Reply to the comments by Reviewer #2:

In this document, the review comments are in black, our responses are in blue.

Reviewer #2

The paper proposes a methodology to reduce the uncertainty in flood hazard for future scenarios by integrating existing data to create a larger dataset, as uncertainty due to the small number of ensemble members is typically an issue for similar applications.

The proposed approach includes the identification of scenarios that bring to the same increase of temperature, which are then used together to increase the number of ensemble members. Then, the discharge is calculated with such enlarged dataset, to compare discharge with historical data. Unbiased variance is finally used to compare the results taking into account the different size of the considered ensembles.

Although I believe that the approach is interesting and may be appropriate in general applications, I disagree with the application to flood event analysis. This because 30-years time-series are analyzed together as they represent a 90-years time-series, which is not the case, just thinking about the difference in term of discharge values with 30 and 100 years of return period.

> Thank you very much for your constructive comments which are very helpful for improving the manuscript. We made a plan for modifying the manuscript, with additional background information and additional quantitative analysis, as summarized below. These are to be included in the revised version of the manuscript.

1. Merging data that brings to the same warning increase seems reasonable. However, the time series used are 30-years long. I am not sure that by putting 3, or 2, time-series together you gain data for 90-year, or 60-year simulation. The three, or two, series, were derived for 30 years, and 90-, or 60-years dynamics may bring to different data, especially in terms of extreme values.

>First, we would like to confirm that the 30-year sample is often used to estimate low-frequency floods, such as 1-in-100-year floods, for future climate flood projections, as demonstrated in previous researches (Hirabayashi et al (2013); Dottori et al (2018); Winsemius et al (2016)) and in numerous recent studies cited in the latest IPCC AR6 (IPCC, 2021).
Hirabayashi et al. (2013) justified this approach by showing that the extreme distribution function provides a relatively reasonable fit even for a 30-year sample (Figure R1). Furthermore, they demonstrated that the changes in the distribution of 1-in-100-year floods estimated from a 30-year sample are almost unchanged

compared to 1-in-10 or 1-in-30-year events (Figure R2). On the other hand, they also claimed that distribution of the 1/150 flood event is not exactly the same as the 1/100 event, indicating that there is no linear law of similarity in the return period.

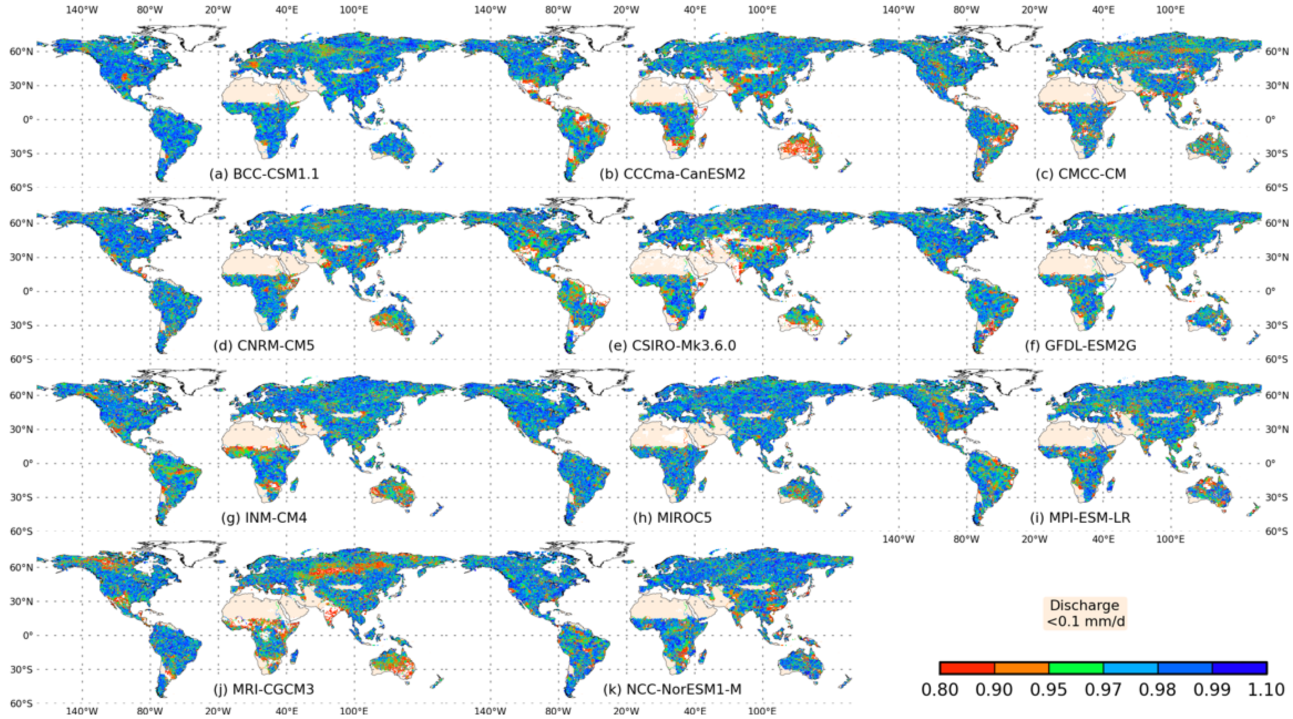

Figure R1: Schematic diagram summarizing the existing finding on 'extreme distribution function provides a relatively reasonable fit even for a 30-year sample.'' Figure taken from Fig.S4 of Hirabayashi et al. 2013, Nature Climate Change.

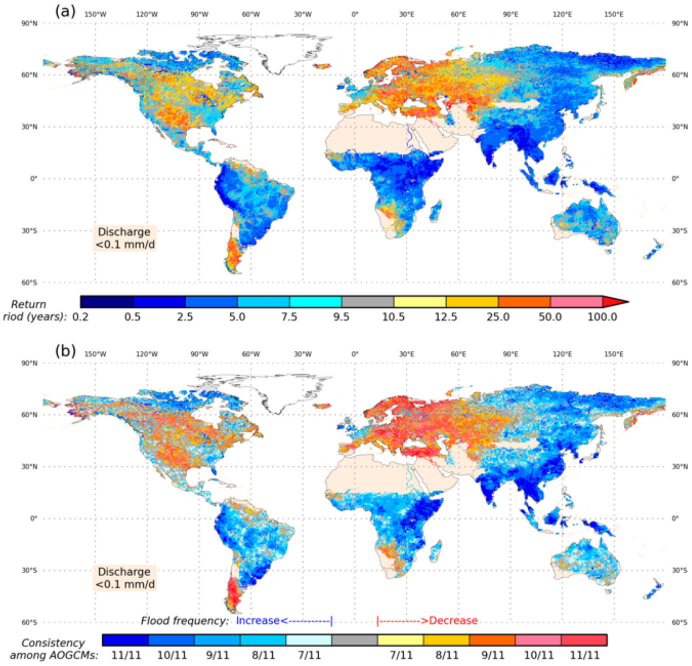

Figure R2: Projected change of 1-in-10-year floods. Schematic diagram summarizing the existing finding on 'the changes in the distribution of 1-in-100-year floods estimated from a 30-year sample are almost unchanged compared to 1-in-10 events.’ Figure taken from Fig.S7 of Hirabayashi et al. 2013, Nature Climate Change.

Other concerns include the reduced likelihood of large extremes events occurring within the 30-year sample. Increasing the number of ensembles by integrating SSP-RCP scenarios increases the possibility of occurrence of low-frequent and high extreme events in the selected temperature range. In Hirabayashi et al. (2013), a global statistical test was conducted to see if river floods can show significant changes in a 30-year sample, and the results showed that in places where there was a high agreement in climate predictions between models, the 30-year window shows a similar trend of increasing or decreasing floods in many GCMs (Fig.R3).

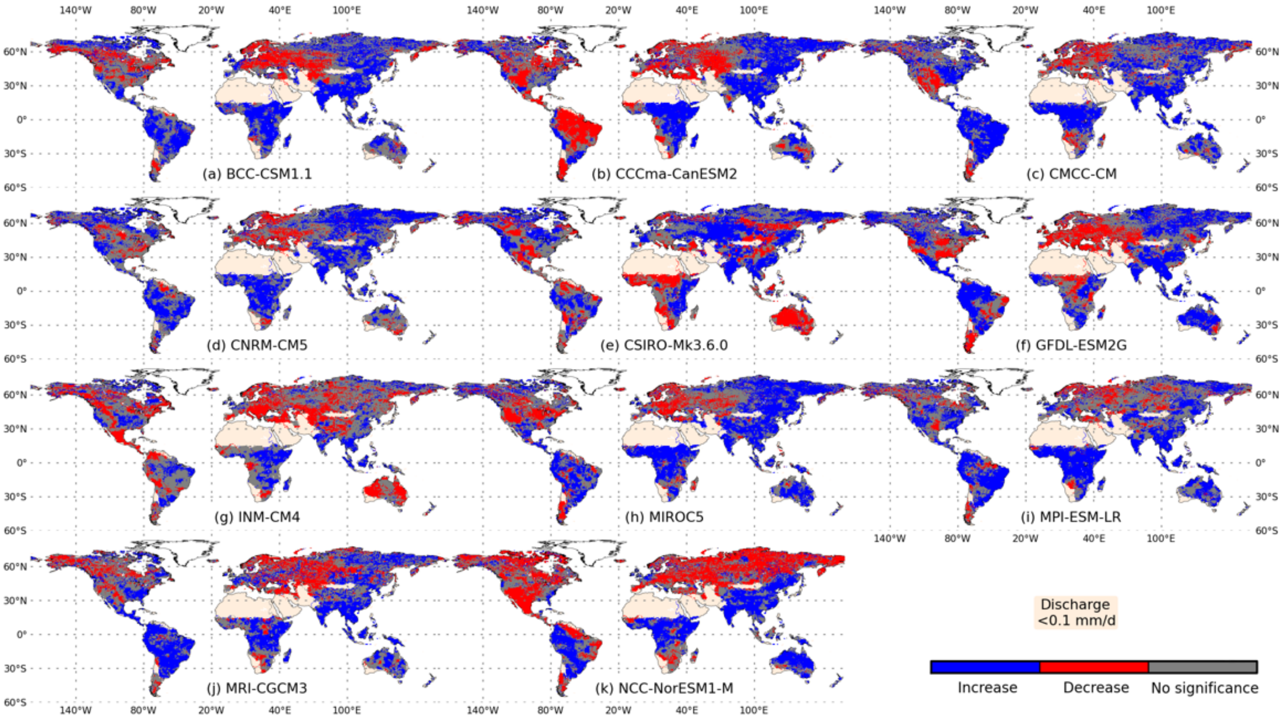

Figure R3: Schematic diagram summarizing the existing finding on 'a global statistical test showed that in places where there was a high agreement in climate predictions between models, the 30-year window shows a similar trend of increasing or decreasing floods in many GCMs’. Figure taken from Fig.S5 of Hirabayashi et al. 2013, Nature Climate Change.

On the other hand, where the climate internal variability is large, the flood projection showed weak multi-model agreements, because the future trend is affected by an occurrence of an extreme event in a simulated 30-years time series. The reviewer’s concern that “if the three-time series were derived for 30 years, 90-years dynamics may bring to different data, especially in terms of extreme values.” is true, and our method is actually proposed to reduce the uncertainty of using a specific 30-year data for low-probability extreme value estimates. The previous study （Kita and Yamazaki,. 2023) also suggested that the use of multiple time-series from ensemble simulation helps to reduce the uncertainty in extreme flood estimation. Note that we cannot use one 90-year time series because the flood characteristics are likely to be affected by the climate change during the 90 years, thus we combined multiple 30-year time series at the same degree of global warming. (Please also see the reply to Reviewer #1)

2. I suggest the Authors to anticipate the lines from 152 to 162 before chapter 2.1, as it may help the reader in understanding the entire methodology.

>Thanks for a good suggestion. In the revised manuscript, we will include a theoretical description of uncertainty in future flood projections in the main text and add an overview explaining the type of experiment proposed to investigate how increasing the ensemble size affects uncertainty. This will be included at the end of the paragraph before Section 2.1.

3. The effect of topography variation could also be taken into consideration.

>The CaMa-Flood model used in this study considers the impact of topography difference on flood hydrodynamics (please see the model description paper, Yamazaki et al. 2011, Water Resources Research). However, as we show the “flood discharge change ratio” which is relative to the present-day discharge, the impact of topography is not clearly shown in the map of our result.

4. There are many references to the supplementary materials. It would be better to think about some metrics that could summarize the behavior of the different conditions tested and move the text that describes the single figures in the supplementary materials themselves. In case the Authors believe that the fomented figures (e.g., 187-189) are necessary to better describe the results, these figures should be included in the main text rather than in the supplementary materials.

>Thanks for pointing out this issue. Based on your comments, we realized that it would be better to add some metrics that could summarize the behavior of the different conditions tested and summarize the results in one table. Thus, we have made the following modifications that we believe will make it easier for the reader to understand the information.

- For figure2 and figure3, we will change the box-and-whisker diagrams to the heatmap visualization as shown in Figure R3. This is to visualize flood discharge change ratio of the grids within small sample number bins, whose information cannot be captured by box-and-whisker diagrams.

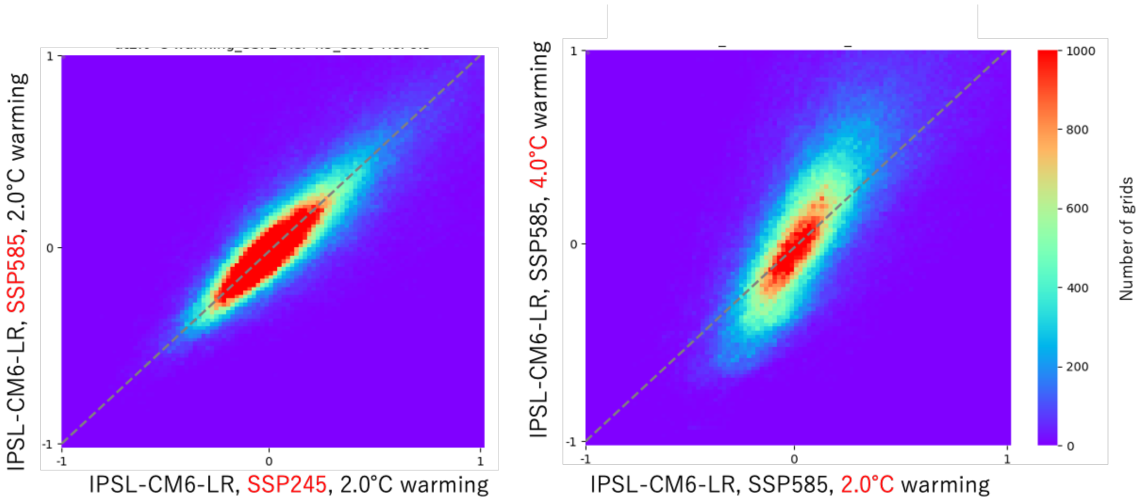

Figure R3: Scatter heatmap plot showing the similarities and differences of estimated flood discharge change ratio between two simulations((IPSL-CM6-LR) (left)among the same warming level but under different RCP-SSP scenarios, (right)among 2.0°C and 4.0 °C under SSP5-RCP8.5;

-The difference and similarity of two simulations shown in the heatmap (including additional heatmaps in Supplementary Figures) will be quantitatively analyzed and the metrics are aggregated into one table. Quantitative results (e.g., Mean Absolute Error and Pearson correlation coefficient) will be summarized as a table (Table R1), so that readers can easily know which simulations are similar/different by checking numbers in the table.

Note that we added the analysis on the comparison between “same GCM, same SSP scenario, different ensemble runs” to discuss the uncertainty due to climate internal variability. For detailed info on Table R1, please refer response to Reviewer #1.

Table R1: Quantitative evaluation of similarities between ensembles. (mean ± standard deviation of MAE between 3SSP or 3 ensemble)

	Mean Absolute Error (MAE)			Pearson correlation coefficient		
Climate Model Comparison setting	IPSL-CM6A-LR	EC-Earth3	ACCESS-CM2	IPSL-CM6A-LR	EC-Earth3	ACCESS-CM2
SSP5-8.5, different ensemble runs, 1.5℃	0.15 ±0.03	0.15 ±0.02	0.18 ±0.01	0.50 ±0.18	0.58 ±0.11	0.52 ±0.03
different SSPs, different ensemble runs, 1.5℃	0.15 ±0.03	0.14 ±0.02	0.19 ±0.00	0.50 ±0.19	0.61 ±0.10	0.49 ±0.02
SSP5-8.5, different ensemble runs, 2.0℃	0.22 ±0.05	0.16 ±0.02	0.20 ±0.00	0.48 ±0.23	0.65 ±0.10	0.57 ±0.02
different SSPs, different ensemble runs, 2.0℃	0.21 ±0.05	0.16 ±0.02	0.20 ±0.00	0.49 ±0.24	0.65 ±0.09	0.58 ±0.02
SSP5-8.5, different ensemble runs, 1.5℃vs2.0℃	0.22 ±0.03	0.17 ±0.02	0.22 ±0.01	0.40 ±0.21	0.60 ±0.07	0.49 ±0.04
SSP5-8.5, different ensemble runs, 1.5℃vs3.0℃	0.25 ±0.02	0.22 ±0.01	0.27 ±0.02	0.38 ±0.13	0.57 ±0.04	0.42 ±0.05

5. Lines 191-192: These lines further reinforce the doubts expressed at point (1). I think it is not statistically, hydrologically and hydraulic accurate to consider the coupling of three 30-years time-series as representative for 90-years. This should be highlighted as a limitation: not a larger ensemble is required, but a longer one (compare to lines 329-331).

>As mentioned above, existing global warming flood projections, such as those by Hirabayashi et al. (2013) and Dottori et al. (2018), often use a 30-year sample to predict low-frequency floods, such as 1-in-100-year floods. Following a brief discussion of existing findings and assumptions in the review responses above, we acknowledge the limitation, as pointed out by the reviewer, that it is challenging to analyze low-frequency, high-impact floods, such as 1-in-150-year events, within the narrow window of a 30-year sample.

We also would like to note that we cannot use one 90-year time series because the flood characteristics are likely to be affected by the climate change during the 90 years. Thus, we combined multiple 30-year time series at the same degree of global warming, where the flood characteristics are considered to be consistent. (Please also see the reply to Reviewer #1)

6. The effectiveness of the methodology is evaluated by applying the unbiased variance computation. However, only qualitative results are shown (Figure 5 reports the comparison and the difference of the unbiased variance). Quantitative comparison should be performed, to highlight the efficacy of the proposed methodology.

>We already conducted quantitative analysis in Figure 5 by comparing the difference of unbiased variance between the proposed method and using one SSP scenario in the original manuscript. However, as the reviewer pointed out, we realized that adding more quantitative analysis would help readers to interpret our analysis results. In the revised manuscript, we will use heatmap by evaluating different conditions tested. In addition, the heatmap information for each model (Figure R1, Supplementary Figures are also to be revised) was aggregated into one table (Table R1), and quantitative comparisons are made for different conditions tested, using MAE (Mean Absolute Error) and Pearson correlation coefficients as indicators.

Minor comments

1. Typo at line 26: “Occillation” should be “Oscillation”

>In the revised manuscript, we will revise it.

2. Line 106: it is not clear how the 100-years return period discharge is computed

>100-years return period discharge is described in lines 103-113 of the original manuscript. To make it easier for the reader to understand, we will provide a detailed method on how to calculate 100-years return period discharge as follows in the revised manuscript.

(1) We calculated specific warming levels (SWLs) as the year each SWL first surpassed a reference temperature relative to the preindustrial period (1850–1900), using a running mean of the 30-year global averaged annual mean temperature (Supplementary Tables S1).

(2) Next, as to each grid, we fitted the Gumbel distribution to the annual maximum discharge of the 30-year used for calculating above SWLs with the L-Moments method (Hosking, 2015).

(3) Then, 100-years return period discharge for each grid point was calculated from the Gumbel distribution.

3. Line 174: it should be “Figure 1” and not “Figure 2”

>In the revised manuscript, we will revise it.

4. Figure 2: “correlation coefficient” of what?

>Thanks for raising issues on this ambiguous point. In the original manuscript, we used the “coefficient of determination” (not Pearson correlation coefficient). In figure2 of the original manuscript, we drew a box-whisker plot delimited to about 30 Bins in order to compare the flood change rate at the same warming level obtained from different SSP-RCPs in IPSL-CM6A-LR. We used the median of X and the median of Y for each Bin to compute an approximate straight line. The coefficient of determination of the approximate line is defined as R.

However, in the revised manuscript, figure2 will be revised to heatmap, and “coefficient of determination” will not be calculated in new figure2. Instead, we used quantitative metrics (e.g., Mean Average Error and Pearson correlation coefficients), which can show the similarities or difference between the two simulations.

5. Line 183: “is larger for 3.0 than for 2.0”. Please write that you refer to 2.0°C and 3.0°C. The same for “1.5 and 2.0” in the subsequent lines.”

>In the revised manuscript, we will revise it.

References:

Dottori, F., Szewczyk W. et al. Increased human and economic losses from river flooding with anthropogenic warming. Nature Clim Change 8, 781-786 (2018).

Hirabayashi, Y., Mahendran, R., Koirala, S. et al. Global flood risk under climate change. Nature Clim Change 3, 816–821 (2013). https://doi.org/10.1038/nclimate1911

Hirabayashi, Y., Tanoue, M., Sasaki, O. et al. Global exposure to flooding from the new CMIP6 climate model projections. Sci Rep 11, 3740 (2021). https://doi.org/10.1038/s41598-021-83279-w

Kita, Y., & Yamazaki, D. (2023). Uncertainty of internal climate variability in probabilistic flood simulations using d4PDF. Hydrological Research Letters, 17(2), 15-20.

Winsemius, H., Aerts, J., van Beek, L. et al. Global drivers of future river flood risk. Nature Clim Change 6, 381–385 (2016). https://doi.org/10.1038/nclimate2893

Yamazaki, D., S. Kanae, H. Kim, and T. Oki (2011), A physically based description of floodplain inundation dynamics in a global river routing model, Water Resour. Res., 47, W04501, doi:10.1029/2010WR009726.