Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-122-RC1, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Are Kenya Meteorological Department heavy rainfall advisories useful for forecast-based early action and early preparedness?" *by* David MacLeod et al.

Anonymous Referee #1

Received and published: 1 July 2020

Referee comment

Review of "Are Kenya Meteorological Department heavy rainfall advisories useful for forecast-based early action and early preparedness?"

Authors: David MacLeod et al.

MS No.: nhess-2020-122

General comments

Printer-friendly version

Discussion paper

This manuscript provides an evaluation of 33 Heavy Rainfall Advisories (HRA) issued by the Kenya Meteorological Department (KMD) during 2015-2019. This analysis is potentially useful for forecasters, practitioners, and decision makers concerned with the prediction of natural hazards and communication of warnings for early action in the region. Since it is essential to evaluate the skill of operational early warning systems, the first assessment of these advisory warnings for Kenya reported in this manuscript is of great practical value for the community. Such an analysis has the potential to provide some evidence-based recommendations for future improvements of similar heavy rainfall advisories in Kenya and in similar contexts.

In general, the article is quite well written and easy to read, even if some specific parts should be improved by making the text clearer, providing some more context and motivations, giving some important details or references to support some statements/assumptions (see major and specific comments below).

However, the manuscript needs some major revisions: the authors should make more efforts in terms of analysis to address the questions posed here more thoroughly, improving some methods to provide more quantitative elements on whether these HRAs are useful and how they could be improved, but also discussing more thoroughly the current barriers and limitations of the HRAs (especially the spatial detail issue, see comments below) and how these fit within the context of current and future co-production efforts. Some justifications used to support a qualitative (or proxy-based) analysis are not convincing. The proxy/qualitative indicators that are used to answer two central questions in the article (questions 1 and 2, see page 5) can provide only partial insights and non-robust indications on the usefulness of the advisories (given unrealistic or not convincing assumptions on relevant trigger probabilities and population exposed to flooding). So far, some parts of your analysis cannot convincingly support a few central points of your conclusions. Thus, major revisions are recommended.

Major comments

1. The first major concern is that to estimate the relative scale of preparedness implied by each advisory, the population exposed and vulnerable to heavy rainfall and consequent impacts (flash floods, water-logging or riverine floods) should be used instead of the total population for each warned county. The total population living in a warned area seems an oversimplified and unrealistic proxy indicator, that does not provide a measure of the number of people likely to benefit from flood preparedness actions in the region and does not allow a comparison of the extent of preparedness action required between advisories. The authors partly recognize this issue, but do not address it properly and do not convince the reader on the value of their 'first-guess' estimates. A proxy estimate based on the total population per county does not seem a sensible approach even to provide a broad indication of the relative amount of preparedness appropriate for each advisory (see L. 170-174). The broad indication derived from this proxy could deliver the wrong message (or maybe the right message but for the wrong reasons), being based on assumptions that are not necessarily true (i.e. there is some correlation between total population of a county and vulnerable/exposed population to heavy rainfall per county, but the distribution of vulnerable population across counties may not match the distribution of total population).

Related to this, it should be also acknowledged that although there is an attempt at overlaying population density and rainfall accumulation observed over each advisory window (L. 262-265 and Figure 6b), in many cases the population living in an area receiving heavy rainfall does not coincide with the population potentially affected by flooding, especially for riverine flooding events. Please consider using some additional datasets of population potentially affected by flooding. For example, you might want to use some datasets that exist to better estimate potentially affected population by flooding at least in some areas, based on data available for past events in some regions of Kenya, at least to provide a case-study example of the extent of preparedness implied by a single or few advisories, e.g. for eastern Kenya you could use the

dataset available in the OCHA's Humanitarian Data Exchange (HDX) platform, based on Sentinel-1 imagery acquired on 2018 : https://data.humdata.org/dataset/potentially-affected-population-by-flooding-in-eastern-kenya-2804

2. A second major concern is that triggers should be defined more realistically, based on some relevant rainfall thresholds and effective probability triggers. You argue that you may avoid considering any specific rainfall threshold or probability because it would not 'provide robust statistics and precludes any meaningful statement' (e.g. see lines 138-140). However, while I agree that the sample size and the inconsistencies of the data from these 33 HRAs preclude any meaningful calculation of some verification metrics such as the reliability of probabilities, I think that the available probabilities and rainfall data could still be used to answer the questions in your paper more quantitatively. In other words, I agree that you cannot compare warnings with different levels of spatial aggregation, different temporal windows for accumulation of rainfall, etc., but you can still test whether the HRAs were useful overall for forecast-based early action based on some minimum quantitative analysis. For example, you can set a minimum rainfall threshold (maybe dependent on the window of accumulation, or a minimum with a larger window) and some significant probability based on the classes available, e.g. probability of heavy rainfall > 33% (and not just above zero). The main problem I see in your analysis is that you have defined the action trigger as the probability (of heavy rainfall) exceeding zero, but an action trigger with a very low probability of unspecified heavy rainfall level seems a very unrealistic trigger even for low-regret actions. Such an approach is likely to lead to overconfident verification results on the value of the HRAs. For example, would a probability lower than 10% of heavy rainfall expected to fall over a big county still lead to any concrete action by government or humanitarian agencies? If you have valid reasons to think so and use a very low trigger probability, you should at least extend the discussion on this point to convince the reader of the validity of this analysis. Maybe you could support this choice based on some literature, or any reported practice in the humanitarian sector, explaining how this would be useful and what actions would be informed – otherwise all the analysis and conclusions seem

to be based on unrealistic assumptions. Still, I believe that the analysis would be more valuable if you could show the performance of the warnings for significant probability levels and considering some meaningful rainfall thresholds that are more likely to be used as triggers.

If you defined triggers in a specific and realistic way, this would allow to make your analysis more concrete and link it to some specific actions to determine the extent to which the KMD HRAs could guide 'worthy' preparedness activity. Your definition of 'worthy' action seems being kept purposely vague (in line with a zero-probability trigger) and not clearly defined, referring to any preparedness assistance and no particular action (e.g. line 181). Using some specific examples of actions and trying to quantify whether taking these actions would be worth would be a natural step forward to give more concrete value to your analysis. Please consider including some specific action-based analysis or examples that could give more value to the article.

3. There is a lack of evaluation of misses (missed events in the warnings) or at least a discussion on it: the evaluation of observed events is only based on seven reported impactful events (the most significant floods events in the EM-DAT dataset), and there is no proper evaluation of 'misses' and 'hit rates' based on a large sample, which is a limitation of the data and period available. Despite the obvious sample size issues, it would be probably possible to include in the analysis in Section 3.1/Section 3.3 some more information on observed events also based on other data (not only EM-DAT).

Are there any other significant flood events beyond the 7 events from EM-DAT (e.g. maybe events with less than 10 fatalities but still high number of affected people / households affected or damaged) that have been missed by the HRAs during the study period? Please discuss this.

If more events were available (beyond EM-DAT), in Section 3.1 you could include an evaluation of hit rates per county using CHIRPS as reference and/or in Section 3.3 a proper evaluation of misses based on the larger sample of reported impactful events.

Printer-friendly version

Discussion paper

Without a full evaluation of hits and misses, the 'hits' picture that is given might be misleading and uncomplete. You mentioned some misses because related to the time windows of advisories issued (e.g. for advisories warning "wrong" counties, see lines 233-234). It would be useful for decision makers if you could calculate a proper hit-rate even if based on a sample of 33 advisories. You could focus on a specific trigger probability and threshold rainfall, for example you could keep the 50 mm nominal threshold case and use a specific probability threshold. Of course, using only a single rainfall threshold across a big country as Kenya is not a proper location-specific indicator of flood impacts, but this would be still useful. Additional analysis with some more observed events (if you had more than these 7) would probably help understand also whether the step change in the advisories in 2017 (access to GHM) reduced the number of misses, as it seems the case from your analysis based on 7 events (Figure 7) and might be expected from the increasing number of advisories per year (Fig. 2a). Section 3.3 could then be more complete by focusing on both hits and misses, by reporting how many observed events were not preceded by advisories.

4. The analysis provides useful insights on the possible limitations of the HRAs and reccomendations, but the discussion should make more efforts in understanding and explaining the current limitations of the HRAs. This is essential to provide more specific recommendations for improvement of the HRAs. One of the major limitations of the HRAs that arises from the analysis is the lack of precise spatial information in the warnings beyond the county-level information. As you suggest, free-shape warning areas should be used instead of administrative county boundaries. However, you also explain that such free-shape warning polygons are currently generated by the GHM and are already in use at KMD. Thus, "KRCS could then overlay these with maps of population exposure and vulnerability to flood risk, in order to further narrow down targets for intervention". Then why more precise warnings are not issued yet?

It is not clear what is the current bottleneck for providing more spatial detail in the advisories, and it would be important to understand whether there are either scientific

/ technical or institutional / economical barriers that prevent this, e.g. either whether it's a lack of resources (e.g. GIS technicians) at KMD, or whether there has not been enough co-production effort made so far to enable a full use of the GHM information at KMD, or whether there are some information layers missing (e.g. flood extent maps which do not coincide with heavy rainfall extent). Without an extensive discussion on this point, it is not clear how the spatial detail in the advisories could be improved.

5. Finally, it would be essential to discuss in this article how the HRAs fit within the need "for strengthened coproduction of forecast information and products" (now widely recognised as you also recognise). Is there any issue of national ownership in the use of GHMs from the UK MetOffice in the HRAs?

From your article, it seems that only the subjective interpretation of the forecasts and the writing of the HRAs summary is carried out within KMD and so within the mandated agency in Kenya. How is this perceived at the national level? I can see that the actual sources of forecast information in the HRAs are not mentioned in the HRAs (see Figure 1, no field on 'data sources') so maybe there is no general perception from the communities involved in Kenya or even from county directors / national policy makers around that issue. However, this point would need attention in such a paper. Also, it would be useful to detail whether KMD get access to raw forecast data from ECMWF and UK MetOffice or only to some end-product maps and the level of spatial detail in these maps (e.g. do KMD get any shapefiles or netcdfs data? At which resolution?). This point might help explain one of the major current limitations (lack of spatial detail in the HRAs) and how KMD and their international partners could deal with it to improve the advisories.

Specific comments

- L. 34: it would be good to specify how this UK-funded project fits within the local context and is linked with other projects mentioned (IARP) or not; are these projects

making some efforts in coproduction and capacity building and how specifically (only by giving the outputs of GHMs models to local agencies or is there something more, e.g. capacity building efforts)? It would sound very sensible to discuss this.

- L. 45: do KMD work on their own on hydro-meteorological forecasting? The mandates and institutions involved in hydrological warnings in Kenya should be clarified, as KMD is the meteorological agency, and there is also a national hydrological agency, the Water Resources Authority (WRA, https://wra.go.ke/) that should be responsible of flood forecasting activities alongside KMD (e.g. see FLOOD ADVISORIES issued by WRA; see also ODI working paper 553, April 2019, "Reducing flood impacts through forecast-based action" by Lena Weingärtner et al.). The links between KMD and WRA are not clear nor mentioned in the paper. It seems an important point to discuss, as probably a closer collaboration with forecasters at WRA could help make the HRAs by KMD more precise, with impact-based focus and hydrologically meaningful. Is there a link between the KMD Flood Forecasting Unit and WRA? Or is this a current institutional barrier? Hydrological forecasting is at the interface between met and hydro agencies not only in Kenya but in many countries and similar questions may arise elsewhere. Some more context about this important point should be provided.

- L. 101: I would suggest specifying "The GHM system", to avoid confusion with other possible meanings, e.g. HRA or other systems just mentioned in the text

- Section 2.2 (Verification Approach): L. 150-219 are difficult to follow and should be reorganised in a more clear or compact way (e.g. with bullet points for all the methods and data associated to each question).

- L. 200-206: the part about the dam collapse of May 2018 seems excessively long in the context of this section and should be kept more concise; as it stands, that part does not flow well within the paper.

- L. 243-244: "As these advisories associate each warning with a probability, these findings are quite consistent" – this sentence is obvious and not specific enough. You

could try to add something more relevant and specific, as the previous remarks in terms of convective rainfall and percentage of warned area are interesting. For example, could you see visually any link between percentages of area warned which receive rainfall accumulation above the 50mm threshold and geographical locations/counties that are known to be more subject to convective rainfall?

- L. 347: “These kinds of actions would have significant costs, so more than ten triggers in a year may not be realistic” – that sounds reasonable but too approximate to be stated in this way, could you provide more details (e.g. approximate estimate of costs and resources) and improve the sentence? Are there any references supporting this sentence (and the number of ten triggers)? Are there any estimates in the scientific/economic literature or in reports of humanitarian agencies on the resources that would be needed / are available for early action and flood preparedness in Kenya or maybe in any specific region within the country?

- L. 405: “For this purpose they are effective” sounds a bit overstated, e.g. given the lack of spatial precision that you highlighted in the paper, the large area warned by identifying counties in the text may not be effective (people in an affected county may not take county-scale warnings so seriously, if these are preceded by warnings that were not followed by any event in their specific area in that warned county).

- "Data and verification approach" section: some final parts do not flow well and could be improved (e.g. more clear organisation by points and questions addressed); some parts need more details or references to the scientific literature or humanitarian reports to back-up some assumptions made (e.g. see also remarks in major points above).

- "Discussion and recommendations" section: there is a lack of discussion on the temporal consistency in the HRA dataset. There is a difference in the source rainfall forecast in the new data in recent years, but also the number of HRA has increased. So, what role does this inconsistency play in comparing earlier years with more recent ones? For example, the number of hits in recent years is expected to be higher simply

because there are more warnings issued.

Technical corrections

- Table 1, column 3 - “Period length” is missing the units (days, probably)? - L. 2 (Abstract): “Forecast-based Action/Finance (FbA)”, it seems that Forecast-based Financing is more used than Finance, please check. - L. 19: Climate risk or better hydro-meteorological risk? - L. 27: repetition to avoid in (see Wilkinson, for. . .) as L. 23 already mentioned it - L. 29: ‘individual expenses’ and/or probably even more ‘community expenses’? - L. 164-165: it’s fine to focus the discussion on results for 50mm accumulation, but maybe you can say here more explicitly that you took this threshold as “working definition for heavy rainfall”, as sometimes later in the results section this is the wording you use (e.g. Line 246), so good to define it clearly from the methods, still mentioning the limitations of it as you do. - L. 167: “To answer question .. we estimate” is missing the question number - L. 399-400: see sentence “We find that an increase in skill over time, and that. . .”, to be corrected. - L. 438: higher-cost actions - L. 456: repetition, "would would" - L. 470-471: repetition of “in Kenya” - L. 484: “mitigating the risk from risks”, I would avoid the repetition - Figure 4a, caption: please clarify whether by “inner and outer quartiles” you mean “inner and outer fences” (which seems more common wording in this context)? – see "(dark/light shading shows inner/outer quartiles and dot indicates the median". (by the way, a parenthesis is missing there).

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-122, 2020.

Printer-friendly version

Discussion paper