Responses to Referee #1

We appreciate the feedback provided by the anonymous reviewer. We respond by highlighting the reviewer's comment in shaded text and our responses in black. The resulting changes from our responses to referee #1 are highlighted in blue in the revised manuscript.

Comments:

1. The authors have done a commendable job in addressing my comments in the revised submission and the overall quality of the presentation has improved substantially. I have one minor suggestions below.

Thank you for your positive feedback and recognition of the improvements in our revised submission. We are pleased that the additional tests we conducted based on your previous comments have enriched the manuscript and bolstered the robustness of our results.

2. L306-314 and Appendix B: The corresponding results that demonstrate the robustness of the choice of parameters to process S/N as the authors presented in the "response to reviewers" should be summarized and included in the paper as well.

Following the reviewer suggestion, we add more details about the sensitivity tests that guided our choices in Appendix B and section “Discussion and conclusions”.

Appendix B now includes one of the test figures and the text was modified as follows:

L415-419: “Our estimation of the S/N ratio follows the method proposed by Hawkins and Sutton (2012), with a thorough parameter selection process that includes sensitivity tests, which guided the rationale for our choices. The key sensitivity test includes the evaluation of different filters, window types, and window sizes (as exemplified in Fig. B1). The tested filters are Butterworth, Chebyshev, Elliptic, FFT low-pass, and Lowess. The window types are Rectangular, Hamming, Bartlett, Blackman, and Hanning. The window lengths are 21, 31, 41, 51, and 61 years. The key findings are summarized as follows:”

L427-434: “… Filters that produce phase distortion can exaggerate differences between the original and filtered time-series, leading to a misrepresentation of the noise term with a higher value. For example, N ranges from 1.18 to 1.23 for filters that introduce phase distortion, whereas N=1.04 for the Lowess filter (Fig. B1a). Similarly, the “blending” of the FFT smoothed signal on the edges of the time-series may make the signal to unrealistically emerge or immerse on the natural variability range. The type and length of the window have a relatively minor impact in comparison. In summary, while we recognize that the selection of window length and smoothing options involves some subjectivity, the resulting N term exhibits relatively low sensitivity to reasonable choices, varying less than 9 % among the entire set of window types and lengths tested for the global signal (Fig. B1b-c). …”

L440-445: "… Lastly, alternative methods, such as estimating local signals by filtering local time-series, can produce misleading results. For instance, we tested this alternative for three rivers with different regional trends in their future mean flows: Congo (positive), Amazon (neutral), and Negro (negative). The results reveal that local filtered signal is affected by the filter's "blending" effects on extremes, regardless of regional trends. Moreover, this "blending" yields to high S/N ratio values in the first year of simulation, which is unrealistic. This issue is avoided when linear regressions are applied."

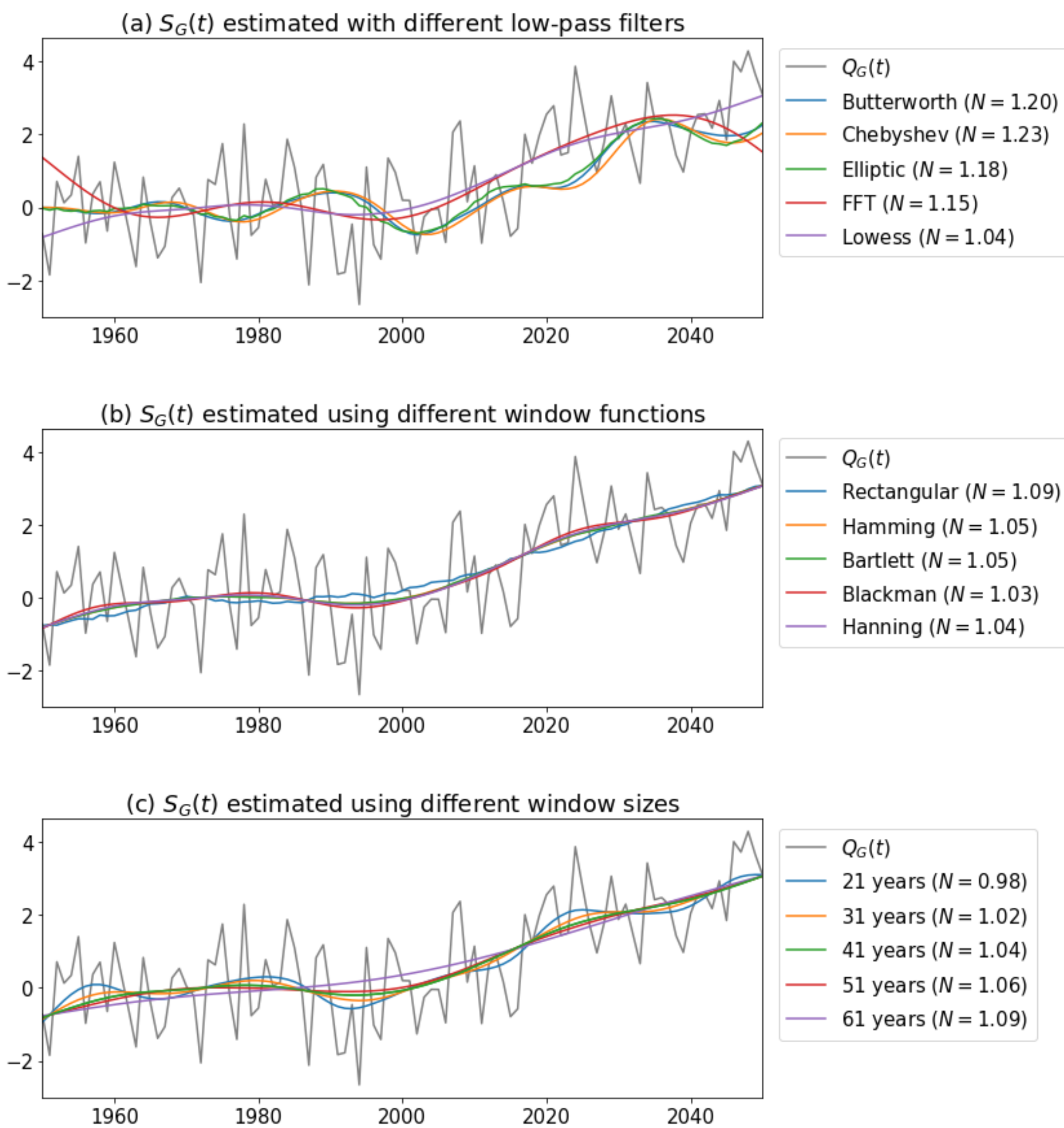


Figure B1. Low-pass filter selection test for the global river discharge anomaly time-series (Q_G) using the CNRM-CM6-1-HR model as a case study. The signal S_G is estimated with various (a) filters, (b) window functions, and (c) window sizes. The (b) and (c) tests are applied for the Lowess filter. The resulting noise value of each filter is provided in the legend.

On the other hand, the paragraph associated to the filtering process in the section "Discussion and conclusions" now states:

L311-325: "The assessment of hydrological simulations involved estimating the S/N ratio and ToE to determine, for which rivers of the world, and when, the climate change signal will emerge from the natural variability. The method used for the calculation of the signal (both global and local) is key, as it determines the noise terms and the ToE. On one hand, the selection of filter, window function, and window length is key for estimating the global signal. Guided by rigorous sensitivity tests of various options, our analysis favored the use of a Lowess filter with a Hanning window of 41-year length. Our results indicated that the choice of filter significantly impacts the signal, as some filters introduce phase distortion and/or a "blending" effect, leading to erroneous high noise values that alter the ToE. These drawbacks are minimized by the Lowess filter, which effectively emphasizes long-term variations without introducing any distortion. Despite the subjectivity inherent in these choices, our analysis revealed that the resulting noise term exhibited low sensitivity to reasonable options of window function and length. A Hanning window of 41-year length was chosen as it effectively emphasizes long-term variations, essential for capturing climate change signals amidst natural variability. On the other hand, local signals can be estimated through linear regression of the global signal or by filtering local time-series. Both alternatives were tested for rivers with different regional trends (positive, neutral, and negative). The results revealed that the local filtered signals are affected by the filter's "blending" effects on extremes, regardless of regional trends. However, this issue was effectively mitigated when linear regressions were applied. These methodological considerations enhance the reliability of our S/N ratio estimation for assessing ToE."

Responses to Referee #2

We thank the new revision offered by the anonymous referee. We respond by highlighting the reviewer's comment in shaded text and our responses in black. The resulting changes from our responses to referee #2 are highlighted in green in the revised manuscript.

Main Comment:

1. The manuscript "River flow in the near future: a global perspective in the context of a high-emission climate change scenario" presents a study of the possible effects of global warming in a high-emission scenario on river flows over the next few decades using model intercomparison. Overall, the goal of the paper is clear, and the writing is well organized. The regional implications of the results are detailed.

We appreciate your recognition of the clarity and organization of our paper, as well as your suggestions for further clarification and improvement.

2. However, my main concern is the lack of comparison of the study's results with similar work (e.g., www.nature.com/articles/s44221-023-00030-7, www.nature.com/articles/s41597-022-01410-6, doi.org/10.5194/hess-21-4379-2017). It is hard to identify the novelty of the study without comparison. Therefore, I would suggest that in the discussion section, instead of focusing on the regional implications again (as some are already stated in Section 3.4), the authors could make a comprehensive comparison of the results with the literature and indicate and interpret the possible disagreement and how it may be related to the methodology used, which may be more valuable for readers.

Thanks for your comment. In response to the lack of comparison with other studies, and the need to reinforce the novelty of our research, we have expanded the Discussion section with the following text (L384-398): “Recent studies reveal significant advancements in understanding the complex interplay between climate change and river flow projections. Bosmans et al. (2022) introduced a high-resolution dataset projecting global river flow under various climate scenarios, which resembles the patterns observed in our findings. Zhou et al. (2023), in agreement with our projections, attribute changes in runoff to shifts in land surface characteristics such as vegetation and soil conditions. Zhang et al. (2023) predict similar river flow anomaly patterns to those found in our work but suggest that global river flow may be lower than projected by GCMs, attributing this discrepancy to the heightened sensitivity of river flow to changes in evapotranspiration, linked to the phenomenon of vegetation greening. Our work complements these studies by employing advanced techniques such as the S/N ratio and the ToE, which are key for identifying when river systems may exhibit conditions beyond their known historical variability. Our findings underscore the pressing need for a paradigm shift in prioritizing water-related concerns in the context of climate change, as emphasized by Douville et al. (2022). Moreover, our study emphasizes the interplay between water cycle alterations and potential hydrological impacts, providing valuable insights for planning purposes. It is concerning that several major rivers are projected to imminently surpass the bounds of their natural variability. However, the hydrological predictions presented in this work should be interpreted in the context of a very high baseline emission scenario, i.e., an outcome only likely if society does not make concerted efforts to reduce greenhouse gas emissions (Van Vuuren et al., 2011). In future work, we will extend the analysis to encompass a broader set of the new SSP scenarios."

We would like to clarify that while Section 3.4 evaluates model projections in regions with strong agreement, the Discussion section delves into real-world consequences (recent observed and potential future impacts) in such regions. Both sections offer distinct perspectives on the topic.

Lastly, we respectfully argue against comparing with Papadimitrious et al. 2017. Their research focuses on biases of runoff generated with uncoupled land surface model simulations forced by GCM atmospheric variables, which do not account for feedback from the land surface to the atmosphere, thereby introducing uncertainty into the simulated runoff fields. In contrast, our study we use the simulated runoff from couple GCMs simulations to force a routing model, which transports the runoff from its origin to the river channels, without altering the land-atmosphere water balance. This fundamental difference precludes extending their findings about runoff biases to our simulations, a point deeply evaluated in Section 3.1 of our manuscript.

Bosmans, J., Wanders, N., Bierkens, M. F., Huijbregts, M. A., Schipper, A. M., and Barbarossa, V. (2022). FutureStreams, a global dataset of future streamflow and water temperature. *Scientific data*, 9(1), 307.

Papadimitriou, L. V., Koutroulis, A. G., Grillakis, M. G., and Tsanis, I. K. (2017). The effect of GCM biases on global runoff simulations of a land surface model. *HESS*, *21*(9), 4379-4401.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G., Kram, T., Krey, V., Lamarque, J.-F, Masui, T., Meinshausen M., Nakicenovic, N., Smith, S., and Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic change*, *109*, 5-31.

Zhang, Y., Zheng, H., Zhang, X., Leung, L. R., Liu, C., Zheng, C., Guo, Y., Chiew, F., Post, D., Kong, D., Beck, H., Li, C., and Blöschl, G. (2023). Future global streamflow declines are probably more severe than previously estimated. *Nature Water*, 1(3), 261-271.

Zhou, S., Yu, B., Lintner, B. R., Findell, K. L., and Zhang, Y. (2023). Projected increase in global runoff dominated by land surface changes. *Nature Climate Change*, *13*(5), 442-449.

Minor Comments:

1. L24: In some places, ET has been shown to be more influenced by vegetation change (e.g., https://www.nature.com/articles/s43017-023-00464-3).

Thanks for noting it. This is now reflected in L24-L26:

"… Evapotranspiration has changed in response to changes in precipitation and warmer temperatures, as well as to the observed vegetation greening in northern high latitudes (Yang et al. 2023), altering the ability of the soil to hold moisture. …"

Yang, Y., Roderick, M. L., Guo, H., Miralles, D. G., Zhang, L., Fatichi, S., Luo, Y., McVicar, T., Tu, Z., Keenan, T., Fisher, J., Gan, R., Zhang, X., Piao, S., Zhang, B., and Yang, D. (2023). Evapotranspiration on a greening Earth. *Nature Reviews Earth & Environment*, 4(9), 626-641.

2. L38: It would be beneficial if the author could clarify the difference between "runoff" and "streamflow" here.

The sentence now states: "This uncertainty extends to runoff, and by that to river flow, which is the local runoff that is subsequently routed from land to oceans through river channels." (see L39). Note that "streamflow" has been replaced by "river flow" (see our response to comment #4).

3. L39, It might be worth mentioning https://doi.org/10.1038/s41558-023-01659-8 here to indicate the role of land surface change as well.

We discuss the suggested paper in the revised manuscript (L39-L44): "Douville et al. (2021) and Zhou et al. (2023) concur on the projected increase in global runoff in the coming decades, albeit attributing it to different factors. Douville et al. (2021) link this rise to global warming, with confidence levels escalating with emissions scenarios. In contrast, Zhou et al. (2023) attribute it to changes in the synergistic effects of vegetation responses to rising CO2 concentrations and land surface reactions to radiative changes, which lead to a shift in precipitation partitioning towards runoff instead of evapotranspiration."

4. L42, Does "river flow" have the same meaning as the previously used "streamflow"? Please consider making the wording consistent in case of confusion.

Effectively, river flow has the same meaning than streamflow. To ensure clarity and consistency, we have replaced all instances of 'streamflow' with 'river flow' throughout the manuscript (L39, L134, L265, 295, 485).

5. L56, "Hydrological model" is confusing here, as it seems to indicate only a stand-alone routing model.

The term "hydrological model" has been replaced by "river routing model" (L59).

6. L71-73, It is hard for me to connect here the provision of resolution (I assume it is by Table 1) and the reconciliation of variability in network topologies. I think the authors wanted to point out that only 50N information is shown because of the different network topologies used.

The reviewer is correct in the understanding. We have clarified it as follows (L76-L80): "To reconcile the variety of grid topologies used by the different GCMs (rectilinear, reduced gaussian, icosahedral, etc) in the comparison of the GCMs' resolution, we provide the atmospheric horizontal resolution at 50 ° N. This mid-latitude serves as a representative point for assessing resolution, particularly given the significant variation in resolution from the Equator to the poles in models using rectilinear grids. The atmospheric resolution at 50 ° N ranges from 25 km to 134 km for the set of simulations (Table 1)."

Note that presenting the GCMs' atmospheric resolution at 50 ° N is a common practice in papers based on HighResMIP simulations (e.g. Vannière et al. 2019, Demory et al. 2020, Müller et al. 2021, among others).

Vannière, B., Demory, M. E., Vidale, P. L., Schiemann, R., Roberts, M. J., Roberts, C. D., Matsueda, M., Terray, L., Koenigk, T., and Senan, R. (2019). Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. Climate dynamics, 52, 6817-6846.

Demory, M.-E., Berthou, S., Fernández, J., Sørland, S. L., Brogli, R., Roberts, M. J., Beyerle, U., Seddon, J., Haarsma, R., Schär, C., Buonomo, E., Christensen, O. B., Ciarlo`, J. M., Fealy, R., Nikulin, G., Peano, D., Putrasahan, D., Roberts, C. D., Senan, R., Steger, C., Teichmann, C., and Vautard, R. (2020). European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci. Model Dev.,* 13, 5485–5506.

Müller, O. V., Vidale, P. L., Vannière, B., Schiemann, R., Senan, R., Haarsma, R. J., and Jungclaus, J. H. (2021). Land–atmosphere coupling sensitivity to GCMs resolution: A multimodel assessment of local and remote processes in the Sahel hot spot. *Journal of Climate,* 34(3), 967-985.

7. L87, it might be useful to explain the basic principle of the method (not necessarily with equations).

The new version of the manuscript includes a brief explanation of how the routing model works (L93-L97): "TRIPpy employs a simple advection method within a water balance model to route total runoff through the topography. This method calculates changes in river channel storage within each grid cell by accounting for the difference between the inflow, which includes both local runoff and contributions from upstream grid cells, and the outflow. The outflow is estimated using a linear function of storage, considering the river flow velocity and the river length between two connected grid cells. Detailed TRIPpy equations can be found in the appendix of Müller et al. (2021a), where..."

8. L106, Why would the range be (-inf, inf) here? Assuming all simulated flows are zero, the percentage difference would be -100% at most, or did I misunderstand the percentage difference?

Thank you for pointing out the error. Indeed, the valid range is [-100, inf). This has been corrected in the revised manuscript (L117).

9. L108, it is not clear what would be the denominator in the calculation.

The Overlapping Coefficient (OC) is a statistical measure used to compare the distributions of observed and simulated river flow time-series. It quantifies the degree of overlap between the probability distributions of the two datasets. The OC score is calculated by summing the minimum values of the relative frequencies of observations and simulations for each bin in the histograms of the respective datasets (see equation 5 in Müller et al. 2021). Essentially, it evaluates how much the histograms overlap, with a score of zero indicating no overlap and a score of one indicating complete overlap. The denominator *N* in the OC calculation is a fixed number which represents the total number of monthly observations for the monitored river, providing the length of the time-series for normalization purposes. We added an interpretation of the score in L119-120.

Müller, O. V., Vidale, P. L., Vannière, B., Schiemann, R., and McGuire, P. C. (2021). Does the HadGEM3-GC3. 1 GCM overestimate land precipitation at high resolution? A constraint based on observed river discharge. *Journal of Hydrometeorology*, *22*(8), 2131-2151.

10. L130, please specify the window width used in the study.

The sentence has been rephrased to clarify this as follows "The filter is based on the convolution of a scaled window of 41-year length with the signal, resulting in a smoothing effect of the inter-annual variability." (L140). Please note that the rationale behind this and other choices related to the filtering processes are explained in detail in Appendix B (as indicated in L147).

11. L156, I would be cautious about calling the consistency "remarkable."

The sentence has been rephrased to "Despite the diversity of these metrics, there is consistency across models, with all exhibiting values in a narrow range." in L166-L167. This modification helps soften the emphasis while still effectively conveying the idea of consistency among models.

12. L161, How is the -1.8% derived? It seems to be an average value, but it is not quite convincing for me to calculate only the algorithmic mean given the positive and negative biases present (may use the mean of the absolute values).

Thank you for your feedback. First, it's important to clarify that we mistakenly wrote -1.8% in the text, whereas the correct value is -1.5%, as accurately depicted in Fig. 1c. The -1.5% relative bias (RB) represents the difference between the river flow of the ensemble mean simulation and the observed river flow in the selected 20 largest monitored catchments. The ensemble mean simulation, being an average of simulations with varying biases (positives and negatives), inherently provides a more stable estimate of model performance, reducing the impact of outliers or biases in any single simulation. This advantage of ensemble mean simulations elucidates the RB value of -1.5%. On the other hand, our choice of relative bias over other percentage scores based on absolute differences (e.g., percentage absolute bias) serves a purpose, indicating whether simulations tend to systematically produce anomalous negative or positive river flows. This information, not provided by the other selected scores, is crucial for ensuring accurate historical river flow reproduction before trusting projections for the future.

Based on the previous explanation, we applied the following changes in the revised manuscript:

a) The value -1.8% was replaced by -1.5% in L172.
b) The term "ensemble mean" is now presented as "ensemble mean simulation" in L109, L167, L171, and in the caption of Figure 1.

c) The analysis of scores for the ensemble mean simulation is now explained as follows in L166-L169: "… Despite the diversity of these metrics, there is consistency across models, with all exhibiting values in a narrow range. The ensemble mean simulation, being an average of simulations with varying biases, inherently provides a more stable estimate of model performance, reducing the impact of outliers or biases in any single simulation. This advantage of the ensemble mean simulation explain why it outperforms most individual models, with RB=6.8%, OC=0.63, r=0.72, and npKGE=0.58. Among the top-performing models are the GCMs of the EC-Earth3P and MRI-AGCM3-2 families."

d) The description of RB now includes its interpretation justifying its choice (L116-L118): "Relative Bias (RB): Measures the percentage difference between total simulated and observed mean flow for all monitored rivers, indicating whether simulations tend to overestimate or underestimate river flows. Range: [-100, inf), Perfect score: 0."

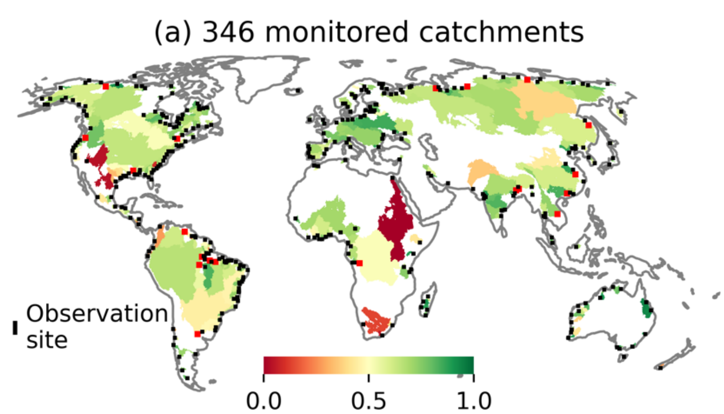
13. L167, The analysis in the paragraph is a bit disjointed from the previous paragraphs; is there some intention to compare the resolution of the model with the performance? If so, what about other models?

Thank you for drawing attention to the disjointed nature of the paragraph in question. After thorough consideration, we have opted to remove it from the manuscript. While our original intention was to reconcile our current findings with those of our previous study (Müller et al. 2021), which focused on models' resolution, we recognize that this comparison may disrupt the flow of the section and could potentially divert the attention of readers from the focus of this paper.

Müller, O. V., Vidale, P. L., Vannière, B., Schiemann, R., and McGuire, P. C. (2021). Does the HadGEM3-GC3. 1 GCM overestimate land precipitation at high resolution? A constraint based on observed river discharge. *Journal of Hydrometeorology*, *22*(8), 2131-2151.

14. Fig 1a) It may be more informative to use one of the four metrics instead of catchment size as the color.

Thank you for your suggestion. We have updated Figure 1a to reflect the overlapping coefficient (OC) of the ensemble mean simulation for each monitored river, as you recommended. The caption has also been revised accordingly.



“Figure 1: Validation of the global hydrological simulations. (a) Monitored rivers with black dots indicating the observation sites for river flow and colors showing the overlapping coefficient (OC) of the ensemble mean simulation for each monitored river. (b) …”

15. L193, what does the ITCZ mean here?

ITCZ is the abbreviation of Intertropical Convergence Zone. It is now specified in its first occurrence (L197).

16. L316, should "increased" here be "increased precipitation"?

Thank you for noting the missing word “precipitation”. It has been corrected in L327.