In the following, we provide detailed point-by-point answers to the reviewers' comments. Original comments are provided in blue, and our responses are shown in black.

This study investigates how climate conditions may impact the rain-on-snow (ROS) floods in Germany using an LSTM model trained on downscaled climate projections and explainable artificial intelligence techniques. The results suggest that while ROS flood frequency will decrease in most regions, the severity of extreme ROS floods is expected to increase across all major river basins, and the role of snowmelt in these floods will become less significant. Overall, this study is comprehensive and well-designed. However, I still have several questions and suggestions for improving the current work.

We thank the reviewer for a positive evaluation of our work and for the helpful suggestions.

1. The "magnitude" of floods is mentioned in the title. However, the absolute magnitudes of floods are represented by the relative ranks.

   Thank you for the comment. We will revise the title and text and use "severity" instead of "magnitude".

2. It is suggested to clearly define the "trans-basin" floods. How to distinguish whether the flood is "trans-basin" or "within-basin"?

   Thank you for the comment. We will clarify the definition of "trans-basin" floods and change the wording to widespread. We define the "trans-basin" (or widespread) flood as follows: there must be a flood at more than two gauges and the total affected area must be larger than 10% of the total area of the study catchments. It is not necessary that those two gauges are within different major river basins. The total are of the "all" category includes all study catchments, the categories for the major river basins (e.g., "Rhine") only include those within this major river basin.

3. Lines 5 and 82: The LSTM model may not be the state-of-art deep learning approach nowadays. It would be better to compare its performance to that of the emerging alternatives, e.g., physics-informed deep learning, and the ensemble-based approach is also beneficial given the limitations of any single deep learning model.

   Thank you for the suggestion. We will revise this part.

4. Lines 18, 471-472, and 523-524: The conclusion that the discrepancy is mainly due to different snow dynamics in climate models should be supported by more detailed arguments and evidence.

   Thank you for the comment. We will revise this conclusion as it simplifies the reasoning for the model discrepancy regarding ROS floods. Indeed, the snow dynamics coming from the land-surface models play a big role for the differences between the models, but is not necessarily the main issue. We will provide a more detailed argumentation and add references supporting it. The modeling process of snow is very complex and depends on many factors, such as the land-surface model of the RCM and its parameterizations, and the forcing data of these land-surface models produced by the RCMs. If an RCM has a cold bias, the land-surface model will likely produce more snow. However, different land-surface models may have a different sensitivity to this bias. A recent study by Risto et al. (2025) shows the sensitivity to snow mass and snow cover produced by a land-surface model to perturbations in the forcing precipitation and the albedo in seasonal forecasts. Variations in both parameters can greatly affect seasonal snow amounts. Terzago et al. (2017) find that RCMs used for downscaling of ERA-Interim and CMIP5 GCMs heavily overestimate the snow pack over the higher-elevated Alpine region, but underestimate the snow pack in the lower-elevated parts of the region. They explain this overestimation with cold biases or wet biases in the forcing coming from the RCMs or, if no biases were apparent in the forcing, to the snow scheme of the land-surface model and the spatial resolution of the RCMs. They also highlight the large spread of the models in terms of snow amount. However, all RCMs agree on the reduction of snow amount in future projections. We will include the findings of these and other related studies in our new argumentation.

5. It would be better to make the Introduction Section more concise and to clearly state the research gaps and study objectives.

   Thank you for your suggestion. We will revise and improve the introduction section.

6. Figure 1 and the other maps: It is suggested to add a north arrow and a scale bar.

   Thank you for the suggestion. However, we prefer to avoid the north arrow and a scale bar as all of our Figures already contain longitude and latitude lines to provide the spatial information.

7. Line 98: How to define the "subsurface runoff" and the "day length"?

   The subsurface runoff is simply the runoff that remains when surface runoff is subtracted from the total runoff produced by the models. In case of ERA5 this subsurface runoff already exists as separate parameter, but must be calculated for the CMIP5 RCMs. Day length is the duration from sunrise to sunset and is used as seasonality. It is calculated for each gauge station individually depending on the latitude. We will clarify the definitions in the text.

8. Line 108: The evaluation metric, NSE, has some inherent limitations. Also, given the sampling uncertainty and measurement errors in both temporal and spatial data, it is suggested to present the values of metrics through a statistical distribution instead of a fixed number for a single evaluation period and to apply the metric to the variable of interest like specific flood peaks. The authors can refer to the article below for more information about the limitations of some commonly used evaluation metrics in hydrology.

   Reference: "Beyond a fixed number: Investigating uncertainty in popular evaluation metrics of ensemble flood modeling using bootstrapping analysis" (https://doi.org/10.1111/jfr3.12982)

   Thank you for the comment and the suggestions. We recognize the limitations of the NSE as a single evaluation metric and will provide the KGE as an additional evaluation metric.

9. Table 1: Could you add a reference for the criterion, "$SWE > 15mm$"?

   Other studies used and tested different thresholds such as 4-10mm (Mudryk et al., 2017), 8cm (Xiao and Liang, 2024), 5-15mm (Matiu and Hanzer, 2022) or density dependent thresholds (Kouki et al., 2023; ECMWF, 2016). Kouki et al. (2023) calculated the snow cover for ERA5 following the formula integrated in the IFS model ECMWF (2016). We used a rather conservative threshold to include the importance of the snow cover for flood generation. We will change the note about the 15mm SWE as threshold for a snow day and will reason our choice.

10. Line 230: How about the potential impacts of human activities on the other gauges? Any thoughts about how to incorporate those infrastructures into the modeling process, especially for the urban areas?

    Thank you for bringing this important point up. Anthropogenic activities likely will impact the other gauges, too. The impact, however, may vary considerably depending on the type of activities and their persistency. If the activities happen on a regular basis over the entire period the LSTM should be able to learn the anthropogenic impact similar to learning a systematic bias within a climate model. If the activities do not happen on a regular basis or are inconsistent within the training data (e.g. structures built during the time period) the model is not able to learn the impact. This is likely the reason why some gauges clearly underperform compared to other gauges with a similar catchment area. Solving this problem for a data driven model is potentially very difficult as it requires additional input that may not be accessible.

11. Figures 3, 7, 8, and 9: The horizontal axis, "Stations", was sorted by different criteria, such as the mean NSE and the median number of ROS floods. Is it possible to keep it consistent? So it is relatively easy to identify each gauge station.

    Thank you for the suggestion. We will change the sorting criteria for Figures 3, 4, 7, and 8 to the median number of ROS floods as shown in Figure 9 to better display the performance of the simulation in catchments with large numbers of ROS floods.

12. Figure 8: Please add the denotation for the shaded areas to the figure caption.

    Thank you for pointing this out. We will add the 10th and 90th percentiles to the description of Figures 4 and 8.

13. Line 370: It would be better to specify the exact p-value rather than "$P < 0.05$".

Thank you for the suggestion. We will add a table containing the p-values in the Supporting Information.

14. Line 479: Is it "GMCs" or "GCMs"?

    It is GCMs. We will correct this issue.

References

ECMWF: IFS Documentation CY41R2 – Part IV: Physical Processes, ECMWF, pp. 108–111, https://doi.org/https://doi.org/10.21957/tr5rv27xu, 2016.

Kouki, K., Luojus, K., and Riihelä, A.: Evaluation of snow cover properties in ERA5 and ERA5-Land with several satellite-based datasets in the Northern Hemisphere in spring 1982–2018, The Cryosphere, 17, 5007–5026, https://doi.org/10.5194/tc-17-5007-2023, 2023.

Matiu, M. and Hanzer, F.: Bias adjustment and downscaling of snow cover fraction projections from regional climate models using remote sensing for the European Alps, Hydrology and Earth System Sciences, 26, 3037–3054, https://doi.org/10.5194/hess-26-3037-2022, 2022.

Mudryk, L. R., Kushner, P. J., Derksen, C., and Thackeray, C.: Snow cover response to temperature in observational and climate model ensembles, Geophysical Research Letters, 44, 919–926, https://doi.org/https://doi.org/10.1002/2016GL071789, 2017.

Risto, D., Fröhlich, K., and Ahrens, B.: Seasonal Snow Simulation: Sensitivity to Initialization, Parameterization, and Forcing, Earth Systems and Environment, pp. 1–12, https://doi.org/https://doi.org/10.1007/s41748-025-00728-6, 2025.

Terzago, S., von Hardenberg, J., Palazzi, E., and Provenzale, A.: Snow water equivalent in the Alps as seen by gridded data sets, CMIP5 and CORDEX climate models, The Cryosphere, 11, 1625–1645, https://doi.org/10.5194/tc-11-1625-2017, 2017.

Xiao, X. and Liang, S.: Assessment of snow cover mapping algorithms from Landsat surface reflectance data and application to automated snowline delineation, Remote Sensing of Environment, 307, 114163, https://doi.org/https://doi.org/10.1016/j.rse.2024.114163, 2024.