Reply to the comments by Reviewer #1:

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

Reviewer #1

The manuscript proposes a statistical method to reduce the uncertainty in flood hazard projections by integrating multiple SSP-RCP scenarios. The authors suggest that this approach can mitigate the limitations posed by a small number of ensemble members in future climate projections, particularly for flood discharge assessments. However, the novelty of the approach is overshadowed by methodological and conceptual issues:

> Thank you very much for your constructive comments which are very helpful for improving the manuscript. We planned 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, The core methodology of integrating multiple SSP-RCP scenarios to increase ensemble size is not convincingly justified. The assumption that different SSP-RCP scenarios can be combined as if they were additional ensemble members is problematic due to the inherent differences between scenarios. SSP-RCPs represent fundamentally different socio-economic pathways and climate forcing trajectories, influencing climate variables in distinct ways. The manuscript does not provide a robust theoretical or empirical basis to support this integration method. Although past studies (referred in this paper) indicate that uncertainty can be reduced by increasing ensemble size, they achieve this by using a wide range of initial conditions and climate model physics, grounded in physical principles rather than statistical manipulation. The findings in this paper may be contingent upon the specific GCM product used.

> Thank you very much for appropriately understanding the main idea of our research and providing suggestions for improving our explanation.

We begin by describing the assumptions of this study and the previous findings that support them in "[1] Strong relationship among temperature rise, hydrological cycle and flood occurrence", and then explain how we determined that the uncertainty was reduced when the ensemble size was increased in "[2] Sources of uncertainties in climate model projections, and use of ensemble data to reduce uncertainties ". Then, we describe how we handled these uncertainties in "[3] Approach to handle uncertainties in this study".

[1] Strong relationship among temperature rise, hydrological cycle and flood occurrence

This study assumed that similar future projections would be obtained for identical temperature increases. To test this assumption, we first verified that the same temperature increase produced similar changes in the hydrological cycle and flood occurrence across multiple SSP-RCP scenario combinations. Next, we

examined whether the closer predictions (i.e., reduced uncertainty) observed with increasing ensemble sizes were driven by differences in initial values from the same GCMs (internal variability), or by larger differences arising from varying SSP-RCP combinations for the same temperature.
The proposal that 'ensembles of different SSP-RCPs can be integrated at the same temperature rise' builds on the existing finding that 'identical radiative forcing produces similar temperature increases,' as well as on the assumption from previous studies that, ‘for the same temperature rise, similar changes in direct meteorological factors—such as precipitation, hydrology, and evaporation’—are expected.

Despite the socioeconomic differences among various SSP-RCP scenarios, there is a strong correlation between surface temperature increases and the rise in accumulated greenhouse gas emissions, as well as the resulting downward radiative forcing since the industrial revolution (IPCC WGI Figure SPM.10, see Figure R1). Even in studies using the CMIP6 experiment, differences in RCP (i.e., radiative forcing) dominate when comparing various SSP-RCP scenarios. For example, (SSP370 where the atmospheric aerosol concentration settings differ from other scenarios) a strong correlation has been observed, with larger RCPs leading to greater changes in temperature and precipitation (see Figure 1 of Shiogama et al., 2023, Nature Climate Change). Thus, it has been highlighted that the larger the RCP, or radiative forcing, the greater and the similar impact on temperature and precipitation changes, regardless of the SSP.

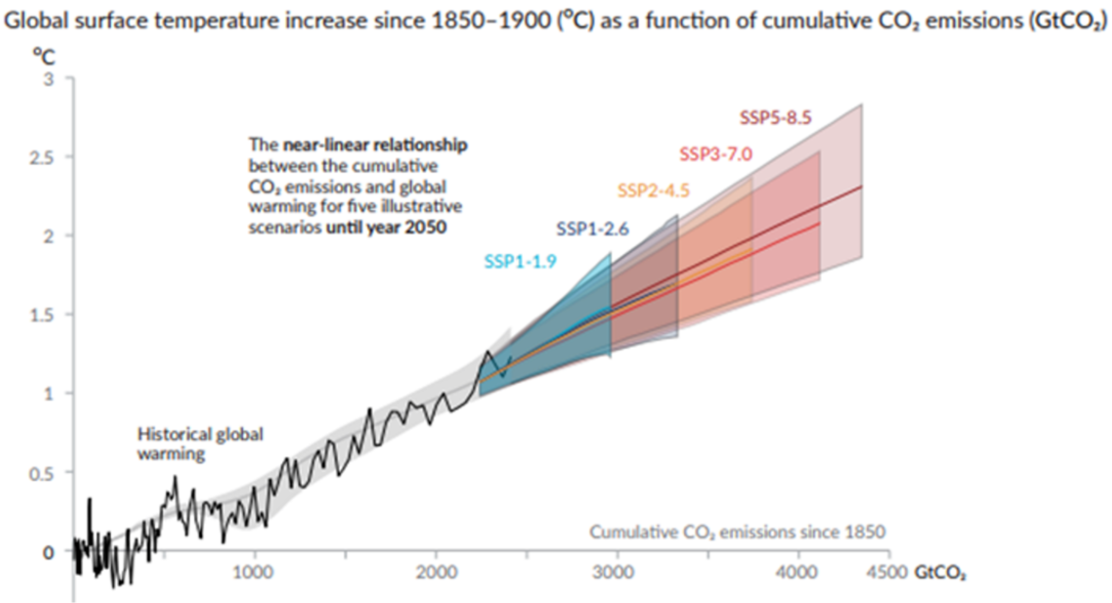

Figure SPM.10 | Near-linear relationship between cumulative CO_2 emissions and the increase in global surface temperature
Top panel: Historical data (thin black line) shows observed global surface temperature increase in °C since 1850–1900 as a function of historical cumulative carbon dioxide (CO_2) emissions in $GtCO_2$ from 1850 to 2019. The grey range with its central line shows a corresponding estimate of the historical human-caused surface warming (see Figure SPM.2). Coloured areas show the assessed *very likely* range of global surface temperature projections, and thick coloured central lines show the median estimate as a function of cumulative CO_2 emissions from 2020 until year 2050 for the set of illustrative scenarios (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5; see Figure SPM.4).

Figure R1: Schematic diagram summarizing the existing finding on 'identical radiative forcing produces similar temperature increases. (Figure taken from “Figure SPM10 of IPCC AR6 WG1 Summary for Policymakers”; IPCC, 2021)

Next, we want to highlight that several studies have shown that changes in average precipitation (Shiogama et al., 2013), heavy rainfall (Alexander et al., 2006), river floods (Hirabayashi et al., 2013; Figure R2), and the cryosphere (IPCC, 2019, SROCC) are highly correlated with temperature changes. Lehner et al (2020) described that “global mean precipitation projections remain almost identical between the different model generations”, and showed a strong correlation of precipitation with temperature increase in CMIP6. Various existing Integrated Assessment Models (IAMs) propose equations to estimate the damage caused by environmental impacts of global warming, based on the assumption that the same temperature increase will result in the same level of impact (damage). Consequently, numerous studies have linked changes caused by climate change to the rise in temperature since the Industrial Revolution. Building on these previous researches, this study assumes that ‘similar changes in river flooding could be expected with the same temperature increase’.

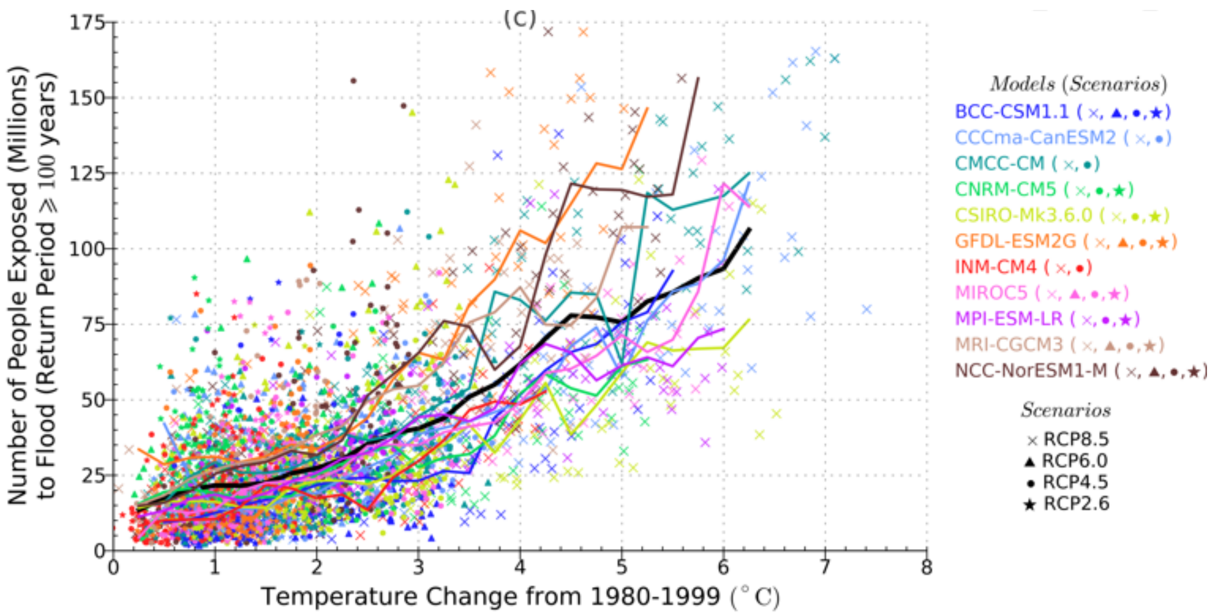

Figure R2: Schematic diagram summarizing the existing finding on 'identical temperature rise produces similar flood exposure increases. (Figure taken from Hirabayashi et al. 2013, Nature Climate Change)

[2] Sources of uncertainties in climate projections, and use of ensemble to reduce uncertainties

Uncertainties in climate projections from GCMs arise from three main sources: 1) internal climate variability (initial values), 2) model uncertainty, and 3) uncertainty in future emission scenarios. Among these, 1) internal climate variability plays a significant role in reducing the uncertainty of projections at low warming levels (e.g. 1.5 degree or 2 degree warming), where the differences among scenarios are small due to the limited response of the climate system to near-term greenhouse gasses and other socioeconomic changes. This finding is supported by the analysis of output results from many climate models. For example, Hawkins and Sutton (2009) and Wu et al. (2024) demonstrate that at low warming levels, prediction uncertainty is primarily driven by internal climate variability. As shown in Lehner et al. (2020), the uncertainty due to internal climate variability becomes relatively smaller compared to other sources of uncertainty as projections extend further into the future (see Fig 2 of Lehner et al 2020).

Regarding climate model uncertainty, it has been empirically confirmed that ensemble averages of model results computed across different institutions are closer to observations than models from any single institution (e.g., Gleckler et al, 2008; Sanderson and Knutti, 2012). Averaging multi-model outputs is expected to offset specific limitations of individual GCMs, such as challenges in representing complex processes and parameterizations, and to help achieve reasonable climate fields, as all climate models are designed to replicate the same Earth climate system in various aspects. Assessment of the impact of higher temperatures, such as those above 3°C, may not be reached by some GCMs or SSP-RCPs. Expanding the ensemble by incorporating SSP-RCP scenarios is expected to not only reduce uncertainty in internal variability but also increase the number of models capable of predicting higher temperature increases."

[3] Approach to handle uncertainties in this study

In this study, we propose the ensemble integration of results from different SSP-RCP scenarios, using the same climate model and assuming similar trends for the same temperature rise, can help reduce the uncertainty associated with internal climate variability, especially at relatively low temperature increases. It is also expected to contribute to reducing uncertainty at higher temperature projection to increase the number of models. As mentioned above, our idea is based on and supported by the previous findings on the relationship between the temperature rise and hydrological cycle changes, and on the source of uncertainties in climate projections.

As the reviewer points out, properly citing the assumptions made and the previous studies that support those assumptions, explaining why different SSP-RCP experiments produce similar flood changes at the same temperature rise, and determining whether the reduction in uncertainty can be confirmed by increasing the number of ensemble experiments with the same model would help clarify the results. In the revised manuscript, we will more carefully explain these above-mentioned background hypotheses and will provide additional analysis focusing on the source of uncertainty (see response to Comment 2 below).

2, While the manuscript aims to reduce uncertainty, it does not adequately address the propagation of uncertainties from various sources, and the uncertainty reduction is not clearly shown. Integrating different SSP-RCPs might introduce new uncertainties, and the manuscript lacks a comprehensive analysis of how these new uncertainties are quantified and managed.

> Thank you very much for pointing out the limitations of the analysis of the original manuscript. In the original manuscript, we discussed that the potential source of uncertainty is likely due to climate internal variability, by visually showing a global map of the projected uncertainties among different SSP-RCP

scenarios. As the reviewer pointed out, we realized that the quantitative analysis was not adequate.

In the revised manuscript, we will include the description of uncertainty in future projections in the main text (as described in reply to above comment). In addition, we will add a quantitative analysis to determine the source of the uncertainties in future flood risk change projections, as shown below.

[1] Use of heatmap plots to show the similarity and difference of two simulation

To quantitatively evaluate the similarities and differences of estimated flood discharge change ratio between two simulations using different GCM input, we made a scatter heatmap plot as shown in Figure R3. Here, the flood discharge change ratio of one simulation is compared to another simulation at each grid. In the top panels of Figure R3, the flood discharge change ratios of the simulations with "Same SSP scenario, different ensemble runs" are compared (note: same model at 1.5°C warming). Thus, differences of two simulations are due to the climate internal variability. Then, in the bottom panel, we compared the flood discharge change ratio between the simulations with "different SSP scenarios, different ensemble runs". Thus, the difference in the bottom panel is considered to be due to both "climate internal variability and scenario difference".

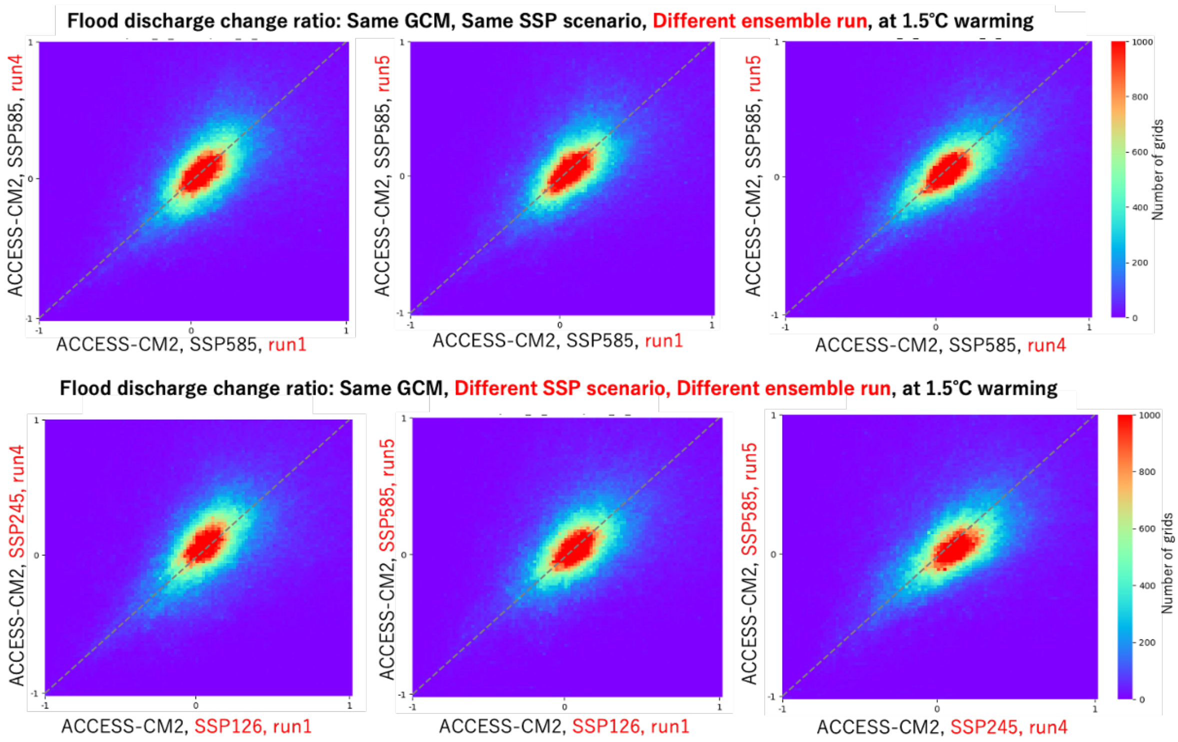

Figure R3: Scatter heatmap plot showing the similarities and differences of estimated flood discharge change ratio between two simulations using different SSP scenarios and different ensemble run input.

[2] Quantitative evaluation of the similarity and difference between two simulations using metrics

Then, we evaluated the difference and similarities of the two simulations by calculating Mean Absolute Error and Pearson Correlation Coefficient. These metrics are calculated for the three possible combinations of two simulations (i.e., three panels in top), and the mean and standard deviation of these metrics are listed in Table R1. We calculated the metrics for three different climate models (ACCESS-CM2, EC-Earth3, and IPSL-CM6A-LR) and at 1.5°C and 2.0°C warming levels.

We found out that the Mean Absolute Errors are almost the same between "comparison of same SSP and different ensemble runs" and "comparison of different SSP and different ensemble runs". This indicates that the uncertainties due to scenarios are almost negligible compared to the uncertainties due to climate internal variability, at least at 1.5°C and 2.0°C warming level. The results for Pearson Correlation Coefficient suggested the same pattern. These results support our idea of integrating different SSP simulations to increase ensemble size, given that the differences are mostly due to internal climate variability, even in different SSP simulations, when the warming level is the same.

We also did the same analysis for the comparison of different warming levels under the same RCP scenario, and found that the difference due to the warming level is larger than the uncertainties due to climate internal variability, for all climate models.

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

3, The manuscript claims that the proposed method reduces unbiased variance in 70% of land grid points. However, the validation of these claims is insufficient. There is a lack of independent verification using observed data or alternative high-resolution models. Without robust validation, the reliability and applicability of the proposed method remain questionable. At the very least, historical or present climate states should be used for validation purposes, even if the main objective is uncertainty analysis. The large uncertainties in the complex climate-discharge system might make the results meaningless, highlighting the need for validation.

> We tried to keep the manuscript concise to focus on our main discussion, but as reviewer pointed out, minimum description of the status of validations in previous studies will be added in the revised manuscript. Please see the background information below.

Firstly, verification of model output is not conducted here, because the aim of this study is to demonstrate a method to reduce variation (uncertainty) in projected floods at each location and/or in each river. The purpose is not to confirm the existence of a climate-change signal (globally or in a specific region) or to validate the simulation of flooding within the model framework used in this study. Secondly, the climate and river model frameworks for estimating future flood changes due to climate change have been widely discussed in various previous studies, particularly regarding their uncertainties and the robustness of future projections.

[1] Validation of the global river model skill in previous studies

For the CaMa-Flood model used in this study, consistency with historical river level, flow, and inundation area data has been demonstrated (Zhou et al., 2024). Additionally, uncertainty among models and scenarios for future projections has been examined (Hirabayashi et al., 2013; Hirabayashi, 2021; Kimura et al., 2023).

[2] Evaluation of the future flood risk projection using global river models.

We agree with the comment that estimating future changes in river discharge, particularly extremes, is a complex process. Nevertheless, previous studies have shown that future changes are, to some extent, highly consistent regardless of the GCM or climate scenario. For instance, under the same RCP (radiative forcing) scenario, flood change trends in CMIP5 and CMIP6 are very similar, even when different sets of climate models and SSP scenarios are used (Figure R4). Moreover, when the same set of GCMs is selected, the projected flood exposure for the same temperature increase is also shown to be similar, even between CMIP5 and CMIP6 with different SSP-RCP combinations (see Figure R5).

In addition, a CMIP4 warming experiment which used a different future scenario (A1B) and assumptions, along with a relatively higher spatial resolution (about 100 km), also shows similar spatial distributions of future flood projections (Figure 6 of Hirabayashi et al., 2008, Hydrological Sciences Journal). Based on these previous studies, we assume that river discharge projections in GCMs exhibit greater consistency - significantly different from analyses of ordinary climate variables, such as local precipitation. This is because large river systems integrate information from upstream grid cells, meaning that changes in river floods are driven by relatively larger-scale changes in climate variables, including precipitation, temperature, evapotranspiration, and snow processes.

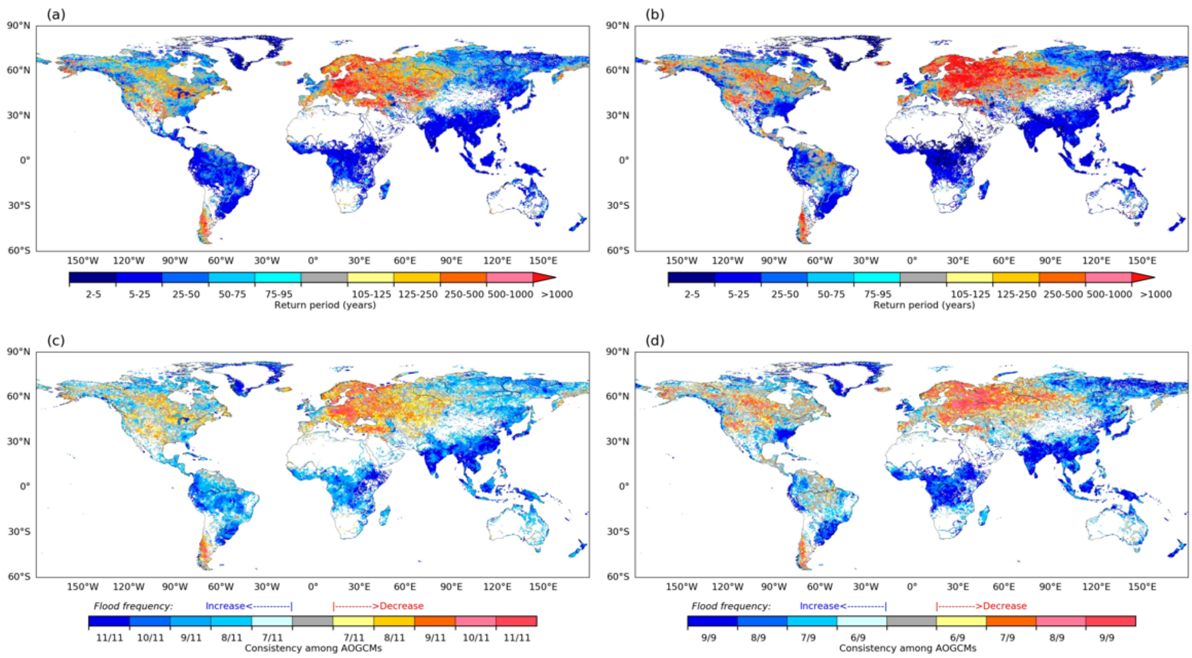

Figure R4: Schematic diagram summarizing the existing finding on 'flood projection is similar between simulations under the same radiative forcing (RCPs) with different SSP assumptions (CMIP5 and CMIP6).' Figure taken from Hirabayashi et al. 2021, Scientific Reports.

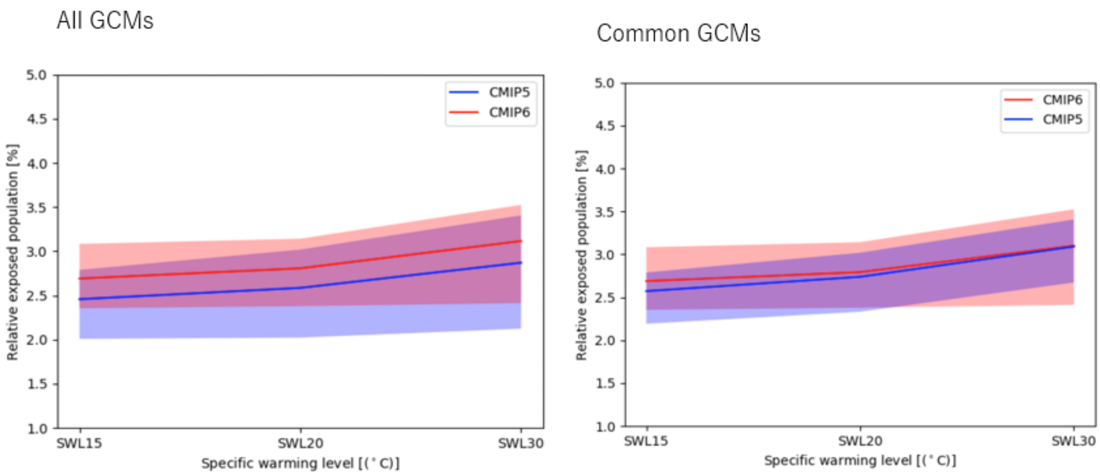

Figure R5: Schematic diagram summarizing the existing finding on "identical temperature rise produces similar flood exposure increases.' Figure taken from Hirabayashi et al. 2021, Scientific Reports.

Against this background, we conclude that there are certain locations where the current model framework can robustly project future floods. We then analyzed the locations where predictions among GCMs do not agree and investigated whether increasing the number of ensembles could reduce the uncertainty in those projections.

We will briefly summarize this information in the revised manuscript and clarify that this study focuses on analyzing uncertainty in future flood projection.

4, Several key explanations in the manuscript are unclear or insufficiently detailed. For instance, the process of determining the similarity of flood hazard projections among different SSP-RCP scenarios is not described in enough detail to be reproducible. Specific queries include: How are the model boundary conditions and initial conditions determined? Is there consideration of evapotranspiration and infiltration? How is the river conveyance capacity estimated? Is there any downscaling method needed to solve the scale mismatch between the coarse resolution climate simulation and the fine resolution hydraulics needed? How about the bias correction in these climate model simulations?

> We will explain the sources of uncertainties in climate model projections (see above, Answer 1) and the potential reasons of the similarity of flood hazard projections among different models/SSP-RCP scenarios (see above, Answer 3) in the revised text.

Additionally, we will revise our manuscript to include the following clarification:
"River streamflow is calculated by integrating GCM runoff with the river model. The runoff itself is calculated by the vertical water and energy balance within the GCM land surface process. The coarse spatial resolution of the GCM runoff is resampled to match the finer resolution topographic data. Bias correction and downscaling are performed simultaneously by calculating inundation depths using the Lookup Method of Kimura et al (2023), without applying explicit bias correction."

5, A typo in line 159, there is an extra).
> Thanks for finding this. We will revise it.

References:

Alexander, L. V., and co-authors, 2006: Global observed changes in daily climate extremes of temperature and precipitation, J. Geophys. Res., 111, D05109, doi:10.1029/2005JD006290.

Gleckler, P.J., Taylor, K.E., Doutriaux, C. Performance metrics for climate models. J. Geophys. Res., 113, D06104. doi:10.1029/2007JD008972.

Hawkins, E. and Sutton, R., The Potential to Narrow Uncertainty in Regional Climate Predictions, Bulletin of the American Meteorological Society, Vol. 90(8), 1095-1108 (2009) (doi:10.1175/2009BAMS2607.1).

Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., & Kimoto, M. (2008). Global projections of changing risks of floods and droughts in a changing climate. Hydrological Sciences Journal, 53(4), 754–772. https://doi.org/10.1623/hysj.53.4.754

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

IPCC, 2019: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M.

Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, 755 pp. https://doi.org/10.1017/9781009157964.

IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32, doi:10.1017/9781009157896.001.

Kimura, Y., Hirabayashi, Y., Kita, Y., Zhou, X., and Yamazaki, D.: Methodology for constructing a flood-hazard map for a future climate, Hydrol. Earth Syst. Sci., 27, 1627–1644, https://doi.org/10.5194/hess-27-1627-2023, 2023.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E.M., Brunner, L., Knutti, R., Hawkns, E. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth Syst. Dynam. 11, 491-508 (2020). https://doi.org/10.5194/esd-11-491-2020

Sanderson, B.M. and Knutti R. On the interpretation of constrained climate model ensembles. Geophysical Research Letters 39, L16708 (2012). doi:10.1029/2012GL052665.

Shiogama, H., Fujimori, S., Hasegawa, T. et al. Important distinctiveness of SSP3–7.0 for use in impact assessments. Nat. Clim. Chang. 13, 1276–1278 (2023). https://doi.org/10.1038/s41558-023-01883-2

Wu, Y., Miao, C. Slater, L., Fan, X., Chai, Y. Hydrological projectins under CMIP5 and CMIP6: Sources and magnitudes of uncertainty. BAMS (2024). https://doi.org/10.1175/BAMS-D-23-0104.1.

Xudong Zhou, Dai Yamazaki, Menaka Revel, et al. Benchmark Framework for Global River Model (Version 1.0). ESS Open Archive . April 08, 2024. doi: 10.22541/essoar.171255743.38640796/v1