The manuscript presents a numerical model for a recent flood event in Mozambique; furthermore, two counterfactual scenarios show what would have happened if a dam that was used to mitigate the flood (Pequenos Libombos dam) had been either larger or absent. While the hydrological/hydraulics model is a standard one, the efficiency of the GPU-based implementation is striking.

However, I am doubtful about the merit of the material as a scientific research paper. The introduction gives the impression that this will be a report of the flood event. As a matter of fact, even though a section entitled "case-study" will follow, most of the introduction is focused on the flood and the role of the dam. At the end of the introduction, the key statements are that (1) hydrodynamic modelling is useful for management purposes and (2) the effect of the flood would have been less if more water had been stored and vice versa, both sounding quite trivial. In order to engage a reader, the manuscript would need more. Which are the challenges of performing a study like this? Which are the distinctive features of this work compared to others? What can be transferred to practitioners and stakeholders? I was hoping that the rest of the manuscript would provide this, but it is actually limited to the report of the results of the simulations for the real scenario and the two additional ones. The only statement that gives a bit more is that a study performed using just open data can be enough reliable, and this needs to be emphasized even though it is not completely new. In summary I strongly encourage the authors to try to enhance the scientific merit of their study.

We disagree with the overall assessment of Dr. Radice regarding the interest and novelty of our work. The two key statements mentioned by Dr. Radice are certainly quite trivial, but those are not the main conclusions stated in the manuscript.

First, we are not using a standard hydrological model (as it can be a lumped or semidistributed model as HEC-HMS, or a 2D distributed hydrological model based on the kinematic or diffusive wave equations), but rather an integrated hydrologic-hydraulic modelling approach based on the 2D shallow water equations (2D-SWE) with an enhanced GPU implementation. The model computes rainfall-runoff transformation in the hillslopes and river flow simultaneously in the whole catchment (discharges, water depths and water velocities). Even if this modeling approach is starting to be used in some studies, it is relatively recent and cannot be considered as standard practice. Moreover, we are not aware of any publication in which this modelling approach is applied to a real extreme flood event in a 5,000 km2 regulated catchment, using only satellite-derived data (i.e. no calibration of empirical parameters with in-situ data), and validated with 3 different kind of data: 1) time series of observed discharges; 2) water marks left by the flood; 3) satellite imagery of the flood extent.

Second, the conclusion of the work is not simply that "the effect of the flood would have been less if more water had been stored and vice versa", as Dr. Radice states. We provide a detailed quantification of the flood control exerted by the reservoir during a flood event that is one of the largest in the last 40 years (as we mention in section 4.2, the peak discharge estimated for this event was 6,000 m3/s, while the maximum discharge registered in the Boane station was 7,250 m3/s on 30 January 1984). In the "Conclusions" and "Introduction" sections we don't give the specific numbers on the effect of the reservoir, which are given in section 4.

Third, we consider that the case study analysed in the manuscript is in itself interesting due to its magnitude (as justified above) and that it occurred very recently. Our analysis can give some additional insight on the flood hazard in an extremely vulnerable region. This is why we refer to the consequences of this event and the general exposition to flood damage in Mozambique in the introduction. Both, the methodology and the analysis of this specific event are, in our view, interesting for the community.

Regarding the specific questions raised by Dr. Radice:

Which are the distinctive features of this work compared to others? As already mentioned above and in the manuscript, the distinctive features are:

1) the integrated hydrological-hydraulic model used, based on a GPU-enhanced implementation of the 2D-SWE.

2) the modelling approach using only satellite-derived global and freely available data and software, with no specific calibration (an approach that could be used anywhere with no requirements of local data)
3) the validation of such a modelling approach using 3 kind of data: discharge time series, water depth marks and satellite imagery

Which are the challenges of performing a study like this? We consider that the methodology followed in our work and its detail assessment (as mentioned in the answer to the previous question), is in itself a challenge that can be of interest to the scientific community

What can be transferred to practitioners and stakeholders? A methodology to quantify flood hazard in data scarce regions using only freely available data and software.

Apart from the general issue, the manuscript is generally well written and easy-to-read, thus I have just few detailed comments (below).

184: the CN value is quite high, probably due to the consideration of a wet soil in the AMC. Was this the case (soil already wet) for this event, based on available records? I mean, apart from the fact that it will later give a good performance of the model.

Yes, as mentioned in section 3.1.4, we considered CN values corresponding to wet AMC conditions because the 5-day antecedent average rainfall depth in the basin was greater than 50 mm, which is the threshold commonly used to distinguish between normal and wet AMC conditions when applying the SCS-CN method.

212: this width of the river section would indicate that the used DEM is detailed enough to have a few points within the river (at least for the high-order stems), which is good. However, it seems that no correction was applied to the DEM, therefore the terrain elevation may be higher than real. It can be mentioned that this issue simply cannot be solved in the absence of a huge data availability.

We agree. Since the aim of the study was to assess what can be modelled using only global and freely available data, we did not modify the DEM in the streams. Moreover, we don't have any local data in order to perform such DEM corrections with a minimum representativeness of the real depth of the streams. However, considering the magnitude of the flood (most of the water during the peak of the event flows through the floodplains) the effect of lacking a detailed definition of the river bathymetry is probably low.

Actually, the fact that "the terrain elevation may be higher than real" is coherent with the positive Mean Error of 0.5 m obtained when comparing the observed and predicted water elevations at the 20 control points (section 4.1.2), i.e. the model tends to overestimate the water elevation, also for the reasons stated in the manuscript. This will be mentioned in the revised manuscript.

249: please specify if this water elevation was maintained constant or was changed based on available information during the event.

The water elevation at the outlet boundary was maintained constant during the simulation. This will be mentioned in the revised manuscript.

303: an equation is missing for the F_1 score (that, besides, does not seem to be used in the following as in the paragraph of line 385 only the HR and the FAR are mentioned).

This is a mistake in the text. We considered that the HR and FAR indices were enough to quantify and interpret the comparison between the observed and modelled water extent. We will remove the reference to F1 in the revised manuscript.

359: it sounds strange that the D-PLD contribution generates a second peak at the end of the event if (line 357) the hydrograph is "earlier" than the release from the dam. Should be mentioned that this, evident from Fig 10, must be due to a second precipitation peak in the lower basin.

Yes, this is the case. The second discharge peak in S2 is due to the second precipitation peak. The discharge generated by this precipitation peak in the U-PLD basin is controled by the reservoir, but this is not the case in the unregulated D-PLD basin. We will mention this in the revised manuscript.

370: while this may be true considering the extent of the model and the resolution applied, it could be acknowledged that in some cases we see overestimation by 2 to 2.5 m (around 100% of a value determined from the water marks).

Yes, this is true and it is already shown in Figure 11. However, due to the inherent uncertainty on the water marks estimations, and given that we compare water depths at 20 control points, we consider that the quantification of the agreement between model and observations should be based on statistics of the error, as it is done in lines 363-369. The fact that at some points the error is over 2 m is as relevant as the fact that at some locations the error is as low as 0.05 m. None of these extremes are relevant given the uncertainty on the depth estimations derived from the water marks. In our opinion, the MAE derived from 20 samples is much more relevant for the overall assessment of the results.