**In this document, the review comments are black color, our responses and our comment) are red color and the revised text in the revised manuscript are blue color.**

5   **<Revised point other than review's comment>**

Author's comment:

Please note that we have found errors in the exposure calculation procedures for reanalysis data results and Lookup method results in section 4.3 and have corrected them. Please also note that this does not affect the conclusions of this paper.

10  Specifically, we have reviewed Table2, Figure 9 and the inundation areas and affected population in the reanalysis data results and Lookup method in section 4.3.

Table 2: Inundation area of the 100-year RP hazard maps and affected population from the 100-year RP hazard maps constructed using various methods; "_no_modification" indicates a map constructed without Backwater_Modification. These results were obtained using IPSL-CM6A-LR.

|  |  | Inundation Area (Unit: Million km^2) | | Exposed Population (Unit: Million) | |
|---|---|---|---|---|---|
|  |  | all | depth > 5m | all | depth > 5m |
| Historical Climate | Reference historical (Reanalysis) | 18.38 | 2.19 | 1,625 | 216 |
| Future Climate | GCM uncorrected | 20.28 | 4.20 | 1,783 | 395 |
| | Runoff Correction Method | 20.45 | 4.02 | 1,802 | 406 |
| | Lookup_Method_no_modification | 19.49 | 3.23 | 1,821 | 417 |
| | Lookup_Method | 19.67 | 3.43 | 1,858 | 441 |

**<Point-by-point response to the review #1>**

20  Line 58: It would be good to add some references on in country hazard mapping. In general as this paragraph is an overview of local-scale flood hazard mapping efforts I found it a bit citation lite. It would be nice to signpost some other review articles as De Moel et al 2009 was published some time ago.

At the risk of self-promotion (no need to include any of this unless it's useful to support your later arguments) we included a UK focused review of local-scale modelling efforts that highlights some of the challenges of using these data sets in a

25  climate conditioned application.

Bates, P. D., Savage, J., Wing, O., Quinn, N., Sampson, C., Neal, J., and Smith, A.: A climate-conditioned catastrophe risk model for UK flooding, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-829, 2022.

Reply: We agree with referee #1 suggestion, so we have added "Bates et al., 2022" and "Wing et al 2022" as a reference.

Bates, P. D., Savage, J., Wing, O., Quinn, N., Sampson, C., Neal, J., and Smith, A.: A climate-conditioned catastrophe risk
30  model for UK flooding, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-829, 2022.

Wing, O. E., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., Neal, J. C, Porter, J. R, and Kousky, C.:
Inequitable patterns of US flood risk in the Anthropocene. Nature Climate Change, 12(2), 156-162,
https://doi.org/10.1038/s41558-021-01265-6. 2022.

"In the research field, for example, Bates et al., 2022 estimated annual flood damage in UK and Wing et al., 2022 estimated
35  flood damage in US by using local-scale flood-hazard maps." (Line 59-60 in revised manuscript)

40  "Line 81: assessment of future flood risks based on the spatial distribution of inundation depths has not yet been established"
I agree at the global scale but there are many local scale examples of such studies that have had to consider the GCM bias
issue. I think that needs to be acknowledged explicitly in this section or the previous paragraph.

Reply: We would like to express our gratitude to referee #1. We have revised the sentence as follows. We clarified that in
this paper we described a "global" study. We also clarified that flood "risk" has been assessed, but flood "hazard" itself has
45  not been sufficiently verified in previous global-scale studies.

"Although climatic and meteorological hazards under future climate change (e.g., extreme temperatures, droughts and
heavy-rainfall events) have been widely assessed using direct output variables of general circulation models (GCMs), such
as precipitation and temperature (Li et al., 2021, Lu et al., 2019), to date, no global high-resolution future flood hazard (i.e.,
inundation depth) has been sufficiently verified. Some studies have evaluated future flood risks (e.g., affected population and
50  GDP) at the global scale (e.g., Ward et al., 2020b); however, it is important to analyze global future flood hazards (i.e.,
inundation depth distribution), and to assess uncertainties such as those caused by various bias correction methods." (Line
80-86 in revised manuscript)

55

For example, this sentence "While bias-correction methods for precipitation and temperature have been studied in detail
(Watanabe et al., 2012, Hempel et al., 2013 and Lafond et al., 2014), such methods have not been established for runoff data
for use as inputs to flood models" is not correct without a more precise focus on the novelty of this study because there are
60  many examples of localized climate conditions hydrological cascades all the way to inundation hazard and impact. These
studies have some of the same issues you so I would have thought there were some useful conclusions from the literature
that might feed into this section.

65   "such methods have not been established for runoff data for use as inputs to global flood models to construct large-domain future flood hazard maps." (Line 94-95 in revised manuscript)

70

Line 113: "Although CaMa-Flood is a global model, it is unique in that it represents the physical processes necessary to reproduce floodplain inundation dynamics" I appreciate there are unique aspects to CaMa-Flood, however there are several global flood inundation models so perhaps remove the word "unique" or be more specific about the combination of capabilities that are unique.

75   Reply: In response to referee #1 comments, we have removed "unique" and revise the sentence as follows.

"Although CaMa-Flood is a global model, it has characteristics that it represents the physical processes necessary to reproduce floodplain inundation dynamics" (Line 116-117 in revised manuscript)

80

Line 137: could you be specific about the simulation length – I suspect its shorter than most readers will assume by the wording long term. Is it 35 years?

Reply: As you understand, the simulation length is 35 years.  Based on referee #1 comment, we have removed "long-term"

85   and added "(time period: 1980-2014)" to the end of the sentence to specify about the simulation length.

"First, we conducted a historical river hydrodynamics simulation with a daily time step using observation-based runoff data (Reanalysis_Runoff) as an input to CaMa-Flood (time period: 1980-2014)." (Line 140-141 in revised manuscript)

90

Line 150: I have no issue with this pragmatic method for getting a spatially consistent hazard map, but I'm not convinced by the description. The hydrodynamic model simulated the backwater effect so I assume that this effect is captured in the hazard/level simulations. Could it be the case that the spatial inconstancy is therefore due to the distribution fitting and other

95   data processing factors rather than a physical backwater effect being omitted? I appreciate the maps will look better with this correction but do you risk biasing the levels higher in the process? On a more fundamental level a hazard map is not an event,

so does the water level even need to monotonically decrease downstream since you simple seek to best simulate the hazard rather than a plausible physical water surface which a hazard map will never be?

Reply: We thank referee #1 for comment. we have revised the sentence to make it clear that it is the case that the spatial inconstancy is due to the distribution fitting at each unit-catchment scale (grid-scale) as follows.

"If reverse-slope revision is not conducted, reverse slope produced by fitting the Gumbel distributions remains and the inundation-depth distribution may not be physically reasonable. For this reason, a novel reverse-slope revision method was applied in this study in the purpose of revising the spatial inconstancy caused by distribution fitting at each unit-catchment scale (grid-scale)." (Line 156-160 in revised manuscript)

In response to referee #1 comment "I appreciate the maps will look better with this correction but do you risk biasing the levels higher in the process?", we added the notes of overestimation to the last paragraph in section 4.2 (Specifically, the following sentence have been added).

"Please note that it is possible that it would be overestimated by applying Backwater_Modification because it performs corrections at all reverse slope occurrence points. Reversed water surface slopes can occur in conventional flood hazard maps, given that these maps are not always constructed by a single flood event simulation. Therefore, we conducted additional validation to investigate whether Backwater_Modification should be applied." (Line 457-460 in revised manuscript)

Section 2.3.1: Could you cite some alternatives to the bias correction method and explain why this one was chosen. I'm not an expert in these methods, but my assumption is that the results of this method would be sensitive to the choice of bias correction. Depending on your response above you might need to caveat the conclusions drawn around line 336.

Reply: We thank referee #1 for comment. We cited some alternatives to the bias correction method and explained why this one was chosen.

"Alternatives to the additive correction method include multiplicative correction method, which multiplies the ratio of GCM to the reanalysis data, and quantile-based mapping method (Panofsky and Brier 1968, Watanabe 2020), which obtains ordinal statistics from the reanalysis data and the GCM and creates an equation relating these statistics. In the future climate, the average monthly runoff may fluctuate significantly due to changes in the humid and arid zones and the timing of the wet and dry seasons caused by increasing temperatures. We selected the additive correction method because it is relatively insensitive to such fluctuations." (Line 197-202 in revised manuscript)

4

Line 383: "On the other hand, the lookup method did produce hazard maps consistent with the changes in flood risk under the future climate projected by the GCMs." I found this a bit confusing. Do you mean flood risk here? I assume you can only derive risk from the hazard maps, do you mean the changes in GCM runoff or precipitation? I think this section needs to be more precise about what you are looking for constancy with because I don't think its risk – or at least I didn't understand how it could be from the text.

Reply: As you pointed out, because we use "flood risk" incorrectly in the sentence, we have revised the sentence to clarify "the changes in flood hazard under the future climate projected by CaMa-Flood simulations". Specifically, we have changed from "the changes in flood risk under the future climate projected by the GCMs" into as follows.

"the changes in flood hazard under the future climate projected by CaMa-Flood simulations with GCM runoff input." (Line 394-395 in revised manuscript)

Line 390: "4.1 Which method is more convenient for generating future-hazard maps?" Convenient is the wrong word here it only applies to some of your arguments.

Reply: We agree with your comment. we have change from "convenient" into "reasonable" in the title of section 4.1.

"4.1 Which method is more reasonable for generating future-hazard maps?" (Line 401 in revised manuscript)

Line 394: I think you could also note here that this approach can also be used with hard mapping methods based on regional flood frequency analysis or machine learning from gauging station data, which at least in historically well monitored river reaches can be more accurate than rainfall-runoff reanalysis based methods (perhaps this is contested) for present day hazard mapping and extreme flow estimation because the modelling chain is much shorter (e.g. Laura Devitt et al 2021 Environ. Res. Lett. 16 064013 DOI 10.1088/1748-9326/abfac4)

Reply:

Thank you very much for your suggestion. As you pointed out, there is no need to limit it to reanalysis-basis as a "historical hazard map". Based on the above, we have added the following sentence to the end of 1st paragraph in section 5 "Conclusion".

"Thus, combining accurate historical hazard maps with information on future flood frequency changes may be optimal in general for generating future hazard maps. Please note that the historical flood hazard maps are not required to apply reanalysis-based simulations using global flood model; the proposed method is also applicable to gauge-based or machine-learning based historical hazard map." (Line 546-549 in revised manuscript)

Line 407 I think this is a separate point from the one above "The lookup method also has the advantage of facilitating research on efficient construction of future climate hazard maps, as it allows for improvement of the reanalysis hazard map by upgrading the model, and the estimated changes due to climate change can be considered separately" and you could expand to discuss using multiple flood hazard models for the analysis as several are introduced in the preceding section.

Reply: We would like to express our gratitude to referee #1 suggestion. We added one point to advantage of the lookup method.

"The lookup method also facilitates research on the efficient construction of future climate hazard maps because historical hazard maps can be prepared separately from the estimation of future frequency change. This is beneficial for two aspects: 1) it allows for improvement of the reanalysis hazard map by upgrading the model; 2) it allows for use of multiple reference hazard maps by using different reanalysis-based simulations" (Line 420-424 in revised manuscript)

Section 4.2: I think it's valuable to discuss this correction (so I'm not suggesting a major revision). But in my opinion it's debatable whether this correction is desirable for a hazard map which is not an event and thus not a physically possible water surface. I think it potentially double counts for the backwater effect and over-predicts the flood inundation levels by biasing locations upstream to any over-prediction errors in the distribution fitting downstream. That said the hazard mapping is improved relative to some high quality validation data so I don't dispute that some form of postprocessing of the levels to aid spatial constancy is the wrong thing to do… but you could smooth for example. Personally I would slightly modify the discussion in section 4.3 to present the results with and without the backwater modification as an indicator of the sensitivity to this issue – which is less than the climate change signal and impact of different GCM treatments. This would also require a small edit to the conclusion ~line 532.

Reply: We would like to express our gratitude to referee #1. Backwater_Modificiation is applied in order to revise the spatial inconsistency that is due to the distribution fitting (fitting Gumbel distribution) at the unit-catchment scale. In response to referee #1 comment "I think it potentially double counts for the backwater effect and over-predicts the flood inundation

6

levels by biasing locations upstream to any over-prediction errors in the distribution fitting downstream." we have added the notes of overestimation (Specifically, the following sentence will be added) in the last paragraph on section 4.2.

200    "Please note that it is possible that it would be overestimated by applying Backwater_Modification because it performs corrections at all reverse slope occurrence points. Reversed water surface slopes can occur in conventional flood hazard maps, given that these maps are not always constructed by a single flood event simulation. Therefore, we conducted additional validation to investigate whether Backwater_Modification should be applied." (Line 457-460 in revised manuscript)

205

Line 465: why choose lower resolution population data rather than say Worldpop? Could this bias the results?

210    Reply: We would like to express our gratitude to referee #1 question.

In this study, as to the estimates of future flood risk change, we compared the differences between the two methods to construct a future hazard map or with/without implementation of bias correction. Since the purpose is to analyze the impact of correction at the global scale, we think that 30 arcsec resolution was enough.

We used Gridded Population of the World; (CIESIN, 2018), which has data that are stored together for the entire world,

215    rather than Worldpop, whose data are broken down by country.

In addition to above, we added the following notes regarding uncertainty in spatial resolution in section 4.3.

"As discussed by Zhou et al., 2021, the spatial resolution of a flood hazard map is a particularly important determinant of its value for impact assessment. Smith et al. (2019) evaluated the population exposure to a 1-in-100 year flood in 18 developing countries, and found that decreasing the spatial resolution of the flood hazard map from 90 to 900 m increased the exposure

220    by 51-94 % for different population products. Although there is uncertainty involved in the choice of the spatial resolution of the flood hazard map, we selected 30-arcsec instead of 3-arcsec resolution to compare future flood hazard map construction methods on a global scale." (Line 497-502 in revised manuscript)

225

<Point-by-point response to the review #2>

L25: "...changes in flood risk..." I would correct in "changes in flood hazard"

230    Reply: As you pointed out, because we use flood risk incorrectly in the sentence, so we have revised it from "risk" to "hazard" as follows.

"On the other hand, the lookup method produced future-hazard maps that are consistent with flood hazard changes projected by CaMa-Flood simulations obtained using GCM runoff input, indicating the possibility of obtaining reasonable inundation-area distribution." (Line 24-26 in revised manuscript)

L29 "we discuss future changes at global scale..."

Reply: We have revised the sentence as follows.

"We discuss future changes at global scale in inundation areas and the affected population within the inundation area." (Line 29-30 in revised manuscript)

L49-50: "To elucidate the potential impacts of flood disasters, a high-resolution map of potential disaster impacts must be developed, commonly named a hazard map." This is not fully correct. In flood risk literature, hazard is a component of risk but it is not a synonym of disaster impact (see for instance Ward et al, 2020). Perhaps you could replace with "To elucidate the potential impacts of flood disasters, high-resolution maps of disaster impacts must be developed".

Reply: In order to make an appropriate sentence that leads to the commonly named a hazard map, we have revised the sentence as follows.

"To elucidate the potential impacts of flood disasters, high-resolution inundation-depth maps must be developed, commonly named a hazard map." (Line 50-51 in revised manuscript)

L70-72 "Uses of large-domain flood-hazard maps include estimation of the affected population within an inundation area and determination of the impacts of flooding on GDP and urban areas in the current climate". This sentence should be reworded. In literature, the maps including impacts such as population and/or urban areas exposed are generally called flood risk maps

Reply: As you pointed out, the sentence was not clear and was not stated correctly, so we have revised as follows to clarify the meaning of "Large-domain flood-hazard maps were used for many applications such as estimation of risks." Specifically, we have revised it as follows.

"Large-domain flood-hazard maps have been used in many applications such as estimation of the affected population within an inundation area and determination of the impacts of flooding on GDP and urban areas in the current climate (Ward et al., 2020a)." (Line 72-74 in revised manuscript)

270

L72-74: Bernhofen et al (2018) compared six global flood models against satellite-derived flood maps, so adding a reference here would be appropriate in my view.

Reply: We agree with your point. We have added Bernhofen et al (2018) to the sentence as follows.

"Hirabayashi et al., 2022, Trigg et al., 2016 and Bernhofen et al., 2018 compared multiple global flood models and analyzed the factors contributing to differences in inundation areas and depths." (Line 75-76 in revised manuscript)

280

L80-81 this sentence should be modified, because several flood risk assessments have been carried out at scales from global to local.

Reply: We have revised the sentence as follows. We clarified that in this paper we described a global study. Also, we clarified that flood risk has been studied, but flood hazard itself has not been sufficiently verified.

"Although climatic and meteorological hazards under future climate change (e.g., extreme temperatures, droughts and heavy-rainfall events) have been widely assessed using direct output variables of general circulation models (GCMs), such as precipitation and temperature (Li et al., 2021, Lu et al., 2019), to date, no global high-resolution future flood hazard (i.e., inundation depth) has been sufficiently verified. Some studies have evaluated future flood risks (e.g., affected population and GDP) at the global scale (e.g.,Ward et al., 2020b); however, it  is important to analyze global future flood hazards (i.e., inundation depth distribution), and to assess uncertainties such as those caused by various bias correction methods." (Line 80-86 in revised manuscript)

295

L147-154: I think that the description of the post-processing method needs more detailing. In particular I have some questions:

300    - My understanding is that the authors fitted a Gumbel distribution on each pixel of the Camaflood 6-arcmin grid, correct? Or do you use different areas for the fitting?

Reply: We have revised the sentence as follows to clarify that fitting a Gumbel distribution on each pixel of the CaMa-Flood 6-arcmin grid.

"We fitted the Gumbel distribution (Zhou et al., 2021) to the time series of annual maximum river water levels using the L-
305    moments method (Hosking, 2015) on each of the CaMa-Flood 6-arcmin grid." (Line 146-147 in revised manuscript)

310    - In lines 155-162 you state that water surface elevation is uniform within each 6-min unit catchment, so I assume that water level in upstream catchment are increased to the same water level of downstream catchment, right? If yes, please specify this in the text

Reply: Referee #2's understanding is correct. As referee #2 suggested that, we have revised the sentence as follows.

"To avoid this issue, if a reverse water slope was obtained in the water surface elevation distribution, we increased the water
315    surface elevation of upstream catchments to match those of downstream catchments." (Line 154-155 in revised manuscript)

320    - Can you also explain why this approach was not needed in previous studies based on CamaFlood?

Reply: We thank referee #2 for question. We would say the risk estimate of the previous studies could be improved by applying this reversal-slope modification approach. As described in the introduction, flood depth distribution itself had not been extensively evaluated in previous studies.

In addition, the purpose of applying reverse-slope revision is written more clearly as follows.

325    "If reverse-slope revision is not conducted, reverse slope produced by fitting the Gumbel distributions remains and the inundation-depth distribution may not be physically reasonable. For this reason, a novel reverse-slope revision method was applied in this study in the purpose of revising the spatial inconstancy caused by distribution fitting at each unit-catchment scale (grid-scale)." (Line 156-159 in revised manuscript)

330

L218: Alfieri et al (2017) actually employed the lookup method, because they used historical flood hazard maps coupled with changes in frequency under future climate scenarios.

335 Reply: We added Alfieri et al (2017) as a reference as follows.

"Alfieri et al., 2017 used the lookup method to estimate future affected population and damage by flood using historical flood hazard maps coupled with frequency changes under future climate scenarios." (Line 225-227 in revised manuscript)

340

L405-406: Based on the results, the authors could maybe draw the conclusion that standard bias-correction techniques of GCM data are not suitable for use in flood hazard estimation (having been developed for different types of climate studies), and that different bias-correction techniques should be used (i.e. more focused on extreme values)

Reply: We agree with referee #2 suggestion. In response to referee #2' suggestion, we revised the sentence in the section 4.1

345 as follows.

"This result suggests that simple bias-correction techniques of GCM data, i.e., additive correction to monthly mean runoff, may not be suitable for use in flood hazard estimation and that various other bias-correction techniques that focus on extreme values should be tested." (Line 418-420 in revised manuscript)

350

Figure 8: Can you specify the unit of measure of the x axis in (d) and (e)?

Reply: We thank referee #2 for comment. We specified the x axis in (d) and (e).



11
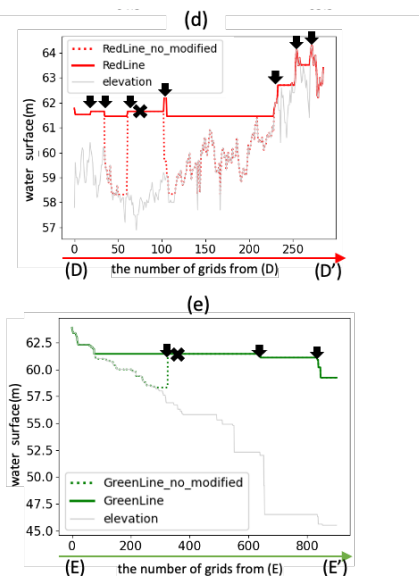
355   L488-503: I do not fully understand this analysis. Combining historical hazard maps with future flood frequency is basically the lookup method, right? (e.g. assuming that present-day 100-year RP will become 50-year RP in the future). Based on the outcome of the paper, I would rather conclude (here and in Conclusions) that historical hazard maps can be used as an indication of future hazard only if changes in flood frequency are properly accounted for.

Reply: We thank referee #2 for comment. As you pointed out, in general, estimating future flood frequency change and
360   combining it with historical multiple-frequency flood hazard map is a good solution to construct the future hazard map, and it can be said as "lookup method". We added the below sentence in the end of the 1st paragraph of conclusion section as follows.

"Thus, combining accurate historical hazard maps with information on future flood frequency changes may be optimal in general for generating future hazard maps." (Line 546-547 in revised manuscript)

365   We'd like to note that the proposed look-up method applying extreme-value analysis at catchment-scale is beneficial at global scale studies since the frequency change varies from basin to basin and sub-basin to sub-basin.