

We would like to start by thanking the reviewer for their comments on the manuscript. These have made a significant contribution to the clarity of the work and the depth of the literature review and discussion. 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.

## Summary

The paper presents a methodology for forecasting the number of people directly exposed to flooding from tropical cyclones in Southeast Africa, aiming to improve anticipatory humanitarian action. It combines meteorological (ECMWF), hydrological (GloFAS), hydraulic (Fathom flood maps), and population (WorldPop) models to create forecasts of flood exposure. The study highlights both the potential and limitations of forecasting flood exposure for tropical cyclones, as well as the experience gained in communicating these forecasts. As such, it is a valuable contribution to the field of disaster risk management. However, the introduction and discussion sections should better position the study within existing literature. Furthermore, important choices made in the model chain should be clarified. Therefore, I recommend a major revision before publication.

## Major comments

Introduction: The introduction should include references to other studies on operational forecasting of tropical cyclone (TC) flood hazard and/or exposure, including those that dynamically run hydrodynamic or surrogate models. This would help better position the current study within the context of existing literature

R: We agree with the review that some more positioning with existing systems would be a valuable addition. How to do this is complicated because the more ‘advanced’ hydrodynamic modelling systems are run in other regions and we are not aware of a global solution. Therefore, there is a significant gap between what is possible and what is typically done in practice for our focus region. We add a paragraph from line 71 in the paper, which is rather long to copy here.

L198–206: More explanation is needed to understand how the preprocessed flood inundation maps are combined with the GloFAS discharge data. Is a single return period map selected for the entire domain or per sub-basin?

R: The return period is applied per pixel by interpolating the GloFAS river return period to the nearest floodplain cells. The reviewer is absolutely correct that this should be noted in the methods and that there are several ways to go about this. Text added:

“Forecast return period changes for every GloFAS grid cell. To project this variability onto the uniform return periods of the Fathom flood hazard maps and generate an ensemble of flood footprints, the hazard maps are sampled per pixel using the nearest available GloFAS return period to choose which hazard map to use at each pixel. Therefore, each 30 m pixel is assigned 51 return periods from GloFAS. Given that the uncertainty in the GloFAS return period was typically large relative to the sampling of return periods by the hazard maps, we select the closest available Fathom return period map that has been simulated rather than interpolating between the hazard maps. The method is conceptually similar to that adopted for operational rapid flood risk assessment in Europe (Dottori et al., 2017) and the event set generation method described by Olcese et al. (2024), except that we calculate a pixel-based return period rather than small sub-basins. “

Olcese, G., Bates, P. D., Neal, J. C., Sampson, C. C., Wing, O. E. J., Quinn, N., et al. (2024). Developing a fluvial and pluvial stochastic flood model of Southeast Asia. *Water Resources Research*, 60, e2023WR036580. <https://doi.org/10.1029/2023WR036580>

How are water levels interpolated between pre-calculated return period-based maps?

Answered in edit above. We kept this very simple because the return period uncertainties in the forecast are very large and we wanted the method to be as computationally lightweight as possible, such that it could be implemented in a desktop PC in the future.

How are GloFAS discharge forecasts linked to the Fathom flood maps?

R: Maps are only linked via return period. Absolute discharge is not used because the two models have different sources for their hydrology – Fathom use a regional flood frequency analysis method and while GloFAS is a hydrological model. We have added text to explain this approach and noted the main benefits and limitations associated with it.

“GloFAS forecasts and the hazard maps are linked only via return period, discharge is not considered because the two methods use different methods to generate discharge. We expect significant bias between the regional flood frequency analysis (Smith et al., 2014) used to estimate discharge in the Fathom flood hazard model and the hydrological model simulations used by GloFAS given previous comparisons (Devitt et al., 2021), meaning that bias between the models would dominate the inundation forecast if absolute discharge were used. Linking via return periods removes this bias, although it assumes both models have the same growth in discharge with return period relative to the mean annual flood.”

Smith, A., C. Sampson, and P. Bates (2015), Regional flood frequency analysis at the global scale, *Water Resour. Res.*, 51, 539–553, doi:10.1002/2014WR015814.

Devitt, L., Neal, J., Wagener, T., Coxon, G., (2021) Uncertainty in the extreme flood magnitude estimates of large-scale flood hazard models. *Environmental Research Letters*, Vol. 16, No. 6, <https://doi.org/10.1088/1748-9326/abfac4>

Is a flood depth threshold used to determine "people flooded"? How is flood protection (e.g., levees or raised houses/thresholds) considered in this process?

R: We use a threshold of 10 cm, although the choice of this threshold is largely spurious within the precision to which we report the results. We take no account of mitigation measures. A household is flooded if the water is at the house in our model and no flood defence measures are considered. Text added

“A household was assumed flooded if the water depth was above 10 cm in the hazard maps. This value was arbitrary; however, the sensitivity of population exposure to this threshold was largely spurious within the precision to which we would advocate reporting the exposure results. No attempt was made to account for flood defences or other flood protection measures such as raised floors.”

### Minor comments

L68: "resolution well below 100m is necessary" — This paragraph focuses strongly on spatial resolution but does not mention vertical accuracy, which is equally important. I suggest discussing the required vertical accuracy in this paragraph as well.

R: Agreed, text added.

“Unlike most other natural hazards, resolution well below 100m is necessary because flood hazard changes materially over a few decimetres and people will avoid settling in hazardous locations if they can (Smith et al., 2019). Early large-scale flood models (often at 1 km resolution) faced a credibility challenge due to simulating divergent estimates of flood extent (Trigg et al., 2016). Recent advances in global flood hazard mapping can theoretically provide inundation depth estimates with sufficient resolution (~30 m) to estimate community level flood exposure when combined with high-resolution population data (Hawker et al., 2024; Wing et al., 2024). However, errors can be substantial at such fine resolutions due, in part, to the limited vertical accuracy of global terrain data (Meadows et al., 2024). Although there have been notable improvements in global terrain data (Hawker et al., 2022), which mean they might, for the first time, be able to support district level hazard and exposure forecasts. It remains the case that only 80% of vertical errors are below 2 m on floodplains in the DEM we use here (FABDEM) (Hawker et al., 2022). When combined with substantial errors in hydrological predictions, we know that inundation forecasts will be complicated by substantial and likely unquantified uncertainties, meaning interpretation and use of these data requires significant care.”

Meadows, M., Jones, S. & Reinke, K. (2024) Vertical accuracy assessment of freely available global DEMs (FABDEM, Copernicus DEM, NASADEM, AW3D30 and SRTM) in flood-prone environments, *International Journal of Digital Earth*, 17:1, 2308734, DOI: 10.1080/17538947.2024.2308734

L172: "Worldpop population data". This should probably be a new subsection.

R Done. We removed this text for the figure and added what was needed to the legend.

L223: "We stop one day after landfall .." Is this standard SOP or specific to the case of Freddy? For TC Idai for example, for example, heavy rainfall and flooding continued for several days after landfall.

R: Yes, this was a decision for this analysis rather than a requirement. We've added the suggested clarification.

“In this analysis, we stop one day after landfall under the assumption that satellite imagery of the flooding is more likely to be used post-event. However, it should be noted that the standard operating procedures are flexible, and forecasting can continue after landfall when heavy rainfall persists (e.g. in the case of tropical cyclone Idai). “

L240: When does a highly uncertain forecast become "misinformative"? Here you state that "the forecasts tend not to be misinformative" but in the discussion around L419 that "Forecasters should consider the possibility that exposure forecasts could be misinformative". Can you elaborate on this?

R: We thank the reviewer for spotting this lack of clarity in the writing. With respect to L240, we have limited this statement to the case where the objective is to identify potentially exposed districts. The previous version of this sentence was too broad. At line 419, we also think the original version lacks precision. We have revised the text to be more specific and avoid the use of the term misinformative:

“Forecasters should consider the possibility that exposure forecasts can change substantially as revised cyclone track forecasts are issued. In this case, Freddy tracks substantially further inland than the early ECMWF forecasts suggest.”

Figure 4: To clarify, do the colors in Figure 4 refer to the same district across each forecast lead time? If so, it would be helpful to state this explicitly in the figure caption and the text around L247. It would also be interesting to assess whether the top five districts remain stable across lead times or if there are large changes in ranking or top-5 districts.

R: Indeed, this is the case. We have clarified in the caption

L329: "In general, the remote sensing is likely to have missed substantial pluvial and flash flooding on smaller rivers." — Please add a reference to support this statement.

R: Done

Matgen, P.; Martinis, S.; Wagner, W.; Freeman, V.; Zeil, P.; McCormick, N. *Feasibility assessment of an automated, global, satellite-based flood-monitoring product for the Copernicus Emergency Management Service*. EUR 30073 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-10254-0, doi:10.2760/653891, JRC119812, Ispra.

L456: "A like-for-like comparison of the inundation modelling and remote sensing was highly problematic in this context because the inundation event is not simulated dynamically". In the results section (around L358) the main reason for not comparing the modelled and observed flood extents is based on limitations of the satellite data, while here in the discussion it is based on limitations of the model. It's probably both, but it would be good provide a balanced and consistent argument in both sections.

R: We agree, it was not our intention to present this issue only in terms of remote sensing limitations. Text added around 358 to balance the rational as suggested.

"In addition, the inundation forecasts are for the maximum return period rather than the flooding at the time of the overpass. Therefore, we believe it is reasonable to assume the remote sensing is a conservative estimate of the actual maximum flood extents and that the inundation forecasts are greater than would be observed at any specific time."

L473: Please consider adding literature from other studies that have done dynamic operational flood forecasting for TCs, e.g. the COSMOS system in the USA.

R: We agree with the reviewer that this is a missing component of the literature in the discussion, and we are very happy to add this. However, we want to be careful about how this is presented and also not make this paper overly technical. While we would advocate for using a dynamic model, and the COSMOS is a great exemplar, this would be a big step for the region and in our experience doing something simple and theoretically transferable to local agencies has real value even if it sacrifices accuracy compared to more advanced approaches. We believe this additional text provides a better balance as suggested by the reviewer:

"A simple hazard map lookup system, as used here, should be expected to perform poorly where there is substantial compounding between a large river and its tributaries, regardless of the skill of GloFAS on individual rivers. Independent simulations of the river flooding on the main river and its tributaries, or a fully dynamic simulation of river flows from each ensemble member that captures river-river compounding effects, are needed. One solution would be to implement a dynamic model run of the inundation model. The technical capability to simulate inundation dynamics from tropical cyclones in a forecasting context has been demonstrated both in research and operationally (e.g. Barnard et al., 2019; Leijnse et al., 2021). However, this is a non-trivial undertaking because dynamic simulations would need to be run for each (or many)

ensemble members to capture the forecast uncertainty from GloFAS. This may i) prove to be computationally expensive if 30m resolution is maintained (although various lower resolution sub-grid approaches are available (e.g. Neal et al., 2012; Leijnse et al., 2021; Sanders et al., 2025), and ii) be difficult to implement operationally because robust code is needed to run in real time along with reliable access to computing resources.”

Barnard, P.L., Erikson, L.H., Foxgrover, A.C. *et al.* Dynamic flood modeling essential to assess the coastal impacts of climate change. *Sci Rep* **9**, 4309 (2019). <https://doi.org/10.1038/s41598-019-40742-z>

Leijnse, T., van Ormondt, M., Nederhoff, K. and van Dongeren, A. Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes, *Coastal Engineering*, 163, (2021). <https://doi.org/10.1016/j.coastaleng.2020.103796>

Sanders, B.F., Schubert, J.E., Martin, E.M.H., Wang, S., Sukop, M.C., Mach, K.J. A fast flood inundation model with groundwater interactions and hydraulic structures. *Advances in Water Resources*, 204, (2025). <https://doi.org/10.1016/j.advwatres.2025.105057>

L475: "This may i) prove to be computationally expensive if 30m resolution is maintained," Please consider adding literature from other studies that use surrogate models or subgrid approaches to balance model speed and accuracy. And how important is a 30m resolution model in the context of other uncertainties in the flood model chain such as the use of a hazard map lookup system, the uncertainty in the DTM, or the uncertainty in the hydrological model simulations?

R: The reviewer is correct that there are sub-grid methods in the literature that would almost certainly be suitable given the forecast uncertainty. We develop one. We have adapted the text to make it clear that this is an option and added some examples. We could add more detail here, but we could not cover everything without making the paper substantially more technical.

This may i) prove to be computationally expensive if 30m resolution is maintained, although various lower resolution sub-grid approaches are available (e.g. Neal et al., 2012; Leijnse et al., 2021; Sanders et al., 2025) that are likely to be suitable given the uncertainties in the tropical cyclone rainfall and river flow forecasts, and ii) be difficult to implement operationally because robust code is needed to run in real time along with reliable access to computing resources.

reviewerResponse