General comments

RC2 GC1. The paper provides the use of a hydrologic-hydralic modeling framework, exploiting a fully shallow water equation solutor (implemented in the IBER software) combined with a SCS-CN runoff model in an ungauged basin in Mozambique which was hit by a severe flood in February 2023. The application is performed by using freely available source of information, i.e. global datasets for DEM, Land cover, CN parameters and precipitation. The basin also includes a large artificial reservoir so that the extent of the flood is affected by spillway regulation whose behavior is analyzed. The paper is well written and organized and the topic is very relevant with both flood management in areas provided with scarce hydrologic/hydraulic monitoring and also with respect to reservoir management.

As a general comment, I believe that the paper results are interesting and satisfactory with respect to an application developed in a data-scarce environment, nevertheless I would suggest the authors putting maybe less emphasis on the goodness of results achieved and more on the uncertainties related to them, for the reasons explained in following points.

We thank the reviewer for his positive assessment of our work.

RC2 GC2. Also, the results of the study case are satisfactory, as just said, but not necessarily easy to straightforwardly address any data-scarce basin in the world. The methodology is, as they say, "reproducible anywhere" but, I would say, not necessarily with the same rate of success. The success of the application is probably also due to the particular study case which is performed in a (I would say) "enough large" basin (almost 5,500 km^2), with a very low rate of anthropical effect (built area is less than 2%) and a homogeneous topography namely, "a very flat topography".

We agree with the reviewer in the fact that the methodology is reproducible anywhere, but not necessarily with the same rate of success. The agreement with reality will depend on the factors mentioned by the reviewer. The methodology will probably perform better in events with high rainfall rates and a wet antecedent soil moisture content. We have clarified this in the conclusions of the revised manuscript in order to avoid misleading the potencial readers.

RC2 GC3. Moreover, the model validation was possible due to the availability of the Sentinel 1 image of flood extent and by ground observations of maximum water levels in a number of points for the flooded area. To my personal experience, both of these datasets are not always easy to find even in Europe or many other areas of the world. That said, the effort they have provided in finding global databases useful for the application is really noteworthy and also, I would agree that the kind of information they have used can be suggested for best practices of "after event" flood management in any area of the world.

It is true that the datasets used for validation are not easy to find, specially in African countries. But those are not needed to implement the proposed methodology. We have just taken advantage of their availability on this event in order to analyse and validate the proposed methodology.

Major comments

RC2 #1. The authors provide three scenarios, MS1 (actual management of the reservoir), MS2 (absence of reservoir), MS3 (what if the reservoir was able to retain the entire flow of its upstream basin). The three scenarios provide interesting insights. The comparison between MS1 and MS2 provides that the presence of the dam had a beneficial effect on the flood extent and depth. The third one also testifies that even if the reservoir was larger the AOI would have been flooded by the second tributary (D-PLD subbasin), nevertheless I notice that in this third case the average water level (1.6 m) is significantly lower than in the MS2 (and actual scenario (2.1 m). I agree that the amount of damage could have been not so different, nevertheless a water level less than the average human height could make a significant difference when life is threatened and has to be saved. Hence, I think it could be of interest to know what would have been the flood extent and (average and maximum) depth if, at the initial condition of the event, the reservoir was empty or if it was at a different level below the NPL, see comment #2 below, in order to see if it could be beneficial or not influential at all (as it probably is, due to its small capacity with respect to flood volume).

We understand the reviewer's suggestion and in fact, we thought a lot about the most convenient scenarios to analyse. We finally decided to include MS2 and MS3 as two limit scenarios on the potential effect of a reservoir on the natural flood regime in Boane, MS2 representing no effect at all, and MS3 representing the maximum effect. We are aware that other intermediate scenarios are possible, and we thought a lot about how these could be defined. The problem is that almost all the reservoir's volume can be controlled by its hydraulic structures, since the dam has bottom outlets, and the crest of the controlled spillways is at an elevation of 24 m, for which the reservoir's volume is of just 8 hm3. Therefore, considering an empty reservoir or an elevation equal to the crest of the spillways is virtually the same for the analysis of this flood event.

So it is complicated to establish a level of the reservoir that could represent a plausible initial condition. Even if such a level could be established (considering for instance an empty reservoir, as suggested by the reviewer), the outlet hydrograph would depend on the way in which the spillway's gates were operated. Moreover, we don't have detailed information about the hydraulic characteristics of the controlled spillways. Under this situation the amount of potential scenarios is huge, considering the different combinations of initial volume in the reservoir and outlet hydrographs.

That's the reason why we decided to keep only the three scenarios included in the manuscript. In our opinion adding new intermediate scenarios wouldn't add too much to the discussion about the role of the dam, and at the same time it would lengthen the content of the paper.

RC2 #2. Figure 7 suggests the need for more information about the operational and the structure of the dam spillway system. In facts, from the figure it appears that in the first four days of observation (6 to 9 February), for water levels up to almost 46.5 m, the daily outflow Qout is zero while, on the 12 and 15 of February for water levels well below 46.5, there is a daily outflow above 500 m^3/s. This observation suggests that the spillways are regulated by some movable device or other hydraulic system that probably were operated (manually or automatically) during the flood event. This is not clearly stated but I believe is necessary for a discussion about reservoir management. The paper does not provide detailed information about the structural and hydraulic operational system for water release. Also there is not a definition of the "Normal Pool Level" (NPL). In particular, it would be of interest to know if there is a minimum level for water release (below NPL) which could be operated by means of such a movable device and, if yes, what is the reservoir volume at that level. These elements could be useful to define a fourth scenario as I have suggested at comment #1

As the reviewer has correctly inferred from Figure 7, the discharge flowing through the spillways and outlets of the PL dam are controlled by movable gates (as we have already mentioned in our answer to comment **RC2 #1**). We have clarified this in the revised manuscript.

Unfortunately, we don't have any detailed information about the hydraulic behaviour of the controled spillways, but even if we had it we would need to know how they were operated in order to derive the outlet discharge as a function of the reservoir's level during the flood event. In fact, for our study the most relevant data are the discharges that were spilled by the dam during the event of February 2023. Those discharges were established by the regional water administration (ARA-Sul) according to the operational rules of the reservoir.

The only information about the discharge structures is that the crest of the spillways is at an elevation of 24 m (the reservoir's volume for that level is 8 hm3), and that bottom outlets can completely empty the reservoir for practical purposes. But we cannot relate the outlet discharge with the reservoir's elevation.

In summary, the data that was available to us for this study were the discharges that were actually spilled by the dam during the event of February 2023, which are those plotted in Figure 7.

The NPL is the maximum operation level, which corresponds to the maximum water level that can be attained during normal operation conditions (i.e. when there is no flooding).

RC2 #3. In section 3.3.3, line 304, a F1 score is mentioned as a combination of HR and FAR, but it is not further defined neither it is used throughout the paper. Also False Negative (FN) cells are mentioned but, if I am not wrong, there is not focus on them in the result sections. I may suggest the use of other indices such as the Critical Succes Index (CSI) and more. Maybe the authors could refine this section by extending the use of these metrics to other indices or explaining the reason why they only focused on HR and FAR.

The mention to F1 was a mistake, and it was removed from the revised manuscript. FN are mentioned because they are used to compute the HR in Equation (4).

Following the reviewer's suggetion we have included the CSI in the revised manuscript. We believe that these three complementary indices, together with Figure 12, are enough to discuss the agreement between the flood extent predicted by the model and derived from Sentinel-1.

RC2 #4. Figure 9 (right) I would add, besides (or replacing) the regression line, the 1-1 line. The regression line, in facts, provide a satisfactory index of determination but suggests a systematic underestimation of the hydrologic/hydraulic model with respect to observation. By this light I don't think the regression line provides a correct information. I see that all points but one are almost perfectly centered. Only in one day the daily average discharge is missed (11 February). This could be a lack of the measured precipitation which is almost absent on that day. To my knowledge CHIRPS values of precipitation may have a high rate of uncertainty and also the CN hydrological model used for evaluating infiltration is rather than perfect.

We agree with the reviewer's comment. In this case the 1-1 line is more relevant than the regression line. We have therefore replaced the regression line by the 1-1 line in Figure 9.

RC2 #5. Section 4.1.1. At line 367-369 it is stated that "the positive ME means that the numerical predictions of the maximum water surface elevation have a positive bias with regard to the field estimations, which is coherent with the fact that the water marks identified in the field work represent a minimum threshold reached by the flood". But at line 280 it is stated "At each point identified, the maximum water depth reached during the flood was estimated". The authors should clarify whether the points were related to minimum or maximum levels of water. I believe they are maximum levels as it would not make sense to perform a field map of minimum water levels reached by water. I would suggest that the positive bias may be due to a number of different explanations not excluded the hydrological model used for runoff generation. It is well known that the CN method may provide overestimation of both volume and rate of runoff. The upper left portion of Figure 10 provides a shaded area of runoff which is practically all over the D-PLD sub-basin, independently of rainfall intensity which looks quite low in some areas of the sub-basin. Even the rate of infiltration in Figure 9 (left) looks low, even considering the possible underestimation of rainfall which CHIRPS may provide as already said in comment #4. On the other hand, the high value of the vertical accuracy of the Copernicus DEM (RMSE= 1,7 m) is not good news, considering it is of the same order of magnitude of the average water depth (2.1 m in MS1, 2.9 m in MS2 and 1.6 m in MS3 as from table 6).

We acknowledge the reviewer for this interesting discussion. Regarding the first part of his comment, we have realized that the statements included at lines 280 and 367-369 might be misleading for the reader. The water depths given in the manuscript are certainly related to the maximum levels of water, since they were estimated from the water marks left by the flood. What we wanted to state in lines 367-369 is that these flood marks might underestimate the real maximum level reached by the water, since the fact that there is a mark means that the water reached that level, but it might have reached a slightly higher level without leaving a significant mark. We have tried to better express this idea in the revised manuscript.

Nevertheless, there might be other alternative explanations to the positive ME. In particular, as the reviewer suggests, in any hydrological model there is always a significant uncertainty related to the estimation of the infiltration parameters (in our case the CN). This is especially true if the results were obtained without any model calibration, as in this study. The associated error might be either positive or negative, but it might well be the case that in our simulations the error is positive considering that we have chosen to use a CN associated to wet AMC, since the 5-day antecedent rainfall depth in the basin was slightly greater than 50 mm (as stated in line 183). Considering the reviewer's comment, we have included this potential explanation in the revised manuscript.

The infiltration rate in Figure 9 is certainly low, partially due to the assumption of a wet AMC. Probably it was slightly higher in the real event. However, as stated in the manuscript, the aim of the study was to analyse the kind of predictions that can be made in data scarce regions in which there is no possibility of calibration and thus, the models must be used with the default parameters, as it was the case here.

We should also clarify that the shaded area in the upper left portin of Figure 10 does not represent the surface runoff, just the extension of the D-PLD subbasin. So it is not related to the rainfall intensity. This might have misled the reviewer, so we have clarified it in the revised manuscript.

We also agree in the fact that the vertical accuracy of the Copernicus DEM is rather high compared to the typical water depths during a flood event. But this is probably the most accurate global DEM that is currently freely available. In any case, it might also explain the differences between observed and modelled water depths, as already mentioned in the manuscript.

RC2 #6. Figure 10 (bottom). Here is probably my major concern. It shows the hydrographs computed with Iber at different locations but it seems that the MS1(S3) line is not an output of Iber but rather a linear interpolation of average daily discharges obtained by water levels registered in the reservoir (they are consistent with those shown in Figure 7red line). As a result, the MS1(S2) line here is the sum of an hourly discharge plus a daily discharge interpolated over different values which appears to me as a critical point of the paper. If my considerations are correct I think this point needs re-evaluation by the authors. If we go back to Figure 9 (left) we see that the daily discharge value (3,780 m^3/s) flowing into the reservoir obtained from the IBER output subtends a much larger hourly peak (5,700 m^3/s). That is the same point (the daily average) we find in Figure 7 as the maximum value obtained by IBER as Qin (in light blue). As I noted before we do not know anything about the spillway size and structure and about hydraulic regulation devices but even considering a very high efficiency of such a structure it is hard to believe that the ratio between maximum Qout and maximum Qin is equal to 2700/5700=0.47. In order to sum up the hourly hydrograph of the IBER output from sub-basin D-PLD with the hydrograph of the spillway discharge the hourly distribution of Qout is needed as well. It should be ideally obtained by knowing the geometry of the spillway structure, and of the lake, to route the hourly IBER output Qin of Figure 9 arising from sub-basin U-PLD into the reservoir and then into the spillway in order to obtain an hourly Qout. If such information is not available at least a peak coefficient could be applied to the daily average value shown in Figure 7, or a feasible ratio between hourly values of maximum Qin and maximum Qout should be searched for. Obviously, such consideration also affects results shown as from scenario MS1(S2) (e.g. Figures 13 and 14). It is likewise obvious that, should the authors state that the dam is provided not with a standard spillway system but with a strongly regulated discharge control system, my concern will be solved and with it also the last sentence of the next comment.

We understand the detailed analysis made by the reviewer. However, as mentioned in our answer to **RC2 #2**, the spillways are controlled by movable gates. The hydraulic behaviour of the outlets and spillways, as well as their operation and percentage of opening during the flood is unknown to us.

The only data that is available is the total discharge (m3/s) spilled by the dam during the event, which is represented in Figure 10 bottom (black line). This was the discharge imposed at the dam location in the D-PLD model, and that's the reason why the green line in Figure 10 bottom MS1(S3) has such a pattern. To further clarify, the discharge imposed in the D-PLD model was the one provided to us by the ARA-Sul as the one spilled by the dam.

At the same time, the discharge imposed seems to be coherent if we consider that the inlet discharge to the reservoir computed with Iber is higher than the outlet discharge of the dam (2,800 m3/s) for approximately 20 hours, during which the volume difference between the inlet (computed from Iber) and outlet hydrographs would be around 90 hm3. The volume of the reservoir the days prior to the event was around 330 hm3, while the maximum volume during the event was around 450 hm3. Those numbers are coherent.

Also, if we look at Figure 12, there is an almost perfect match between the flood extent computed by Iber and the one derived from Sentinel-1 in the river reach between the dam and the confluence of Movene and Umbeluzi rivers. We should note that the discharge in this reach is mainly the one spilled by the dam (since the contribution of infiltration and rainfall might be neglected in that small area). If there was a relevant error in the discharge spilled by the dam, the agreement would probably be worse.

We understand the concerns of the reviewer, but another assumption about the discharge spilled by the dam with the available data would be purely hypothetical and not necessarily more precise than the values used here.

RC2 #7. Figure 11. In this Figure I see that the northernmost point, ID 6 in Table 3, if am not wrong is the one that provides the highest overestimation (second highest value) in water level h obtained by IBER vs h from field observation: 5.4 m (table 5) vs 2.8 m (table 3) of water depth over the ground level). If I read well Figure 8 this is also the only one placed on a reach affected only by the flood of sub-basin D-PDL. Consider now the significant FAR value (0.37) found in section 4.1.3 and look at Figure 12. I see that a good portion of the False Positive cells affecting FAR are in the same northern reach coming from sub-basin D-PDL. Considering that the discharge coming from the reservoir outflow may be affected by an underestimation error (see #6) both the high FAR value and the overestimation of water depth in point ID 6 may be an effect of the overestimation of runoff arising from the use of the CN infiltration model. Such overestimation may compensate the underestimation of the daily outflow hydrograph from the reservoir in the remaining points.

The reviewer is right in his considerations about the areas in which the highest number of False Positives occur. These are the floodplains of the Movene tributary, upstream its confluence with the Umbeluzi. It is also true that in this area the flood extent is barely affected by the discharge of the dam. Therefore, as the reviewer suggests, the overestimation of water depth (and flood extent) in this area is probably an effect of an underestimation of the infiltration rate in the model. This is very related to the reviewer's comment **RC2 #5**, so we refer also to our answer to that comment. We have introduced this consideration in the revised manuscript, in the discussion of Figure 12 and of the FAR value obtained.

On the other hand, we don't think that the daily outflow hydrograph from the reservoir is underestimated. As we mentioned in our answer to the previous comment (**RC2 #6**), the reservoir discharge was provided by the water administration ARA-Sul, and the good agreement between the flood extents estimated from Iber and from Sentinel-1 in the river reach located just downstream the dam suggests that the discharge spilled by the dam is correctly imposed in the model. Lastly, there are also several areas with False Positives after the confluence, between Boane and Matola (Figure 12). This suggests that the overestimation of the discharge of the Movene tributary is not compensated by an underestimation of the reservoir's discharge (or at least this cannot be easily inferred from Figure 12).

RC2 #8. the rainfall event that generated the flood of February 2023 was particularly severe also by the light of its spatial distribution. In facts, the presence of the highest rainfall intensity in areas close to the reservoir generated a very quick response that did not gave any possibility of operating on the reservoir by releasing water at a discharge compatible with river conveyance with the aim of providing more storage in the reservoir available to mitigate the peak flow. Nevertheless, considering the basin size and travel time of water, a different rainfall distribution may provide this operational time. I would suggest mentioning this possibility, practicable by mean of this hydrologic/hydraulic operational tool, as a discussion item for best practice in reservoir flood management.

We fully agree with the reviewer's statement. In fact, we mentioned in lines 143-148 that the spatial distribution of rainfall during this event contributed to reducing the response time of the basin, thus increasing the peak discharges flowing into the PL dam. Nonetheless, the comment made by the reviewer is more precise and very appropriate, so we have included it in the Conclusions section.

Minor comments

RC2 #9. Line 85. The CHIRPS acronym is only used in this line and it is not explained. I suggest expanding the acronym and explain in section 3.1.2 the relationship with GPM-IMERG.

Following the reviewer's suggestion, we have expanded the CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) acronym in the revised manuscript. In fact, we have only used CHIRPS to characterise the catchment in terms of average annual precipitation, because has rainfall estimations since 1980. However, it was not used as input data in the model because it only provides rainfall estimates with a maximum temporal resolution of 1 day, which is not enough for relatively short and intense events. In section 3.1.2. we describe the GPM-IMERG data set in detail because it is the one used in our study. We don't describe in detail other rainfall products because there are many others well described in the literature, and that would lengthen unnecesarily the section.

Nevertheless, considering the reviewer's suggestion we have included a brief mention to the spatial and temporal resolution of other freely available rainfall products in relation with GPM-IMERG.

RC2 #10. Figure 8. What is the shaded area in the background?

The shaded area represents the maximum flood extent for the scenario MS1, in order to have a reference about where the points are located with respect to the flooded area. We have specified it in the caption of Figure 8 in the revised manuscript.

RC2 #11. Line 426. The reference, if I am not wrong, should be to figure 14.

The reviewer is right. We have modified the reference in the revised manuscript.