

Authors' response to anonymous referee #1

Authors: Luca G. Severino, Evelyn Mühlhofer, Nishadh Kalladath, and David N. Bresch
Date: 20 October 2026
Journal: *Natural Hazards and Earth System Sciences*
Revised manuscript: *Beyond the flood map: an open-source framework for forecasting cascading impacts of natural hazards on people and essential services*

1 General comments

Referee comment: This paper details an approach to estimate indirect impacts from flooding: specifically loss of access to health centres due to inundation. Clearly the work is relevant to improving specific information for advance protection of communities as well as response. As the authors show however, integration of hazard, exposure and vulnerability also compounds uncertainties, such that in the studied contexts (Sudan and Somalia) even aside from the difficulty forecasting flood hazard, there is substantial uncertainty even in locating health centres (as an aside it suggests that the most fruitful path to reducing uncertainty may be generating up-to-date information on the location of exposed elements - which in principle is simpler than the uncertainty associated with predicting convective rainfall). The paper is well written and structured and I have no significant concerns to address before publication.

Authors' response: We thank the reviewer for the kind and constructive words. We have taken into account the different suggestions to improve our manuscript. In particular, the revised manuscript contains the following changes with regard to the reviewer's comments:

- An extended discussion regarding the discriminating power of the forecast and the robustness of the assessment of the impact-based trigger has been added.

Below you will find our detailed responses to the different comments and suggestions. The changes refer to the revised version of the manuscript that we are planning to resubmit. We are grateful for the time and attention that the anonymous referee #1 and the editor gave to our submission and hope that the revisions will improve the quality of the manuscript.

2 Major comments

2.1 Major comment 1

Referee comment: My only major comment is on the interpretation of the P(number of people

indirectly affected > threshold) being above $P(\# \text{ directly affected} > \text{threshold})$ which in turn being above $P(\text{discharge} > \text{return period})$. i.e. figure 6 and associated discussion. It is implied that translation to indirect impacts enhance trigger decisions (e.g. title of section 3.1.3). However I am not sure this is the case. Fig 6 shows that the probabilities of indirect /direct impacts reach much higher levels than discharge return period forecasts, with the latter “too weak to trigger anticipatory action” (l. 316). Yet the critical value of a forecast to triggering a decision is in its discrimination. That is: high probabilities before ‘events’, but also low probabilities (ideally, mostly when events do not occur). In this case it’s my understanding that this critical discriminating information is coming entirely from the GloFAS forecast (and tracking it back, from the rainfall forecast ultimately). Everything else in the workflow provides more detail. In the case of comparing probabilities of discharge level and probabilities of impacts, is it not just a non-linear scaling? i.e. instances of higher likelihood of discharge will presumably give higher likelihood of impacts. Put another way, the fact that higher probabilities found after translating to impacts is arbitrary and depends on threshold used to determine # people (indeed “at least 1 person affected” in a flood prone region is quite a low bar to compare with a N-year return period event). All of that is to say that I believe the impact layer will not actually enhance the trigger decision itself by improving discrimination of the model to advise on event occurrence between forecasting instances. Indeed this cannot be checked directly since the authors only consider one season per location - a more systematic look at generating and evaluating an impact ‘hindcast’ would be need for this (which I appreciate is out of scope but might be worth considering in future). Having said that I do believe there is clear value in this method for decision-making. Although I am not convinced it enhances the trigger decision itself, I believe the value lies in the enhancement of the information available for planning, e.g. mapping populations likely to be cut off from services given a specific forecast. Before publication I’d welcome the authors take on this point, and where relevant to incorporate it into the discussion.

Authors’ response: We agree that most of the discriminating power comes from the hydrological forecast itself. The performance of the impact forecast depends on the performance of the discharge forecast, and the impact modelling pipeline will only be of value to anticipatory action if a certain skill can be attained by the hydrological forecast. This specific level of skill depends on the implementation of the actions, where lower levels of discrimination can be tolerated in certain cases, for instance for low-regret actions. However, we do believe that the impact modelling pipeline can improve discrimination and then enhance trigger decisions. This improvement comes from both the inundation modelling, and the impact modelling. For instance, a high level of discharge will not always lead to a high inundation, as this depends on the morphology of the river basin and the river banks. Additionally, a high level of inundation will not always lead to a high level of impacts, as this depends on the exposure and vulnerability. By including exposure, vulnerability, and a geographically-explicit representation of floods, our framework can help discriminating between these situations, where high discharge levels lead to no floods, or floods lead to no impacts. We

also believe that our comparison between impact-based and discharge-based threshold is fair, as low and high thresholds are considered for both methods. For the discharge-based threshold, we believe that a threshold of a 2-year return period event is comparable with the impact threshold of at least one person affected, as both thresholds correspond to lower-impact and high frequency events.

This being said, we agree that a skillful meteorological or hydrological forecast is a prerequisite to a useful impact forecast. As you highlight, this aspect is not sufficiently discussed in our manuscript and we have revised the manuscript to include a more developed discussion on this aspect (in section “Application to anticipatory action”). We also agree that our analysis would benefit from a more systematic assessment looking at longer time series to more robustly assess the value of incorporating impacts to trigger anticipatory action. This is discussed in section “Limitations, improvements, and future research” in the revised manuscript.

In section “Application to anticipatory action” in the revised manuscript:

A comparison of probabilities of exceedance of hypothetical impact-based triggers with probabilities of exceedance of discharge-based triggers, which are used operationally, suggests that forecasts of impacts correlate better with observed flood impacts than discharge forecasts. We explain this by the fact that the impact model adds spatially-resolved exposures and an inundation model, which can help determine whether a predicted extreme river discharge will actually lead to impacts. This indicates that impact forecasting could be used to help refine existing anticipatory action based on hazard forecasts. However, we note that the discriminating power of the impact forecast remains intrinsically dependent on the discriminating power of the input hazard forecast. Hence, the inclusion of impacts only helps refine an existing signal, which depends on the skill of the meteorological or hydrological forecast. Our analysis does not reveal that including indirect consequences of infrastructure failures and mobility restrictions modifies the predicted timing of flood impacts. However, the inclusion of indirect impacts does increase the probability of exceeding certain impact levels, which can be used to refine trigger mechanisms.

In section “Limitations, improvements, and future research” in the revised manuscript:

The lack of impact data on post-disaster healthcare access and infrastructure damage restricted our analysis to a couple of case studies. We expect an analysis based on a more extensive number of case studies, spanning longer time frames and expanding over different regions to provide more robust results, yielding a clearer picture of the actual benefits of impact forecasting and the inclusion of indirect impacts. This is particularly true for our analysis of how the impact forecast can be used to trigger anticipatory action, which is based on only two locations for a single event and is therefore not robust. An extended analysis based on multiple events and locations requires structured data documenting impacts on service accessibility, at specific levels of spatial and temporal disaggregation.

3 Minor comments

Referee comment: l21 First paragraph of introduction is long, can split it somewhere

Authors’ response: This has been changed in the revised manuscript.

Referee comment: l66 This final paragraph is a little repetitive of what came before: could instead be used to signpost the structure of the paper

Authors’ response: Thank you, this has been rephrased in the revised manuscript.

Referee comment: l80 “spatially-explicitly” can be rephrased more simply without a novel hyphenation

Authors’ response: Replaced with *in a probabilistic and spatially explicit way*.

Referee comment: l87 please note which GloFAS version? v4 had serious problems in Somalia in 2023

Authors’ response: For this analysis, the archive of operational forecasts is used. This corresponds to GloFAS version 4.0 for the Somalia case study and GloFAS v4.1 for the Sudan case study. This has been clarified in the revised manuscript

Referee comment: l88 please give a quick explainer of what Dottori provide (is it satellite or model based - presumably inundation?). Also more explicit detail of the ‘model’ of Riedel would be appreciated, how do the combine stream flow with flood hazard maps?

Authors’ response: A more detailed description of the model of Riedel, and Dottori is provided in the revised manuscript.

In section “Flood forecast”:

We use the model introduced in [Riedel et al. \(2024\)](#) to model river flooding. This model combines daily river discharge data from the global hydrological model Global Flood Awareness System (GloFAS; [Alfieri et al., 2013](#)) with flood hazard maps from [Dottori et al. \(2016\)](#) to obtain event-specific flood inundation maps. In an initial step, the statistical distributions of daily river discharge values are obtained for each grid-cell of the GloFAS domain by fitting Gumbel distributions using 36 years of river discharge reanalysis from GloFAS. These statistical distributions can then be used to derive event-specific return period maps for reanalysis or forecasts of the river discharge data that GloFAS produces for specific days. These event-specific return period maps are then bi-linearly interpolated on the grid of the flood hazard maps, and flood depths are estimated using linear interpolation to link the re-gridded return periods to the corresponding flood depths from the flood hazard maps ([Riedel et al., 2024](#)). The flood hazard maps from [Dottori et al. \(2016\)](#) have been produced by incorporating hydrological model output obtained from GloFAS for events corresponding to specific return peri-

ods into a two-dimensional hydraulic model. We use the updated version of the flood hazard maps released on 14 March 2024 at an initial resolution of 3 arc-seconds ([Baugh et al., 2024](#)). For computational purposes, these maps are up-scaled using an aggregation factor of 7. For this analysis, we use GloFAS v4.1 for the river discharge reanalysis, and the archive of operational forecasts for the river discharge forecasts. This corresponds to GloFAS v4.0 for the Somalia flood event of 2023 and GloFAS v4.1 for the Sudan flood event of 2024 ([Copernicus Emergency Management Service CEMS and European Centre for Medium-Range Weather Forecasts ECMWF, 2025](#)). We use the probabilistic forecasting ensemble of 50 members and forecast lead-times of five and 10 days. This results in a daily reanalysis and a daily probabilistic ensemble of 50 forecasts of flood extents and flood depths with a resolution of 21 arc-seconds.

Referee comment: l95 worth mentioning here what lead time is used

Authors’ response: The specific lead-time considered in this analysis is now mentioned in the revised manuscript.

Referee comment: l100 is ‘residential level’ a defined term in OSM? And, do you filter out so that these are not included, or you this is the lowest level included? Also ‘filter out’ doesn’t need a hyphen).

Authors’ response: The exact road types used for the analysis are now described in the revised manuscript.

In section “Population and critical infrastructure exposures”:

OSM uses a classification system for road data that categorizes roads into 9 types, from the highest level; motorway, to the lowest level; track ([OpenStreetMap contributors, 2026](#)). To simplify the processing of the road network, we filter out roads at the residential level and below, thus only keeping the first 6 categories of roads; motorway, trunk, primary, secondary, tertiary, and unclassified.

Referee comment: l104 Would welcome an overall comment on spatial resolution - populations scaled to 2km yet the flood data is 20 arc-seconds. At what resolution are the layers when they’re actually interacting with each other, and what is the impact of losing the population resolution by up-scaling?

Authors’ response: In CLIMADA, hazard data is spatially overlaid with exposure data by using a nearest-neighbor algorithm in order to compute the impacts. Hence, each exposure data point will be assigned its geographically closest hazard centroid. In our set-up this means that each exposure data point, on either a 2 or 4km grid is assigned the nearest hazard centroid on the 21 arc-seconds grid (roughly 630m at the equator). We included a short comparison of results obtained with different population resolution in the appendix. We show that using lower resolution for the

population layer does not change our results qualitatively. Additionally, we remind in the revised manuscript that our results are demonstrative of the entire modelling pipeline rather than a specific best situational advice, and that users should ideally use the best data available to them.

Supplementary section “Sensitivity to the resolution of the population data”:

We assess the sensitivity to the resolution of the population data by computing the difference between modelling results obtained with WorldPop at its native resolution of 1km, and with the up-scaled resolution of 2km. Table B1 shows the difference between the two resolutions as:

$$\Delta S(\text{resolution}) = S(\text{WorldPop}_{2\text{km}}) - S(\text{WorldPop}_{1\text{km}}) \quad (1)$$

where S represents one of the access statistics (healthcare access loss, healthcare access undisrupted, healthcare no base access), expressed in terms of proportion of population affected. The comparison reveals that the resolution of the population data only limitedly affects the accessibility statistics, with an average difference below 0.05 of population affected for each statistic across the four regions. Comparing the influence of population resolution across the different regions, the strongest influence of the population resolution is in Jubaland, with an average difference across the three metrics of 0.08. In Galmudug, Southwest, and Hirshabelle, the average difference is below 0.02. This suggests that the resolution of the population data does not affect our results qualitatively, and that results obtained using a higher resolution for the population data are likely to be contained in the uncertainty intervals displayed in Fig. 7. However, we are not able to rule out the effects from parameter interaction with this complementary analysis, which might increase the total influence of the population resolution on the results.

Table B1: Comparison between population resolution

Region	Access loss	Access undisrupted	No base access
Galmudug	−0.0	−0.03	0.03
Southwest	0.0	−0.02	0.02
Hirshabelle	0.0	−0.01	0.01
Jubaland	−0.04	−0.08	0.11

Referee comment: l108 Can the threshold function be clarified, presumably a binary filter where anything over 50cm is considered affected, otherwise not?

Authors’ response: This has been clarified in the revised manuscript.

Referee comment: l116 ‘inter-dependent’ is said several times in a few lines. Consider a simpler rephrasing, especially since the dependencies are explicitly detailed, hence introducing them as

‘inter-dependent’ is redundant.

Authors’ response: Thank you for the suggestion, we have simplified the phrasing in the revised manuscript.

Referee comment: l136 “These impact forecasts are...” I didn’t quite follow this step, consider being more explicit.

Authors’ response: The re-aggregation was referring to the step described in section 2.3, where it is explained that obtained geo-located lines or polygons [from OSM] are then disaggregated within CLIMADA for the impact calculation and later re-aggregated at their original shape. The ambiguous phrasing has been removed for clarity.

Referee comment: l152 please describe what FLOPROS is

Authors’ response: We have added an explanation of FLOPROS in section Flood forecast:
The model of [Riedel et al. \(2024\)](#) also proposes an additional flood layer that uses the FLOPROS dataset to model the effects of flood protection measures on inundation ([Scussolini et al., 2016](#)). FLOPROS is a globally consistent database of flood protection standards which has been derived by combining information from different sources such as policy documents, technical reports or modelling results. [Riedel et al. \(2024\)](#) assigns flood depths of 0 to locations where the modelled flood return period is lower than the flood protection standards from the merged layer of FLOPROS at the same location.

Referee comment: l153 “medium-range” is a fuzzy definition but more generally describes forecast lead time beyond one week. Most forecasters would probably consider 5 days belonging to the weather timescale.

Authors’ response: This term has been removed in the revised manuscript.

Referee comment: l212 “170 thousand” check journal style, likely better as 170,000

Authors’ response: Thank you for pointing this out. We have standardized the formatting of numbers as suggested.

Referee comment: Figure 2 please add subfigure labels and corresponding caption description. Currently also assumes readers can tell the difference between Sudan and Somalia on a map! Also the top left plot shows the two countries floating in the same box - this could be interpreted as representing an exact geographical position, which it is not. Suggest separate boxes for these. Finally, the inundation data on the map (source, date?) is not identified in the caption.

Authors’ response: The figure and captions have been improved in the revised manuscript fol-

lowing your suggestions.

Referee comment: l227 Simulated “directly affected” is compared with OCHA-reported “total affected” as a way of verifying the realism of the model. I’m not sure this is a reasonable comparison point however - simulations consider directly affected if they are standing in at least 50cm of water: is this comparable to the way OCHA is counting? Or is OCHA inclusive of people who are indirectly affected as well?

Authors’ response: We have amended the discussion section of the revised manuscript to address this limitation:

Furthermore, our definition of people directly affected may differ from the definitions adopted by the different organisations, which can bias the comparison. In particular, UN OCHA estimated the number of people affected in Sudan by multiplying by a factor of 5 the number of houses damaged or destroyed, and adding the reported number of people killed or injured. This statistic is driven by the number of people affected by damages to houses (118468 houses damaged or destroyed and 209 people injured or killed at the country level), which then makes a comparison with our estimation of directly affected people sensible. However, our framework does not include people killed or injured, and estimates from other sources might be subject to other biases, as for instance IFRC reports do not systematically indicate the methodology used to estimate the number of people affected.

Referee comment: Figure 3 x axis labels might be clearer to show just 1st of months and leave out the year (1-Jul, 1-Aug etc)

Authors’ response: Thank you for the suggestion. We have considered it to update the figures.

Referee comment: l228 696,000 being ‘slightly below’ “about a million” is a favourable reading of the data. By eye the total affected number is closer to 1.2m which makes the simulation “almost twice” the observed estimate.

Authors’ response: The phrasing has been changed to: *This is below our cumulative number of directly affected people by the end of the event, which amounts to 1,200,000.* in the revised manuscript.

Referee comment: Fig 4, 5 can be made bigger, perhaps put in a two panel figure

Authors’ response: We have made a two panel figure and increased the figure’s size. The updated figure can be found below.

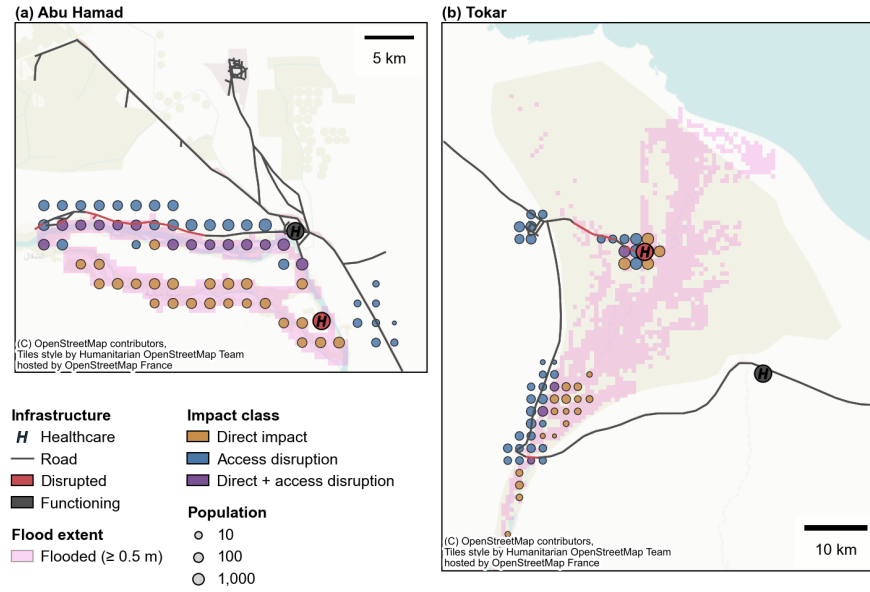


Figure 4: Forecast maps of directly impacted roads and healthcare facilities and population nodes losing access to healthcare services for the towns of Abu Hamad (a), and Tokar (b). The forecast shows the impacts for the maximum flood extent over a lead-time of 10 days for the ensemble member showing the 90th most extreme event from the forecast ensemble initiated on 1 August 2024 in Abu Hamad, and 20 August 2024 in Tokar. Pink shading show location for which a flood depth higher than 50cm is forecast. Disrupted infrastructures are colored in red. Population nodes are represented as dots with sizes proportional to the population counts and are colored according to the forecasted impacts; grey no impact, orange direct impacts, blue disrupted access to healthcare, and purple direct impacts and healthcare access disruption.

Referee comment: Fig 6b,d includes “additional indirect impacts” in the legend but no data on the plot, please check.

Authors’ response: This shows that the additional indirect impacts do not change the probability of exceeding the thresholds. On panel d), it can be seen that the inclusion of indirect impacts modifies the probabilities of exceedance by a small fraction, around 18 July 2024 and 26 August 2024.

Referee comment: Section 3.2 The result of large underestimate of actual impacts for Somalia chimes with experience of poor GloFAS performance as noted in the paper. In fact this was specifically due to GloFAS v4 which entirely failed to flooding on the Shabelle at peak flood time. Post-event analysis by the GloFAS team showed this was not a ‘forecast bust’ (i.e. the heavy rainfall in the basin was correctly anticipated) but was somehow related to the soil simulation in the model - indeed even forced by real rainfall observations it would fail to produce a flood peak for 2023

Deyr and even the record-breaking 1997 Deyr. Unfortunately I do not know of a reference where this has been documented, and as far as I am aware the problem is still unfixed. No particular suggestion for an edit here as there’s nothing citable, but just an FYI.

Authors’ response: Thank you very much for the detailed explanation. Unfortunately, we also did not find a reference to include this in the publication. We hope that this issue will be fixed in the next version of GloFAS.

Referee comment: Figure 9 might be visually clarified: currently the different bars are distinguished both by labels (Access loss, access uninterrupted, no base access) as well as encoded in colour. To make it clearer just one of these might be chosen. Also, including region names along with the sub-figure labels would be helpful.

Authors’ response: The figure has been updated including these recommendations in the revised manuscript.

Referee comment: l425 “impact forecasts are able to provide a clearer predictive signal” see main comment above. Is it really ‘clearer’ in the sense of true discrimination (which you have not technically shown, having only included one flooding season per region) or, is it just “a larger probability” which scales with the strength of the discharge forecast signal. The latter would end up issuing high probabilities frequently, which does not in fact clarify the signal.

Authors’ response: This has been clarified in the revised manuscript (see previous response to the major comment).

References

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS - global ensemble streamflow forecasting and flood early warning, *Hydrology and Earth System Sciences*, 17, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>, 2013.
- Baugh, C., Prudhomme, C., D’Angelo, C., Dottori, F., Neal, J., Colanese, J., and Salamon, P.: Global river flood hazard maps, <https://doi.org/10.2905/JRC.VD32YWG>, 2024.
- Copernicus Emergency Management Service CEMS and European Centre for Medium-Range Weather Forecasts ECMWF: GloFAS v4.1, <https://confluence.ecmwf.int/spaces/CEMS/pages/473828729/GloFAS+v4.1>, 2025.
- Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., and Feyen, L.: Development and evaluation of a framework for global flood hazard mapping, *Advances in Water Resources*, 94, 87–102, <https://doi.org/10.1016/j.advwatres.2016.05.002>, 2016.

OpenStreetMap contributors: Road Classifications Guide, https://wiki.openstreetmap.org/wiki/Kaart:_Road_Classifications_Guide, 2026.

Riedel, L., Rösli, T., Vogt, T., and Bresch, D. N.: Fluvial flood inundation and socio-economic impact model based on open data, *Geoscientific Model Development*, 17, 5291–5308, <https://doi.org/10.5194/gmd-17-5291-2024>, 2024.

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., and Ward, P. J.: FLOPROS: an evolving global database of flood protection standards, *Natural Hazards and Earth System Sciences*, 16, 1049–1061, <https://doi.org/10.5194/nhess-16-1049-2016>, 2016.