

We would like to start by thanking the reviewer for their constructive comments on the manuscript. These have made a significant contribution to the clarity and readability of the work. Below, we provide a point-by-point response to each comment by the reviewer in green and starting R: We believe we have addressed all the comments in full but welcome any feedback.

The study drew on experiences from multiple cyclone events in Southeast Africa to forecast the number of people exposed to these flood events, and highlight challenges in interpreting and communicating forecasts due to substantial uncertainty across complex hydrologic and hydraulic modeling processes. Overall, this study is meaningful for more informed flood risk management, at least for the study area. However, I still have several comments and suggestions as follows.

1) The reasons why Southeast Africa was selected as the study area should be elaborated in more detail. The lessons learned from the cyclone flooding events may be only applicable to the specific regions of Southeast Africa, and generalized guidelines beyond the study area are suggested to be discussed.

This is a good question. The tropical cyclones we responded to were driven by requests from our funder (FCDO), we did not respond to all tropical cyclones. The responses tended to favour locations where the need to provide forecast information was perceived to be greatest. We did respond to other TC outside of Southern Africa (e.g. TC Amphan). However, as these are small in number, we decided to focus our analysis on Southern Africa, where we had experience from multiple cyclones. Our approach can and was implemented globally, but the reviewer is correct that some conclusions might not be transferable. For example, GloFAS accuracy is not globally consistent, and this is also likely the case for the Flood Hazard layers. We added to the introductory text to signpost these limitations and decisions.

“Given their on-demand nature, bulletins were not provided for all tropical cyclones. The majority of bulletins were requested for tropical cyclones marking landfall in Southern Africa, and therefore, we decided to focus the analysis in this paper on the region. Not all conclusions will be transferable to other regions, however we would need to conduct more bulletins to assess this.”

2) It is suggested to add a list of acronyms mentioned in the manuscript. The full term of the acronym is only presented the first time it appears, e.g., GloFAS. Some acronyms seem to be unnecessary, e.g., JRC.

Thank you for the suggestion. On careful reading of the manuscript, we have been able to remove half of the acronyms and retain only those we think are in common use. So rather than add another table, we have simply reduced their usage to improve readability.

Removed acronyms: ADAM, JRC, IFS, SAR, SOP

Retained acronyms: FCDO, ECMWF, UNOSAT, GloFAS

3) Line 139: Why were different versions of GloFAS used for different events? Would this lead to another uncertainty source about streamflow estimates?

Indeed, it's unfortunate that the data version changed during the study. This was a relatively minor update to the system and we do not expect it to have major implications for our study.

4) Line 172: Is “WorldPop population data (population counts)” a section title?

Yes, sorry, this was an editing oversight. corrected

5) Table 1: Five or six events were investigated? Six or seven landfalls were investigated? Please indicate the format of the date, e.g., DD/MM/YYYY.

Table corrected and date format added.

Table 1: Summary of recent tropical cyclones included in this study, including four tropical cyclone systems and seven country landfalls.

6) Figures 2-3: What is the unit for population? It is suggested to improve the quality of most of the figures in the manuscript, e.g., keep the labels clear and the font size consistent, and the sub-figure title is overlapped with the vertical label in Figure 4.

R: Thankyou for spotting this omission. Units added to captions for the figures.

7) Lines 230-231: The statement that “these outliers are typically driven by ensemble members that cause flooding over areas of high population, rather than being ensemble members with exceptionally intense rainfalls” should be supported by some evidence.

R: This is a good point. We have removed the comment as we don't feel we have enough evidence or TC examples to make such a strong statement

Removed :“, rather than being ensemble members with exceptionally intense rainfalls”.

8) Lines 246-247: Considering that the ensemble mean would be sensitive to outliers, why not use ensemble median for the further analysis?

R: We mostly work with percentiles so we think this is OK in the manuscript.

9) Figure 4: Why is the lead time not the same among the seven landfalls, 5~-1 days and 4~-1 days? What do different colors stand for in the lower plots? Maybe a typo: “Gombe Malwai” or “Gombe Malawi”?

R: Thanks for spotting this. We removed the 5th day if no exposure was forecast, but it should be clear from the figure caption. Text added, “The five-day lead time is omitted where no exposure was forecast”.

10) Figure 5: Figures 5a-5d are for the total precipitation $\geq 200\text{mm}$, while Figures 5e-5f are for the total precipitation $\geq 300\text{mm}$?

R: Thank you for spotting this. Corrected.

11) Line 278 and Figure 5c: 30th February is not true.

R: Corrected to 27th

12) Figure 6: It would be helpful to present the population and the percentage of urbanization for each district?

Figure 6 has been updated to show district-level total population estimates based on Worldpop data (2019 estimates).

13) Table 3: The equation used for calculating the “Diff %” may not be the same, e.g., those for “Other (<1000 exposed)”. It would make more sense if the max exposure from observed extents is used as the denominator.

R: We agree that this is a clearer way to present these differences. The table was updated to the percentage change from remotely observed. The caption was also updated to clarify how Diff % was calculated.

14) Lines 404-405: Given the various uncertainty sources that were not considered in the case study, the findings may be subject to change after more comprehensive analyses are conducted. Also, the uncertainty in model evaluation should not be ignored. The authors can refer to the article below for more information about the limitations of some commonly used evaluation metrics in flood models.

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>)

R: Reference added, and we have also included the caveat regarding model uncertainties changing our findings.

15) Lines 429-430: It would be better to quantify how large or how small the district is for the corresponding conclusion.

R: Done.