Savona, 26/6/2023

MS No.: egusphere-2023-804

Title: Impact-based flood forecasting in the Greater Horn of Africa

Reply to Referee #1

GENERAL COMMENTS

The paper is well written, well-structured and clear. The topic is surely of interest for the readers of Natural Hazard and Earth System Sciences (NHESS) as the paper describes an important effort to develop a large-scale flood forecasting system in an African region. The authors made a great job in developing the system and I believe the paper deserves to be published.
However, I have some major comments that, in my opinion, need to be addressed before the publication.

Reply: We thank the reviewer for the careful reading and the constructive comments which helped us improve the article. We do not disagree with any of the reviewer's comments so, the vast majority of those will result in an addition to the text, a clarification or a change. Our reply to each comment is shown below, interspersed with the reviewer's comments. We noted that a revised manuscript version is not yet required at this stage, hence to make replies clearer we include below portions of modified/updated sentences that will be used in the revised manuscript version.

MAJOR COMMENTS

The development of the system has required a number of choices with respect to input data, meteorological forecasts, and hydrological modelling. The paper only describes the system currently running without considering possible alternatives. For instance, why satellite precipitation from GSMaP? Why the GFS forecasting system? Have the authors investigated alternative options? I believe that a discussion on the decisions made to develop the system is needed.

Reply: We understand the reviewer's point that further clarifications and motivations must be included in the article to justify the use of the main tools and data. In the revised version we will work to include those details, so that all such choices are motivated and follow a coherent reasoning.
Regarding the hydrological model, we will clarify that Continuum has been developed at CIMA Foundation over the past 20 years and has already been implemented in several research applications and in operational forecasting chains. Having an in-house model, is very important for full system customization, and to learn from past experiences of implementation in different geographical configurations.
Regarding the use of GSMaP and ERA5, we will clarify that those datasets were chosen following a set of criteria driven by the operational nature of the system to build: 1) real-time production and release with minimal latency (a few hours at most); 2) availability of a historical dataset to maximize the coherence between the operational runs and the past data and related warning thresholds; 3) use of free products, to enable system continuity after the project completion; 4) data availability over the entire focus region with spatial and temporal resolution relevant for the desired application; 5) skillful performance in the simulation region.
Regarding the use of GFS and GEFS forecasts we will add that those products were chosen as they are freely available at the original resolution and with short latency for operational implementation, as well as the historical archive of past forecasts from 2015 onwards. This is important for simulating events occurred before the start of the operational forecasts, such as the Sudan floods case study in summer 2020. At the time of the start of the system implementation this

was the main choice available with regard to operational weather forecasts. More recently, some global centers started to share a limited set of forecast products for non-commercial uses, hence we are considering if some of those can be added to turn the system into a multi-model approach. However, this needs to be considered carefully as some centers such as ECMWF (known for producing skillful forecasts) releases a spatially aggregated product (at 0.4° resolution) and with a delayed release schedule.

It is not clear how the system works in real time. If I understand correctly, the hydrological model is run every day with last day satellite precipitation from GSMaP (1 day behind now) and 5-day GFS forecast. But in the text it reads ERA5 is used. Presumably the model is run every day starting from N-days before the "now" and ERA5 is used until it is available. Something it reads at the beginning of section 2.5.1, but it seems that ERA5 is not used at all. However, this is not specified in the text and it should be clarified.

Reply: The reviewer's comment is very pertinent, hence we will work to improve and clarify the operational methodology in Sect. 2.5.1. We will specify that hydrological states are updated to the 00 UTC of the current day through a 1-day run starting from the previous day conditions and taking as input the GSMaP 24-hour precipitation and the other atmospheric variables of the last 24 hours from the GFS forecast run of the day before. Such filling with 1-day forecast data is performed on average over the last 5 days, due to the latency of ERA5 data.

The criteria used for parameter regionalization should be specified.

Reply: We will clarify that parameter regionalization was performed on those domains with no calibration stations, according to criteria of proximity and climatic conditions, i.e., where parameter sets are taken from the closest calibrated donor domains with the same dominant Köppen-Geiger climate class taken from Beck et al. (2018).

Did the authors check the agreement between ERA5, GSMaP and GFS precipitation data? It is a very important and critical aspect in the development of a flood forecasting systems.

Reply: ERA5 precipitation was not used in this work, as clarified in the reply above. Regarding the choice of the precipitation product for model update and for forecasts, it has been motivated in the reply to another question above. Indeed, being for operational implementation we had to take a decision also conditioned by other factors, although we acknowledge that there may be discrepancy between the statistics of the two datasets. This is inevitable, given that one (GSMAP) is a remote sensing product while the other (GFS) is the output of a forecast model.

The impact assessment is carried out by defining several indices. However, it is not clear how the indices are calculated and how they are integrated. I assume that normalised indices have been calculated, but this should be clarified.

Reply: This part will be improved to clarify what the system produces with regard to impact data. Units will be included for all terms of the equations (1) and (2). I_{AR} *is the potential impact for any considered administrative region (AR) and have the same units of the considered exposure category. It is obtained as a double summation over all pixels within AR and over each of the three considered hazard classes (Hc), where Lcc is a constant value for each country and the product (H E V) is computed at the pixel level for each Hc and then added to the sum.* RI_{AR} *is calculated as the ratio between* I_{AR} *and the total amount of each exposure class in each administrative region, hence it is a dimensionless number ranging between 0 and 1. Impact classes are not integrated among each other, hence each impact class can be visualized individually in the myDewetra interface. Some details will be added on what myDewetra is, clarifying that results of flood impacts are* *displayed in the myDewetra geospatial visualization web platform (https://www.mydewetra.world/), developed by CIMA Foundation to support forecasters and decision makers in hazard monitoring, early warning as well as during emergencies.*

The authors say that correlation is a suitable indicator to measure the model capability to detect flood events and it is good if threshold exceedances have to be assessed. I would agree, but it should be shown in the paper. Is the model able to detect flood event correctly in terms of threshold exceedances? A dedicated paragraph should be written on this point.

Reply: We will add a reference to Alfieri et al. (2013), already cited in the paper, but useful to support this statement. In addition, we will clarify that correlation is sensitive to even a few outlying data pairs, thus highlighting significant shifts between the timing of simulated and observed flow peaks (Wilks, 2006).

SPECIFIC COMMENTS (L: line or lines)

L161: “Alfieri et al. (2022a) is missing in the references list.
Reply: We thank the reviewer for spotting it. Full reference will be added in the reference list.

L181: GEFS is not defined, please check all the acronyms.
Reply: We will add the full name of the GEFS acronym, i.e., Global Ensemble Forecast System.

L248: It is not clear how many stations are used for calibration and how many for validation.
Reply: Here we will clarify that validation was performed over 78 river gauges, hence including 22 stations in addition to the 56 used in the calibration phase.

L269: The Supplemental Material should be cited more clearly, which figure exactly? Which paragraph?
Reply: In the revised version we will be more specific whenever citing content of the Supplement. In the example raised by the reviewer it will be changed to “see example in the Supplement material, Figure S6”.

L325: It would be interesting to show stations located downstream large reservoirs to assess the reservoir impact.
Reply: Upon the reviewer’s comment we have made some tests on adding the reservoir location in maps of Figure 2, but the result creates confusion. We think it is a better solution to refer the readers to the map of Figure 1 to assess the position of lakes and reservoirs, which is directly compared to the stations location. Sentence in line 325 will be modified accordingly.

Figure 3: The figure is too small and hardly readable. Moreover, the stations shown in the figure should be highlighted in the map. The last panel (bottom right) shows a strange behaviour of river discharge; is there any explanation for that?

Reply: In the revised version we will add a label to each panel and show the corresponding label in one of the maps, or in an additional map if it turns out more readable. The idea behind this figure is to show a general comparison between observed and simulated flow for the entire validation period and representing all modeled domains. By zooming into the pdf the readability improves, though to fully capture the differences over specific events we would need to plot shorter portions of the time series, which is different from the initial aim. Anyways, if the reviewer considers it an important feature we can produce html versions of these graphs to add as supplementary files, so that the interested readers can zoom into specific events interactively. The increasing trend of the flow in the last panel is caused by the increasing levels of the big lakes along and upstream the White and

Victoria Nile in the late 2010s-early 2020s (see http://www.fao.org/3/cc0474en/cc0474en.pdf also cited in the paper as FAO and WFP (2022)), which caused increased flows in the White Nile and persistent flooding in a large portion of South Sudan (where the Mongalla station is located). Unfortunately, no observed flow data was available in the recent years for a more extensive quantitative evaluation.

L364: Do the authors have an estimation of peak river discharge? Can the authors make a comparison between observed and modelled peak discharge?
Reply: The station Blue Nile at Khartoum is one of those used in calibration and validation, though observed flows end in 2016, as shown in the Figure S7 of the Supplement, hence does not include estimates for the 2020 event. This will be added to the text for clarification. In addition, as stated in Sect. 3.2 "The 20-year hydrological reanalysis forced by GSMaP satellite precipitation correctly identifies the flow peak of September 2020 in the lower Blue Nile as the largest in the available simulation record"