Answer to comments of anonymous referee 2

This manuscript compares simulated flood occurrence and damage under different flood-protection assumptions with satellite observations and reported impacts across Europe. Among the tested protection scenarios, uncertainty is primarily associated with flood protection, and model agreement improves in larger catchments. My main concerns involve the temporal inconsistency among input datasets, the rationale for the continental scale, and several claims in the Conclusions.

We thank the reviewer for the interesting points of discussion that were identified, and also for the raised minor comments, that we will integrate in the revised manuscript.

Major comments:

1. The input datasets refer to different years or periods. These temporal mismatches introduce uncertainty into the exposure and damage estimates, but the manuscript does not assess or discuss their effects. The authors should explain how differences in temporal coverage and reference years affect the comparison. In addition, the flood analysis covers 1992–2024, whereas the authors use the mean protection value for 2015–2020. Flood-protection standards and infrastructure may have changed during the study period. State whether the analysis accounts for these changes. If it uses a fixed protection level, justify this choice and discuss how it may affect estimates for earlier events. A summary table would make these differences easier to evaluate by reporting each dataset's source, reference year or period, spatial resolution, and other relevant information.

We thank the reviewer for raising this point. Discussing the effect of changing the reference period of each single dataset is indeed not a trivial process. Despite the aim of the current manuscript not being to provide an accurate assessment of flood damages in Europe in the last decades but rather comparing different ways of assessing flood damages and datasets (Lines 81 to 84), the discussion can indeed benefit from a better discussion of the effects of the reference period choices for the different datasets.

Currently the rationale behind the choice of the fixed mean protection value for 2015-2020 is briefly presented only in Appendix 1 (lines 683 to 692) and it consists primarily in the fact that this choice allows a comparison with the fixed protections of the FLOPROS dataset (which was produced in 2016 with the most up-to-date data at the time). In addition, the methodology employed for the estimation of the protection return periods in the original work by Paprotny et al. (2025) consists in first estimating the flood vulnerability of each NUTS region and year based on reported data (from Hanze) and various predictors and then retrieving the protection return periods that allow to obtain a similar vulnerability. As a consequence of this methodology, for most regions, flood protections have been estimated to have decreased between 1992 and 2020 (figure A5), which besides being physically inconsistent, it just implies that flood vulnerability has increased during the same period.

In any case, using the yearly protection values would have allowed to obtain a slightly lower number of floods and total damages compared to those presented in section 3.1, but would not have changed the interpretation of the results, since with the current setting reported damages are exceeded by 8 times.

In the revised manuscript we will expand this discussion and report it also in the main text of the article. We will also add a summary table reporting the main information regarding the original data in the cases of the population, land cover, protection, simulation and remote sensed datasets as suggested.

2. The manuscript presents continental coverage as a key contribution but does not establish why this scale is necessary for the research question. The authors should explain what the European-scale analysis reveals that studies of selected countries or river basins could not show. National differences in flood protection, exposure, reporting practices, and data availability may affect the comparisons. Discuss how these differences influence the results and whether they limit comparisons across countries. The continental scope also creates an

opportunity to discuss transboundary flood-risk management, including implications for communities located near national borders.

We thank the reviewer for allowing us to better clarify the raised points. In fact, we did not intend to imply that the continental scale is necessary for the research questions. In the introduction (L.73), we just state that this is the first assessment at the European scale, analysing and comparing reported, simulated and "remotely sensed" flood impacts.

In addition, the continental scale makes sense and is necessary with our approach as both EFAS and GFM are services provided coherently at European scale and hence allow comparisons at this large scale. Adopting a continental scale also allows to obtain datasets large enough to extract robust statistics. More local comparison studies using these same data would in fact not benefit from their main strengths, furthermore the resolution of more targeted and local hydrological models and drone/airborne remote sensing would far exceed that of the systems here analysed, allowing for much finer comparisons and identifying issues of a nature different from that we considered here. We will add these points in the introduction.

3. The flood-hazard maps cover return periods up to 500 years. The Methods do not explain how the authors handle simulated discharges with return periods above this threshold.

We thank the reviewer for identifying this missing piece of information. Damages and water depths are capped at the values calculated using the 500 years RP maps, discharges with a RP higher than this were assumed to cause the same damages and water depths.

At lines 132-133 we briefly state that the 500-year return period flood hazard map is used to represent the potential flood extent across Europe but indeed we did not say that it is also used to cap damages and water depths from the simulations.

We will add this point in the revised manuscript.

4. At the end of the Introduction, the authors state that the study will "provide recommendations for improving large-scale flood risk assessments." The Discussion and Conclusions do not present a defined set of recommendations. List the recommendations and connect each one to evidence from the analysis. Relevant categories may include flood-protection data, satellite observations, event-matching methods, hazard maps, and reported-impact databases. The authors should also explain how practitioners could apply these recommendations in transboundary assessments. Moreover, the Conclusions mention AI as a potential way to address the limitations identified in the study. The claim remains too broad because the manuscript does not specify the tasks, data requirements, or validation strategy. Identify which parts of the workflow could benefit from AI, what training and validation data those applications would require, and which uncertainties AI would leave unresolved. Otherwise, remove the statement or narrow its scope.

At lines 85-92 we state that, as one contribution among many, we provide recommendations for improving large-scale flood risk assessment. We thank the reviewer for this comment, as the statement might be better formulated as "provide recommendations for improving large-scale flood risk assessments and their interpretation", rather than citing the allowance of "more robust interpretation of flood risk estimates" only in the following sentence, as this might better convey the contribution.
Concerning the set of recommendations, we present it at lines 648-656, without explicitly connecting it to the contribution in the introduction, which we will correct in the revised manuscript.

In addition we may add a sentence suggesting the development of more specific flood protection datasets, reporting not only the protection return period, but the actual characteristics of the flood protection, as works exploring this concept have recently started to appear.

Concerning the discussion on transboundary flood risk management and assessments, in order to remain within our intended scope for the manuscript, we will mention that the continental scale nature of the datasets used in the work naturally allows to coherently assess risks in a spatially consistent way also in transboundary areas.

Finally, concerning the expansion of the paragraph mentioning AI as a step forward, we will expand this in the revised manuscript following also the suggestions of the anonymous referee #1.

Minor comments:

1. Line 11: Replace “observations” with “satellite observations” to distinguish these data from reported flood-impact records.
2. Lines 33–34: “Thanks to” sounds informal in this context. Consider “due to” or “because of.”
3. Line 39: Define the “simplifying assumptions and incomplete information.” One or two examples would make the research motivation more specific.
4. Lines 48–49 and line 157: The references appear to use inconsistent formatting. Check the citation format throughout the manuscript.
5. Line 96: Capitalize the figure reference as “Figure 1.”
6. Figure 1: Expand the caption to describe the main datasets, processing components, and outputs. Readers should be able to understand the workflow from the figure and caption.
7. Lines 114–116: Define NUTS level 3.
8. Line 120: Add a web address or citation for the phrase “documentation available online.”
9. Lines 121–122: Insert spaces between values and units, including “km” and “km².” Check unit formatting throughout the manuscript.
10. Lines 155–165: Explain why the population and land-cover datasets represent different years. Discuss how this mismatch may affect the exposure and damage estimates.
11. Lines 166–167: Specify the spatial resolutions of NUTS 2 and NUTS 3.
12. Line 214: Specify the reference point for the radius.
13. Line 216: Define convex hull.
14. Figure 2: The labels describe processing stages rather than time steps. Replace “time steps” with “processing steps” or “workflow stages” to avoid confusion with the temporal resolution of the flood data.
15. Lines 237–239: The sentence combines several points. Divide it into shorter sentences.
16. Line 377: Introduce the No Protections scenario and the rationale for comparing it with the two protection datasets in the Methods before presenting the results.
17. Figures 6 and 14: Increase the font size of the axis labels and annotations. Check all other figures as well.

We will correct all the minor comments that were highlighted by the reviewer.