Replies to reviewer #1

The revised article is substantially different from the previous submission, and addresses most of my initial comments. Also the quality of the english language has improved, with only some typos left. Also please check the ending "s" for third person singular form. I recommend that the article is published provided that the comments below are adequately addressed.

P1 l21: It is not observable over large areas, but it is observable over specific river reaches.

Re: Thanks, to be accurate, we revised the sentence as

> "However, FHM is a theoretical map of a global-identical reoccurrence (e.g., 1-in-100 year return period), and thus it is ***difficult to be observed, especially at a large scale***."

P9 l 18-19: Here I would call it "the ranking" rather than the relative magnitude

Re: "***The relative magnitude***" is now replaced by "***the ranking***".

Figure 5: Why are VARIABLES, RUNOFFS and FUNCTIONS uppercase? Also, fix the typo on the y axis, it should be "Proportion". (Same for Figure S1-S5). Also, increase the visibility of the yellow cross.

Re: We intended to emphasize the different groups. Now the uppercases are revised and the figure titles are presented in just a natural way. The typo "proportion" has been fixed in all figures and we have enlarged the yellow cross in the figure.

P15, l3: e.g, "should be taken with caution".

Re: Thanks, this sentence is revised as

> "This inundation will lead to migration and economic losses, and the impact ***should be taken with caution*** because of the uncertainties in inundation estimations."

P18 l2: I don't understand why "calibration will ruin the designed sensitivity test with different runoffs". That shouldn't happen if you use the same calibrated setup for all runs to compare.

Re: Sorry, the words lead to some confusions. We meant that we cannot re-calibrate our model for each specific runoff. For instance, if we apply one new runoff and we target to minimize the bias in discharge peaks by tuning model parameters, the final results related to the floods (e.g., flood water depth, flood inundation) will be useless for the sensitivity test. As the reviewer said, we have optimized the model parameters and validated our results driven by another given runoff (Lin et al., 2020) against GRDC records globally (although the results are not shown in this paper). Previous publications (Hirabayashi et al., 2013, Yamazaki et al., 2011) have also validated that CaMa-Flood is able to be used for flood assessment globally.

In the new submission, we revised the current sentence as:

> ***"However***, in this study, CaMa-Flood ***was not calibrated against observations for each specific runoff*** because ***additional*** calibration will ruin the designed sensitivity test."

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4), 1–21. https://doi.org/10.1029/2010WR009726

Replies to reviewer #2

The authors are to be commended for the improvements to their manuscript and the interesting results they present. A number of issues still remain outstanding, however, which require addressing before publication is to be considered:

1. [a]Language remains an issue in this manuscript. I appreciate the efforts of the authors, but the language as a whole remains difficult to follow and poor translation has led to some questionable commentary in some places. The copyeditors of the journal will be able to help with some of the more cosmetic errors, but the substantive meaning of what has been said needs to be checked by the authors. In particular, [b.]I'm not sure diagnostic/prognostic variables is a useful descriptor, and [c]in the abstract it states that runoff is 80% of total uncertainties -- please make clear this is the total of considered uncertainties, given an analysis of very large uncertainties (e.g. terrain data errors, river bathymetry) is not considered.

Re: [a.] Thanks, we checked throughout the manuscript and corrected/refined the texts we found. We think this will help improve the readability of the paper.

[b.] The prognostic variable (water storage) and diagnostic variable (water level) are specific for the river model (CaMa-Flood). Although we treat the two variables in a same way in sensitivity analysis in this paper, they mean differently in the CaMa-Flood as the primary causes to their variations are different as well. For instance, the variations as well as the uncertainties in water storage will be mainly caused by the water flux, while the bias of water level can also be significantly affected due to bias in the river bathymetry.

In the section 2.2 Global river routing model (CaMa-Flood), we added explanation:

> "Therefore, the estimation of water level will additional contain uncertainties in river bathymetry and topography, while uncertainties in the water storage are dominant by the water flux."

[c.] Yes, thanks the reviewer's comment. The quantification is only among what have been investigated in this paper (i.e., variables, runoffs and functions).

> "Our results show that deviation in the runoff inputs is the most influential source of uncertainties in the estimated flooded water depth and inundation area, contributing more than 80% of the total uncertainties ***investigated in the study***."

2. I still do not think the section on the AIC analysis provides useful conclusions in understanding model skill. The floods being simulated are extrapolations from these distributions, and the degree to which they fit 35 years of data does not illustrate how useful they are for extrapolation.

If I am wrong about this, then please add further discussion explaining why this is a relevant test of the model, as I am presently unconvinced.

Re: Thanks, we agree with the reviewer that the degree of AIC does not necessarily mean how useful they are for extrapolation. In practice, we need to find a historical flood record with a certain reoccurrence (e.g., 1-in-100 year, 1-in-500-year) to judge if the fitting and extrapolation is good or not. Obviously, we cannot do this for the global scale. Measurement of the fitting performance with current available CaMa-Flood estimates is the only way to interpret the variations among different fitting distribution.

Moreover, the variation of the AIC will be reflected in the extrapolation. Small difference in AIC will be enlarged when extrapolated. For example, AIC for GUM always deviates from other distribution functions (Figure 3). And we can observe from Figure 6d and 7d (and other figures in the supplementary), GUM tends to provide a much higher water depth for rarer floods compared to others.

Therefore, we still think the section of AIC analysis is necessary for the analysis, and we only made slight changes in this section to include the reviewer's comments.

3. Some interpretations of the results and comparison with the literature do appear quite tokenistic, rather than contributing to a rich discussion. [a.] E.g. P18 L3-13 loosely says routing is and is not important. [b.] P18 L14-24 is a little rushed in its review of these papers and it is not clear how this relates to the study in question. P11 L5-10 suggests the spatial patterns are not consistent with Schellekens et al., but I disagree. Although the proportion of the contribution to the uncertainty by runoff forcing is fairly consistent in space (Fig 4e), the magnitude of the uncertainty (Fig 4c; the relevant figure for the point being made about Schellekens' conclusion) is larger in mountainous and arid regions.

Re: a. Thanks, we listed in the Discussion a few studies related to the river routing. However, because none of them specifically investigated the sensitivity of the flood water depth or flood inundation to the river routing, we can hardly conclude how important the river routing is. This requires further studies.

b. We have revised the Discussions. First, we combine the P18 L14-24 with the previous paragraph as both of them are discussing the potential factors affecting the flood hazard mapping. Then we add one new paragraph to discuss what are the uncertainty source can be considered in the future to the flood impact (i.e., population exposure, economic exposure). This is also a response to the reviewer's next question. Please check the text with tracks.

c. Here we just want to mention that the contribution of runoff to the uncertainties is not sensitive to the climate zones. Both Fig 3c and the figures in Schellekens et al., 2017 show that the magnitude of the uncertainties have spatial variation in mountainous and arid regions, however, the spatial variation is not revealed in the contribution map (Figure 4e). So, we modified our texts to avoid this confusion.

> "the spatial patterns of runoff spread in their results ***and the variation in Figure 4-c*** are not seen in ***Figure 4-e***, indicating that the contribution of runoff to the total uncertainty in flood water depth is not sensitive to climate zones or topography."

4. [a.]Section 3.3 is still a fairly long-winded way of saying that runoff is most important. I would suggest condensing 3.3.1 and 3.3.2 without using so much text, as the conclusion is quite simple -- consider what the reader truly needs to know. [b.] A GRDC gauge is mentioned, but then no analysis is done using this, which is confusing. [c.]Section 3.3.3 is interesting, but the authors should be much more explicit about the uncertainties in the exposure data they are using: how accurate is 1km resolution population data? If you analysed sensitivities with the 90m flood model with higher resolution population data, I expect conclusions would be very different. There is lots of literature on the accuracy of these global population datasets and the impact of resolution that should be discussed.

Re: [a.] Thanks for the suggestions. In section 3.2, the results show that the runoff is the most important for the 1-in-100 year flood. In sections 3.3.1 and 3.3.2, we investigated if this importance is the same for all return periods. Although the results still demonstrate that runoff is the most important for all return periods, we have to use sufficient words to interpret the figures (Figure 6 & 7). Despite the conclusion for the importance of runoff, we also found the importance of the functions increases from normal return period to the rarer floods, as the uncertainty range for the rarer floods is larger than that for floods with higher reoccurrence.

We have reviewed the two sections and revised the texts but not too much. Please check the text with tracks.

[b.] The flood water depth is the variable that can be accurately measured if suffering a flood. However, as one of the reviewers mentioned in the first round reviewing, we have to select one representative location, so that we select the GRDC gauge. This will help if others conduct similar results and want to compare the results with this study.

[c.] Thanks, yes, we then found a perfect paper (Smith et al., 2019) for discussion about the resolution.

Table 2 Total population living in the 1 in 100 year floodplain (millions) summed across all 18 countries, for varying resolutions of both hazard and population data

		Population data					
		30 m	90 m		900 m		
		HRSL	HRSL	WP	HRSL	WP	LS
Hazard	90 m	101	102	122	124	130	134
	900 m	196	196	205	197	205	203

In their paper, they evaluated the population exposure to the 1-in-100 year flood in 18 different countries but with different population products. Meanwhile they aggregated the hazard map and population data from high spatial resolution (30m or 90m) to lower spatial resolution (900m) to investigate the sensitivity of the population exposure. From their Table 2, when the hazard map is at 90m, the exposure due to resolution changes of population from 30m-90m is negligible, while increased 8-22% if population resolution decreases to 900m. When the hazard map is 900m, the exposure does not change with population resolution, while the total exposure increased by 51% to 94% compared to 90m hazard map. Therefore, the spatial resolution of the hazard map is the most import factor to the final population exposure.

In the revised manuscript, this part has been added to Discussions.

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications*, *10*(1), 1–7. https://doi.org/10.1038/s41467-019-09282-y

5. Some figures require improvement. The x axes on Figs 6+7 make no sense. Fig 5 is too small also (where is the yellow cross?).

Re: We have cut the x axes and only the floods corresponding to return period larger than 2 years are remained. Figure 5 has also been enlarged and the yellow cross is enlarged. Figures in the Supplementary are revised as well.