Answer to comments from anonymous referee 1

The manuscript presents a comprehensive assessment of the uncertainties associated with continental-scale flood risk estimation in Europe. By comparing EFAS-LISFLOOD simulations under three flood protection scenarios (FLOPROS, P25, and No Protection) with satellite-derived observations (Copernicus GFM) and historical impact records (Hanze database), the authors examine the characteristics and limitations of each data source.

The analysis shows that flood protection assumptions are a major source of uncertainty, substantially influencing estimated losses and the relative risk ranking of European countries. The manuscript also examines the limitations of different reference datasets, including the tendency of the model to overestimate inundation due to static hazard map aggregation and the tendency of satellite observations to underestimate flooding in urban areas because of SAR-related limitations.

The manuscript is well organized, and the conclusions are generally supported by the presented analyses. I recommend this manuscript for Moderate/Minor Revisions. The following comments are intended to further clarify the discussion and strengthen the manuscript.

We thank the reviewer for the valuable comments and the positive evaluation of our work.

Spatial mismatch in flood protection (limitations of FLOPROS): While the manuscript demonstrates the influence of using FLOPROS versus P25, the discussion could further address the implications of applying polygon-based protection standards at the NUTS3 level to localized hydrological processes. Applying a uniform protection standard across an entire NUTS3 region may overrepresent protection in rural and agricultural areas where such infrastructure is limited. Discussing how this spatial mismatch contributes to differences in agricultural damage estimates between the model and GFM would improve the interpretation of the results.

We thank the reviewer for raising this very interesting point, that we agree is something worth discussing. The current approach most likely overestimates flood protections in agricultural areas. However, this approach makes sense and is in our opinion the most reasonable given how the protection levels of P25 and FLOPROS were obtained. Until more refined protection datasets identifying the precise location and geometric characteristics of flood protections become available, it would not be possible to apply different protection standards to urban and agricultural areas without any strong assumptions. For example with the current datasets some critical cases nay include towns located in the vicinity of rivers but with an agricultural or rural area in between the urban cluster and the river; in this case, if the hypothesis of a uniform protection at the NUTS3 level is discarded, it is in fact not clear which protection standard should be used or why the rural and agricultural areas should have two different protection levels if only river levees are accounted for. Another potential problem might be represented by the fact that flooding originating in agricultural areas at the outskirts of urban centers might expand to the adjacent urban centers themselves, even if these are theoretically protected. In the revised manuscript, we propose to add a few sentences in section 3.1 to address these points.

Addressing the low return-period gap: The manuscript identifies that the absence of hazard maps for return periods between 2.5 and 10 years accounts for nearly 20% of the unmatched events. Because EFAS is an operational system, the authors may consider emphasizing this finding in the Abstract or Conclusions and suggesting the development of hazard maps for return periods below 10 years in future Copernicus/EFAS releases.

We thank the reviewer for highlighting this finding. We agree that it is indeed an important point with operational implications that will be further emphasized in the conclusions of the revised version of the manuscript.

Satellite limitations in urban areas: The manuscript attributes the under-detection of flooding in residential and industrial areas by GFM to SAR backscatter limitations and the 0.1 m water-depth threshold. The discussion could briefly mention potential directions for improving urban flood mapping, such as integrating multiple remote sensing data sources (e.g., SAR and optical imagery) or data assimilation and AI-based approaches.

We thank the reviewer for this valuable suggestion. In the revised manuscript, we will extend the discussion to outline potential directions for improving flood detection in urban areas.

Our finding that GFM under-detects flooding in residential and industrial areas indicates that flood-mapping approaches relying primarily on Sentinel-1 SAR observations could benefit from complementary data sources and methods, particularly where complex backscattering from buildings or shallow inundation reduces the visibility of floodwater.

We will refer to several relevant studies illustrating these opportunities. For example, Munasinghe et al. (2023) evaluated the combined use of Sentinel-1, Sentinel-2, and Landsat observations at the global scale and showed that multisensor integration can substantially increase the probability of obtaining a usable observation during a flood, thereby reducing temporal gaps associated with individual satellite missions. Fawakherji and Hashemi-Beni (2025) combined Sentinel-1 SAR data with very-high-resolution UAV optical imagery using machine-learning classifiers, demonstrating how information from sensors with different spatial resolutions and imaging characteristics can improve the discrimination of open water, inundated vegetation, and other land-cover classes in complex environments. Finally, Liu et al. (2026) integrated Sentinel-1 SAR, Sentinel-2 optical imagery, high-resolution optical observations, and terrain information within a machine-learning framework. Their results illustrate the complementarity between the all-weather monitoring capability of SAR and the finer spectral and spatial delineation provided by optical imagery, particularly in heterogeneous environments containing vegetation and built-up surfaces.

Together, these studies suggest that the under-detection identified in our analysis could be reduced through multi-sensor approaches that combine SAR with optical or very-high-resolution imagery and use machine learning to exploit their complementary information. We will also clarify that these approaches would not eliminate all sources of uncertainty: optical observations remain affected by clouds and acquisition timing, very-high-resolution imagery may not be consistently available at continental scale, and data fusion alone would not resolve the limitations in detecting inundation below the 0.1 m water-depth threshold adopted by GFM.

References

Fawakherji, M., & Hashemi-Beni, L. (2025). Flood detection and mapping through multi-resolution sensor fusion: integrating UAV optical imagery and satellite SAR data. Geomatics, Natural Hazards and Risk, 16(1). https://doi.org/10.1080/19475705.2025.2493225

Liu Y, Wang X, Kong X, Ye Y, Xu X, Zhao B, Zhu W and Wen Q (2026). Machine learning–based flood inundation mapping using fused optical and SAR remote sensing: a case study of the Miyun flood. Front. Earth Sci. 14:1805240. doi: 10.3389/feart.2026.1805240

Munasinghe, D., Frasson, R.P.d.M., David, C.H. et al. (2023). A multi-sensor approach for increased measurements of floods and their societal impacts from space. Commun Earth Environ **4**, 462. https://doi.org/10.1038/s43247-023-01129-1

Integration of multiple data sources: The manuscript demonstrates that each dataset (hydrodynamic simulations, SAR observations, and reported impacts) has different limitations, suggesting that combining multiple information sources could improve flood risk assessment. The discussion could mention recent developments in Earth system foundation models or flood foundation models as a possible research direction. Such models could potentially integrate hydrodynamic simulations, remote sensing observations, and historical impact records to improve consistency among different data sources and reduce the uncertainties identified in this study.

We thank the reviewer for these insights. In the revised manuscript we will strengthen the discussion of these points. In particular we will mention the possibility of using different observational data sources to better calibrate and validate flood models, considering the limitations of each, including within the scope of earth foundation models. For example GFM can be especially useful to calibrate flood models in agricultural areas, but not in urban areas, while to this aim local water depth information could be used. Regarding Earth System foundational models, multi-source datasets, such as the catalogue we built in our work, could be used for training multimodal foundation models, as heterogeneous flood data can be processed by such models. More broadly, AI-based flood models could benefit from multiple target datasets, and their agreement could provide probabilistic information.

To support this discussion, we will mention the work and findings of Xu et al. (2025), Nazari et al (2026) and Bodnar et al.(2025) .

References

Bodnar, C., Bruinsma, W.P., Lucic, A. *et al.* A foundation model for the Earth system. *Nature* 641, 1180–1187 (2025). https://doi.org/10.1038/s41586-025-09005-y

Xu, Q., Shi, Y., Zhao, J. et al. FloodCastBench: A Large-Scale Dataset and Foundation Models for Flood Modeling and Forecasting. Sci Data 12, 431 (2025). https://doi.org/10.1038/s41597-025-04725-2

Nazari, B., Ahmadisharaf, E., Bates, P. D., Moradkhani, H., Merwade, V., Wahl, T., et al. (2026). Synergistic integration of flood inundation modeling methods: A review of computational, data-driven, observational and experimental, and conceptual models. *Reviews of Geophysics*, 64, e2025RG000898. https://doi.org/10.1029/2025RG000898

Recovery Period: The use of a 90-day window is justified based on economic recovery considerations. However, from a hydrological perspective, this criterion may merge independent flood events occurring within the same period. A brief discussion of the potential influence of this assumption on event matching with the Hanze and GFM datasets, or a short sensitivity statement, would help clarify its implications.

We thank the reviewer for identifying this point. Despite already providing a sensitivity analysis of the effect of the 90 days window on the total number of simulated floods and total damages in the No Protection case, we indeed did not directly address its effect on the event matching procedure. In the revised manuscript we will address this by adding a short section reporting the number of matched events with the same windows already tested in current sensitivity analysis (windows from 30 to 360 days). Additionally, in the methods section (around line 280 of the original manuscript), we will clarify that the current matching procedures already allows to match a single simulated event (that includes multiple peaks) with several GFM or Hanze events (which can instead identify different peaks as separate events).

Hanze discussion: In Section 3.2, the large number of events detected by both GFM and the model but absent from the Hanze database in Scandinavia is attributed to sparse population and limited reporting. It may be useful to further characterize Hanze as an exposure-dependent dataset, whose spatial coverage is influenced by population distribution and reporting practices in addition to hydrological hazard. This distinction would provide a more precise interpretation of the observed discrepancies.

We thank the reviewer for this point. The evidence provided by the current manuscript indeed supports the raised point, that we will emphasize in the revised manuscript.