Dear Editor,

We thank you and both reviewers for your time and effort to review and handle this paper. We apologise for delays in our response caused by a combination of personal, work and health factors, and thank you for allowing us to resubmit following an extension.

We thank R#1 for approving our previous edits and clarifying that they want no further changes.

We are especially appreciative of the generous time given by R#2 who has read the paper closely and identified some important points for clarification and correction.

We explain how we have addressed their comments point by point below and feel the paper is consequently far more robust. We have also added further clarification and emphasised the key point of this work which is focused on global rainfall product intercomparison.

Kind regards,

Fergus McClean and co-authors

1. Lack of river channels in the hydrodynamic model. The revised paper argues that representation of river channels is not necessary for large events and cites two papers that it is claimed back this up (new text on lines 88-93). I'm a co-author on one (Neal et al, 2021), and I have to say this is absolutely not what our paper shows. In Neal et al (2021) there is no comparison of models with and without bathymetry, but there is in his Jeff's 2012 paper:

Neal, J., Schumann, G.J.-P. and Bates, P.D. (2012). A simple model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resources Research, 48, Paper no. W11506, (10.1029/2012WR012514

And what this shows is crystal clear: representation of channels is critical for correct simulation of flood propagation. In fact, the reason for developing the different ways of representing channels in Jeff's 2021 paper was precisely because the 2012 findings showed that this was so necessary. The paper by Dey et al (2019) has also been misunderstood. Like Neal et al (2021), Dey et al compare different ways of implementing the bathymetry in large scale hydrodynamic models and also do not ever consider the 'no bathymetry' case. The statement quoted in the text that "the choice of bathymetric model becomes irrelevant at high flows for predicting hydraulic outputs" does NOT mean that bathymetry irrelevant at high flow as claimed. Rather, it says that the choice of channel shape and how much longitudinal spatial variability needs to be included is unimportant, but the paper is clear that a channel of some form is still needed. In fact, two sentences on from the quoted test, Dey et al make the statement "1D hydraulic models for high flows do not require incorporation of geometric variability in channels and even using a simple triangular or rectangular shape is sufficient for flood modeling purposes". Dey et al's opening sentence is "Accurate representation of river bed topography, commonly referred to as bathymetry, plays a critical role in a variety of hydrologic and hydraulic applications including but not limited to flood modeling". The papers by Dey et al and Neal et al are remarkably consistent, but do not show what you claim. Moreover, the importance of bathymetry has been recognised all the way back to the very earliest papers on large scale hydraulic modelling e.g. Kate Bradbrook's work for JBA in the early 2000s.

The use of citations in the revised paper to support not including bathymetry is therefore both wrong and selective: you must surely have read the rest of the paragraph in Dey et al to see that they flatly contradicted the point you were seeking to make only a couple of sentences further on?

Moreover, there are good physical reasons why channels are central to flood routing and inundation prediction even during extreme flows. First, even at high flows a good proportion of the discharge still goes through the channel because of the high velocities there. From flume experiments and field observations we know that floodplain flow velocities are often an order of magnitude lower than those in the channel, so channel conveyance is still a major component of the total flux. Let's just think about the Carlisle 2005 event. In Carlisle, the Eden is ~70m wide and ~5m deep. With a conservative average channel flow velocity during the event of 1m/s that's 350 m3s-1 out of a total flow of around 1600 m3s-1. And in terms of channel flow:floodplain flow ratio, Carlisle 2005 is likely to be an extreme (low) end member. Channel velocities during the 2005 flood were probably even higher, and the event itself was 150 year return period so there was a lot of floodplain inundation. If you have no channel, then the unaccounted for channel flux has to be distributed over the floodplain and that must lead to a significant overestimation of flows there. Second, flood waves do not propagate in a physically realistic way without a fast-moving channel flow filament set within a slower moving floodplain background field. The Neal et al (2012) paper shows this very clearly, even when momentum exchange between channel and floodplain is not included.

I fully accept that your model is getting reasonable results, but the justification for not including bathymetry does not stand up to scrutiny, and actually runs counter to everything we know about channel-floodplain hydraulics. The errors so generated will be significant (my back of an envelope calculation using the real geometry suggests ~20% of the discharge for Carlisle 2005) and leads to the suspicion that your model may only be getting the right results for the wrong reasons.

I'm sorry, but you are going to have to provide improved text to justify not including channels. Iif plausible arguments cannot be found then you will have to be clear about the errors this can generates and explain why your model still appears to perform well in spite of these (i.e., explore what might be going on to compensate).

Thanks for this very important clarification and identifying our error. We apologise and have amended the paper to correctly represent these publications.

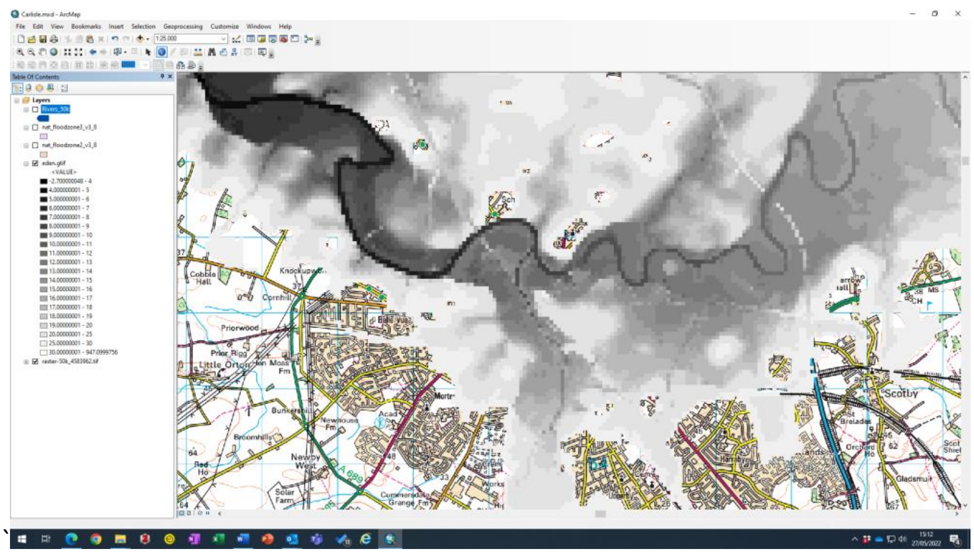
We have added further explanation about channel bathymetry including discussion about uncertainties. We have also referenced the Neal et al. (2012) paper you cite.

We have addressed the above point(s) in several sections throughout the paper. To summarise the key points we have included are:

- No editing of the DEM data has been undertaken as the focus of this work is on the sensitivity of model performance to different global rainfall products and we have, by choice, not adjusted any other data to focus on the rainfall.
- We note that a number of studies highlight the importance of the bathymetry is reduced at higher flows, we now clearly state our assumptions will likely lead to an overestimate of discharge.

- At lower elevations, and within valleys a river channel is captured directly within the DEM. For example, as in the screenshot below the shape of the channel is rectangular. Channel depth varies from 1.5-4m, the river width ranges from 50-100m (i.e. one to two DEM grid cells).
- OS Terrain 50 is well conditioned for slope, and some key features, and whilst absolute accuracy is low, its precision (i.e. relative accuracy over a small area) is therefore much higher.
- Further discussion is added about the implications of these assumptions and around uncertainties more generally, including on the accuracy of rainfall data, drawing from some of the literature cited by the reviewer and elsewhere.

We show below a screenshot of the DEM, with the scale limited to show elevations between 4m (darker) - 30m (lighter) for the Carlisle case study site. The river channel can be clearly seen as embedded in the DEM. Whilst we hope this is useful context to show in our response, we do not believe a plot like this is required in the paper but would be happy to produce something if the reviewer and editors think this would be helpful.



I think you also have to better explain how the model can get the stage-discharge relationship correct, as Figure 7 seems to show, when you do not have any channel bathymetry in the model. My understanding of the model simulations is that you run a ‘warm up’ period to get ‘normal flow’ in the river channels (line 116) which I assume means below bankfull discharge for the real geometry. The event simulation then takes the flow from what would, in real life, be in- to out-of-bank flow. Without a channel geometry (however approximated) how can stage in the model respond correctly to changing discharge? Figure 7 shows that the change of stage with discharge in the model is similar to the observations, but the latter implicitly includes channel bathymetry whereas the model does not. I am at a loss to explain this and just don’t understand how this can work.

We have added explanation to define normal flow.

As noted in response to point 1, we have clarified details about the river channel. However, Figure 7 plots a timeseries of stage, we do not plot stage-discharge, which hopefully allays the reviewer's concerns as to what Figure 7 is showing.

2. Calibration. One thing that might be a compensating effect which you don't discuss at all in the paper is calibration. How was this done, if at all, and what model parameters values did you end up using? I should probably have identified this in my previous review, but could you also add a short section outlining how the model was calibrated and commenting on the extent to which the optimum parameters are in the right physical range?

Our focus here is on the performance of different types of globally available rainfall data. There is therefore no specific calibration as this would compensate, differently for each catchment being studied, for the errors and differences in the data that this work is seeking to understand. Further discussion has been added to Section 4.

The only hydrodynamic modelling parameter that has been set is Manning's *n* which was uniformly defined as 0.03 based on Chow (1959), and is consistent with many other studies. We have now clarified this in the text and cited some other studies that justify this selection.

3. Lack of infiltration. I was more convinced by your justification for not including infiltration during extreme events as that seems at least physically plausible, but the paper you cite to justify this (Hossain Anni, 2020) is a small-scale study in an urban catchment, so again does not show quite you claim. Please could you replace with references showing the same effect at catchment scales.

We agree that this paper alone is insufficient on its own. We have therefore included further evidence from other more directly comparable studies to support this statement and understand the implications of this assumption (e.g. Ni et al. 2020; Hou et al. 2021). These report that peak flow and/or extent are insensitive to infiltration rates, but has an impact on outflows as the flood wave falls. The effect of this is greater for longer floods, and would be more significant in semi-arid or arid regions. Explanation has been added to Section 2.2 and in discussion in Section 4.

In the case of the Carlisle 2005 event it is widely reported that the ground was saturated, we have added a citation to this effect.

4. Lack of flood defences. I was surprised how you got such good results in Carlisle without including flood defences. Even in 2005 Carlisle was well defended and I don't think these were always overtopped, so this needs a couple of sentences of explanation in the paper.

Figure 4 compares observed and modelled depths, whilst Figure 5 shows how the models perform against the observed flood extent from the detailed mapping of the 2005 Carlisle event that was undertaken by: Neal, et al. (2009) Distributed whole city water level measurements from the Carlisle

2005 urban flood event and comparison with hydraulic model simulations, J. Hydrol., 368, 42–55, https://doi.org/10.1016/j.jhydrol.2009.01.026.

We have been unable to find more accurate information than that reported in the Neal et al. paper. Further, we have not been able to locate the historical defence levels for Carlisle in 2005, nor find studies that report which defences were overtopped and which were not.

Without further data it is not possible to answer this question with certainty, however, we would suggest that if there were lengths of defence that did not overtop in 2005 then those are located in parts of our model where the floodplain elevation is higher or rapidly rises, for example in the NW of the study area; or areas are flooding as floodplain flows allow water to go round the back of higher flood defences. Consideration of the latter mechanism is not something we specifically studied in this rainfall intercomparison, and would probably benefit from a different methodology if it were to be analysed.

This has been clarified in the paragraph on the results from the Carlisle flood in Section 3.

5. Line 116. Could you define 'normal' flow? I assume it is somewhere below bankfull discharge (see above) but I think this needs to be clear.

Thank you, we have added an explanation to clarify what is meant by normal flow: "*Antecedent rainfall is necessary to initiate normal flow in the river channels, which requires the water from all upstream cells to reach the outlet of the basin. Normal flow here refers to the flow in the channel before the flood event took place. If no spin-up period is included, then flood magnitudes would be underestimated, and the flood wave would not propagate in a physically realistic way.*"

6. Digital Elevation Model. I still can't quite understand what is happening with the DEM data. Thanks for confirming that OS Terrain 50 is photogrammetry and not LiDAR, but that does mean you are using a DEM with 4m RMSE of vertical error. If that is genuinely the case, how did you get inundation and water depth predicted as well as you do? Either the actual vertical error over the areas you simulate is substantially lower than 4m or there is something odd going on in the model to compensate.

First, I think you need to be clear about the DEM error in the current paper. You don't mention this anywhere, but it is huge (DEM RMSE >50% of the observed flood wave amplitude for all catchments).

Second, you need to explain how the model is still able to produce the results it does despite these terrain data errors.

The RMSE for OS Terrain 50 is reported to be 4m across the whole country, being calculated as the absolute accuracy relative to a number of GNSS points across Great Britain. This combines systematic (e.g. block linkages between photogrammetric observations) and random errors from one end of the country to another and is therefore not an absolute measure of accuracy for a given area of interest.

Local precision is more important here, this is the relative accuracy from point to point. This will generally be much better than RMSE, especially over the relatively small area visualised in Figure 15. OS don't publish the local precision of OS Terrain 50 but the Royal Institute of Chartered Surveyors suggest (Table 5, P24) that for an Area Of Interest it would likely be of the order of decimetres: www.rics.org/globalassets/rics-website/media/upholding-professional-standards/sector-standards/land/earth-observation-and-aerial-surveys-global-guidance-note-6th-edn.pdf

Although they don't report their local precision, OS Terrain 50 has been validated to meet the positional requirements for key features such as waterbodies, and to capture slope: (www.ordnancesurvey.co.uk/documents/product-support/support/os-terrain-50-overview-v1.5.pdf).

The combination of relative accuracy and positional validation against key features explains why the DEM is able to capture the relative positioning of key features that dominate the flood extent of major storms in this catchment. Further discussion on this and other uncertainties raised by the reviewer has been added into Section 2.4 and 4.

Lastly, your statement regarding OS Terrain 50 that it "has been shown to perform best for flood risk modelling in a comparison with other DEMs (McClean et al., 2020)" derives from a paper that did not include LiDAR data in the comparison. Instead, your previous work only compares OS Terrain 50 to a basket of global DEMs, which unsurprisingly have higher error. Numerous previous studies, starting with Sanders (2007):

Sanders BF. 2007. Evaluation of on-line DEMs for flood inundation modeling. Adv.Water Resour. 30(8):1831–43.

Have shown airborne LiDAR data sets to have considerable advantages for flood inundation modelling. In the above text from the paper you have used the word 'best' when you mean 'better' and have not qualified the subset of DEMs that your study refers to. Because of this your statement is misleading.

We fully agree, and have amended the text to use "better (relative to global DEMs)" rather than "best".

The Yunus study you quote is also not the greatest evidence here because: (a) it is a bathtub model which ignores hydraulic connectivity and (b) the events simulated are extreme and valley filling such that inundation extent becomes a very insensitive metric for comparing DEMs. There are now many papers on DEMs in flood inundation modelling and you need to consult a range of these to revise this text instead of using a self-citation in an inappropriate way.

As an aside, I still don't understand why you are not using LiDAR data where it exists? Most river floodplains in the region are now mapped and you could use these data and patch with OS Terrain 50 in hillslope areas away from the floodplain. LiDAR vertical errors are ~10cm RMSE and it is the gold standard.

We agree that high resolution LiDAR is the gold standard. The focus of the work overall has been to understand the impact of uncertainties on open and widely available data. The particular emphasis of this paper is on globally available rainfall products. OS Terrain 50 provides a convenient DEM product for this purpose because it limits additional processing as it does not require coarsening of high resolution LiDAR and/or merging of different DEM datasets which can introduce different uncertainties;

and a 50m resolution is sufficient to characterise flood extent whilst also limiting the computational expense for 2D physically-based hydrodynamic simulations.

We have added more explanation as to why we have used OS Terrain instead of LIDAR, including a reference to Sanders (2007) and other literature.

7. Figures. Some of these are difficult to read because the panels are small compared to the detail you are trying to show or because you have used a low-resolution raster file format instead of vector. The latter issue means that text, axes and other figure elements are quite pixelated. You need to redo these images at higher resolution or (better) as vector graphics.

Thank you for highlighting this, the resolution of all figures has been increased to 300 dpi and additionally provided separately as vector graphics.