkregen'). We will reformulate the text to make it clearer.