Responses to Referee #1

We appreciate the detailed and constructive feedback provided by the anonymous reviewer, which has been valuable in enhancing the manuscript. We respond by highlighting the reviewer's comment in shaded text and our responses in black.

General comments:

The aim of this work is to provide a global assessment of near future river flow changes within the context of a high-emission climate change scenario. The authors leverage the TRIP model to simulate river flow dynamics, which is forced by runoff from HighResMIP simulations. They then quantify the changes by using the signal to noise ratio and time of emergence metrics. I have two major concerns on this work: 1) the need for a thorough justification, validation, and demonstration of signal and noise estimations; 2) the absence of logistical and substantive content that elucidates the implications for metropolitan areas and water resource management. Please see my detailed comments below.

We appreciate the reviewer's detailed feedback and their two major concerns regarding our manuscript. The comments provide valuable insights into the clarity and robustness of our work. First and foremost, we would like to address the following:

1) We understand the importance of providing a strong rationale for our signal and noise estimations. To address this concern, we provide a comprehensive and detailed description of our signal-to-noise and time of emergence methodology in the responses to the main concerns, and explain the main points of that in the revised manuscript. This expanded explanation sheds light on the utility of these techniques for our research and includes specific validation tests, ensuring a thorough justification, validation, and demonstration of our estimations.

2) We recognize that it is challenging to establish cause-and-effect relationships between simulated climate change signals and particular extreme events and/or sectoral impacts. Following reviewer #2’s suggestion, we will replace the section “Rivers susceptible to strong changes” by “Regions susceptible to strong changes”. In this new section we analyze the discharge projections grouping rivers in four regions with (a) trends that significantly deviate from the historical period, and (b) clear consensus among the model simulations. The regions are: Central Africa, Arctic, South Asia, and Patagonia. This regional analysis provides a more concise way to convey the broader implications of our study. The information about the potential risks for individual rivers will be briefly summarized at the end of the new manuscript.

Major comments:

1. Abstract: the flow of the abstract from L9 to L14 is hard to follow. They did not conclusively reflect what has been discussed in the main text. Please consider refine the abstract.

The abstract will be rewritten to more accurately reflect the main findings of our research, in particular, those referring to the new analysis at regional scale (explained in major comment 7).

2. L77-79 and Appendix A: additional evidence is necessary to support the hidden assumption that all models share similar internal variability as HadGEM-GC31. Do the remaining models contain ensemble members of precipitation? If so, how does the precipitation spread across the ensemble members? Do they share the similar internal variability as the precipitation spread of HadGEM-GC31? Consider exploring such variables that could influence runoff to earn the statement.

We agree with the reviewer regarding the assumption of similar internal variability across all models. To address this concern, we obtained access to additional realizations of the HighResMIP-CMIP6 from the Laurence Livermore National Laboratory (LLNL) Earth System Grid Federation (ESGF) Node (esgf-node.llnl.gov), which includes total runoff among the variables. This extended our internal variability analysis to a total of 58 individual realizations across different GCMs.

The breakdown of realizations per GCM is provided in Table R1.1:

GCM	AMIP realizations	COUPLED realizations
CNRM-CM6-1	10	3
CNRM-CM6-1-HR	10	3
EC-Earth3P	3	3
EC-Earth3P-HR	2	3
HadGEM3-GC-31-L*	3	3
HadGEM3-GC-31-MM	3	3
HadGEM3-GC-31-HM	3	2
MRI-AGCM3-2-H	1	—-
MRI-AGCM3-2-S	1	—-
NICAM16-7S	1	—-
NICAM16-8S	1	—-
Total	38	20

Table R1.1: Number of realizations per GCM used to analyze the GCM's internal variability.

Note that GCMs from the MRI and NICAM families provide only one realization, limiting the extension of the analysis to these specific model families.

The internal variability results for CNRM-CM6-1 and EC-Earth3P GCMs confirm our earlier findings for the HadGEM3-GC31 family. While the internal variability tends to rise over time, it is smaller than the inter-model variability and comparable to the variability given by the GCMs' resolution. These new results strengthen the robustness of our assumption that the set of simulations is adequate for the proposed objectives and that more realizations would not substantially alter the presented results.

New panels showing the internal variability of CNRM and EC-Earth3P GCMs' families (see Figure A1) along with their interpretation will be added to the Appendix A: “Internal variability in projections of runoff anomaly”.

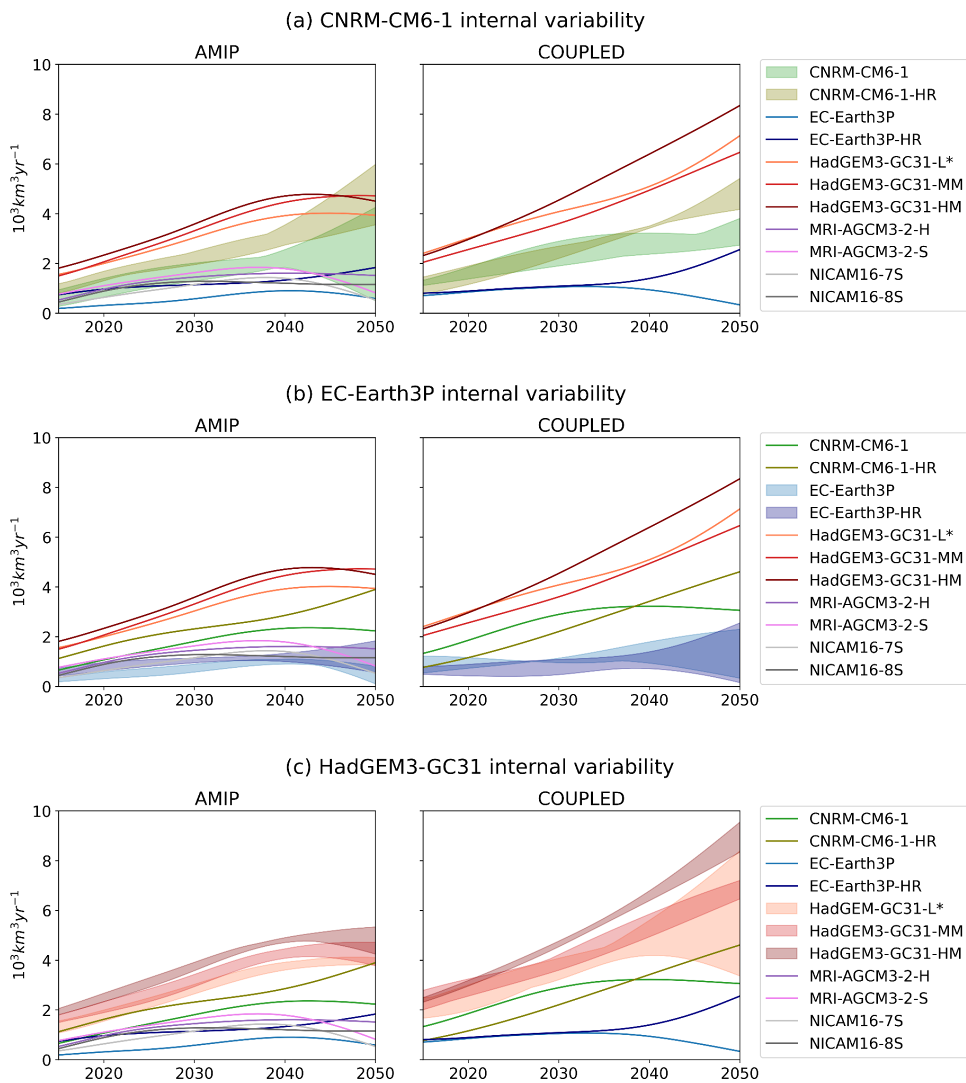

Figure A1. Low-pass filtered projections of global runoff anomalies for AMIP and COUPLED simulations of (a) CNRM-CM6-1, (b) EC-Earth3P, and (c) HadGEM3-GC31 simulations. The solid lines show the projections of individual GCMs. The shaded bands show the internal variability of the GCM realizations at different resolutions.

3. While the authors direct readers to Müller et al. (2021a) for details on the TRIP model, it would be beneficial to provide a brief overview of TRIP's global-scale performance and the rationale for its selection.

Thanks for making this important point. To address this concern:

(a) We will add more information about the model origin and features that make it ideal for our purposes in the section "2.1 GCM simulations and river routing model". The model's key attribute lies in its simplicity, enabling long-term global simulations with minimal computational resources, all while delivering commendable performance (as shown in item b).

(b) We will include an explicit validation of simulated river flows following the suggestion of reviewer #2. It includes the calculation of four different metrics: (relative bias, overlapping coefficient, correlation coefficient, and non-parametric Kling-Gupta efficiency) for the 18 hydrological simulations and the ensemble mean. The metrics are based on the comparison of our simulations against monthly observations of a selection of 346 monitored rivers, which cover about 42% of the global land and their flows contribute to about 45% of the global river discharge. The validation results clearly demonstrate the robust performance of the model in this study (please see them in our response to reviewer #2 main comment #1). Further details about the validation methodology will be included in the section "2.2 Assessment methodology", while the scores will be analyzed in the new section "3.1 Validation of the hydrological simulations".

(c) We will provide a link to the GitHub repository of the used model in the "Code and data availability" section. The repository includes access to the source code, documentation, and examples for running at the global and regional scales.

4. L110-114: The demonstration of the estimation of signal and noise ratio needs more details:

a) the mathematical details of the low-pass filter, justification for its choice, and an explanation of how the choice of low-pass filter might impact the noise term;

b) regarding the noise term, why this could be a representation of "natural variability"? Do the variation of the river flow in the PRESENT period include impact of human water activities? If so, how would this impact be separated from the "natural variability"?

c) why the signal of a local river flow anomaly can be linearly regressed on the global signal given large spatial heterogeneity in local river flow variability.

These terms need to be better demonstrated in the method section and the robustness of the estimation of signal to noise ratio need to be proved.

The estimation of the signal-to-noise ratio follows the method proposed by Hawkins and Sutton (2012), with a thorough parameter selection process that includes sensitivity tests. We appreciate the reviewer's feedback and would like to provide a clear response to each concern.

a) Choice of filter:

The choice of the low-pass filter is indeed a critical step in our analysis, as it directly impacts on the noise term N, and subsequently, the Time of Emergence. In our approach, the climate change signal S_G is estimated as $S_G = Q_G * w$, i.e., convolving the global river discharge anomaly time-series Q_G with a Hanning window w that has a length of 41 years. This operation performs a smoothing of the given time-series. The filter is chosen considering the following factors:

(i) it effectively attenuates high-frequency noise without introducing phase distortion, unlike Butterworth, Chebysev, or Elliptic (see their phase shifts around 2020 in Fig. R1.1a), and minimizes boundary effects, unlike the FFT low-pass filter (see the blending effect in the first and last decades in Fig. R1.1a);

(ii) the Hanning window provides similar smoothing results to other filters, except for the rectangular window, which introduces some interannual variability (see Fig. R1.1b); and

(iii) a window of 41 years highlights the long-term variations.

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. Similarly, the "blending" of the FFT smoothed signal on the edges of the time-seres, 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. Moreover, Fig. R1.2 demonstrates the insensitivity of local signal and noise to different filtering options tested for the global signal.

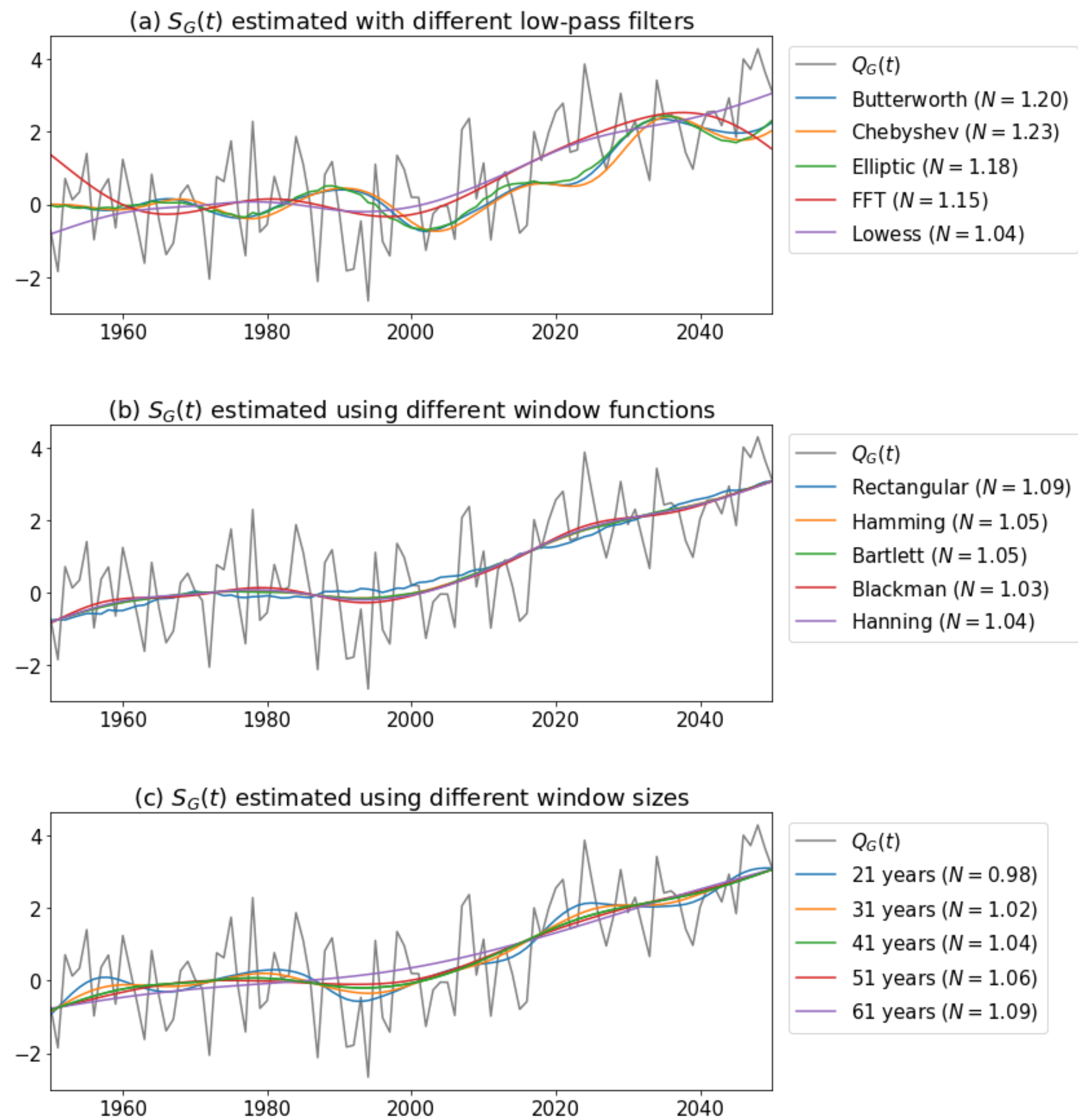

Figure R1.1. Low-pass filter selection test for a river discharge anomaly time-series (Q_G) using the CNRM-CM6-1-HR model as a case study. The signal S_G is estimated with various (a) filters, (b) window functions, and (c) window sizes. The test of parameters in (b) and (c) is applied for the Lowess filter. The resulting noise value of each filter is provided in the legend.

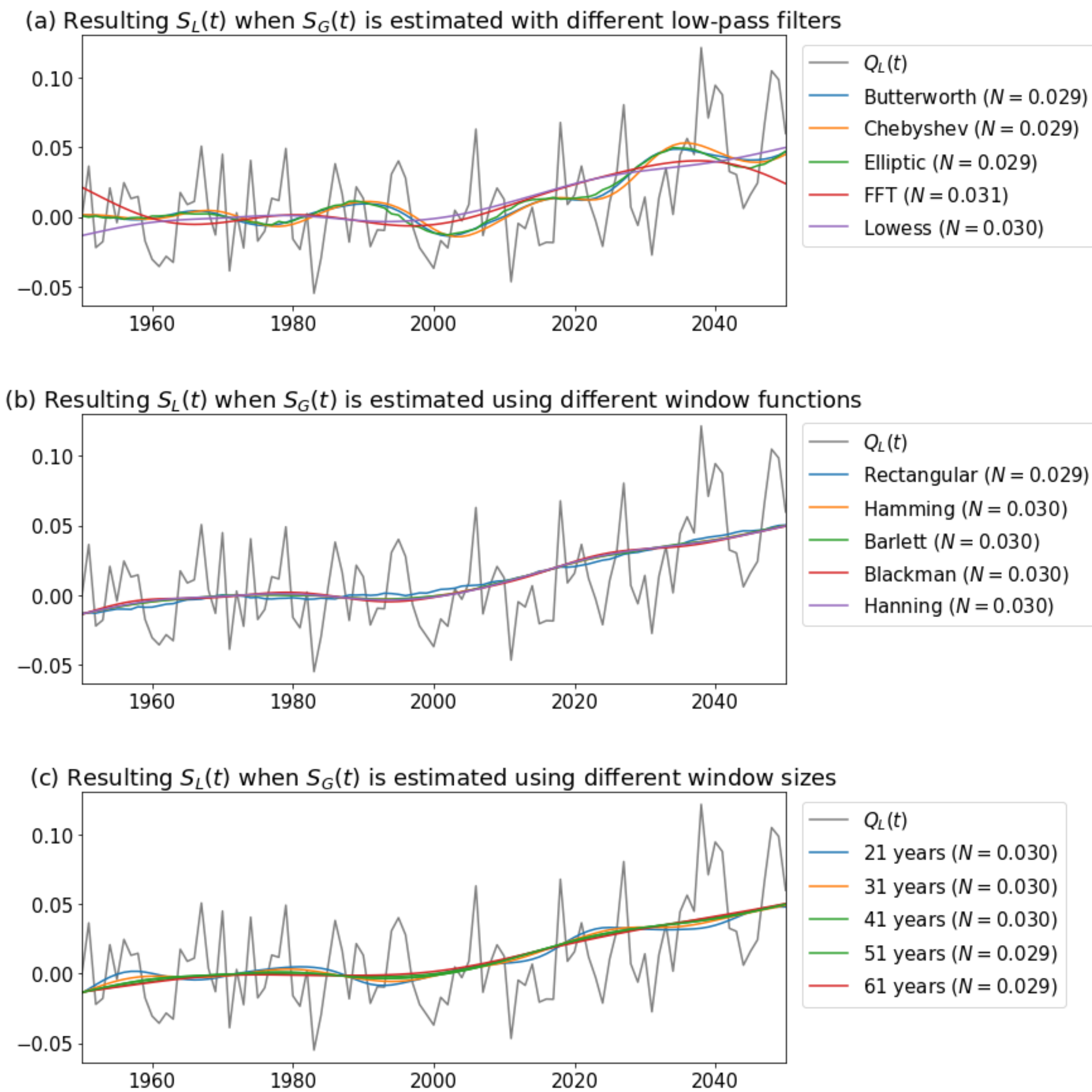

Figure R1.2. Illustration of the robustness of local signal (S_L) and noise (N_L) derived from various filter options applied to the global signal. The time-series represents the Yukon discharge at its river mouth.

b) Natural variability:

In our study, we adhere to the definition of "natural variability" as proposed by Hawkins and Sutton (2012), also applied in Hawkins et al. 2020, i.e. noise is the local component that is unexplained by long-term global changes. We acknowledge that in some studies, the definition of noise has been broadened to encompass the uncertainty in the climate response to anthropogenic forcing and the uncertainty in future anthropogenic emissions (Giorgi and Bi, 2009; Hawkins and Sutton, 2009, 2011). However, the GCM simulations used in our study do not include anthropogenic water activities. Thus, our primary focus remains on the natural internal variability of climate, as this serves as the key source of noise relevant for the analysis of the Time of Emergence. The base period for the anomalies calculation covers 6.5 decades (1950-2014) capturing different variability time-scales from interannual to multidecadal. In summary, while we recognize that river flow variations in the ‘real world’ include both natural and human-induced elements, our study primarily aims to assess when simulated changes in river discharge become distinguishable from the background of this natural climate variability, following the approach outlined by Hawkins et al.

c) Linear regression:

We agree with the reviewer that river flow anomalies are not homogeneous worldwide. However, we justify the estimation of local signals based on linear regression of local anomalies with respect to global anomalies considering that:

(i) It has been applied for precipitation which is strongly heterogeneous, even more than river flow. See Hawkins et al. (2020) and the IPCC AR6 Ch4 WGII (Caretta et al. 2022).

(ii) River flow anomalies are not largely heterogeneous, indeed they present consistent spatial responses at the catchment scale.

(iii) The alternative method to estimate the signal, i.e. estimating local signals by filtering local time-series produces misleading results. The local filtered signal is impacted by the filter's 'blending' effects on the extremes. This is clearly shown in Fig. R1.3 for Congo, Amazon, and Negro, three rivers with different regional trends. Figure R1.4 shows how the blending yields high SNR values in the first year of simulation, which is unrealistic. This issue is avoided when linear regressions are applied.

Following our above responses to each individual concern of comment 4, we will summarize the main points of this discussion and include it in the revised manuscript. This helps ensure that the methodological details are robustly presented and that the reader can follow the logic behind our decisions.

Caretta, M.A., A. Mukherji, M. Arfanuzzaman, R.A. Betts, A. Gelfan, Y. Hirabayashi, T.K. Lissner, J. Liu, E. Lopez Gunn, R. Morgan, S. Mwanga, and S. Supratid, 2022: Water. In: *Climate Change 2022: Impacts, Adaptation and Vulnerability.* Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 551–712, doi:10.1017/9781009325844.006.

Giorgi, F., and X. Bi (2009), Time of emergence (TOE) of GHG-forced precipitation change hot-spots, Geophys. Res. Lett., 36, L06709, doi:10.1029/2009GL037593.

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, Bull. Am. Meteorol. Soc., 90, 1095–1107, doi:10.1175/2009BAMS2607.1.

Hawkins, E., and R. Sutton (2011), The potential to narrow uncertainty in projections of regional precipitation change, Clim. Dyn., 37, 407–418, doi:10.1007/s00382-010-0810-6.

Hawkins, E., and Sutton, R. (2012). Time of emergence of climate signals. Geophysical Research Letters, 39(1).

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., and Sutton, R. (2020). Observed emergence of the climate change signal: from the familiar to the unknown. Geophysical Research Letters, 47(6), e2019GL086259.

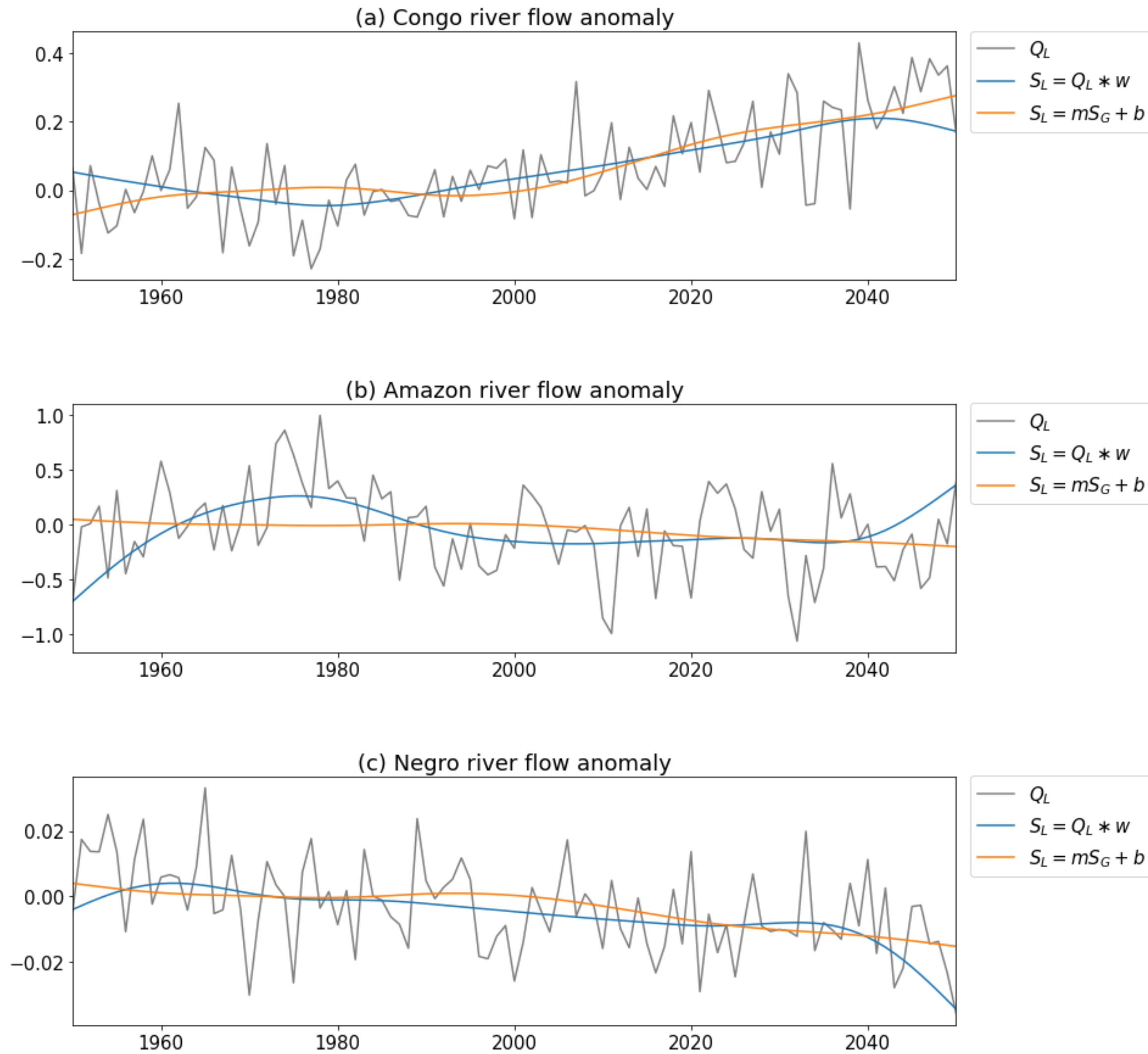

Figure R1.3. Local signal (S_L) estimated by filtering with the Lowess filter (blue) compared to the local signal based on the linear regression of local anomalies with respect to global anomalies (orange). The signals are estimated for time-series of rivers presenting different trends (positive, neutral, and negative): (a) Congo, (b) Amazon (), and (c) Negro river flow at their river mouth.

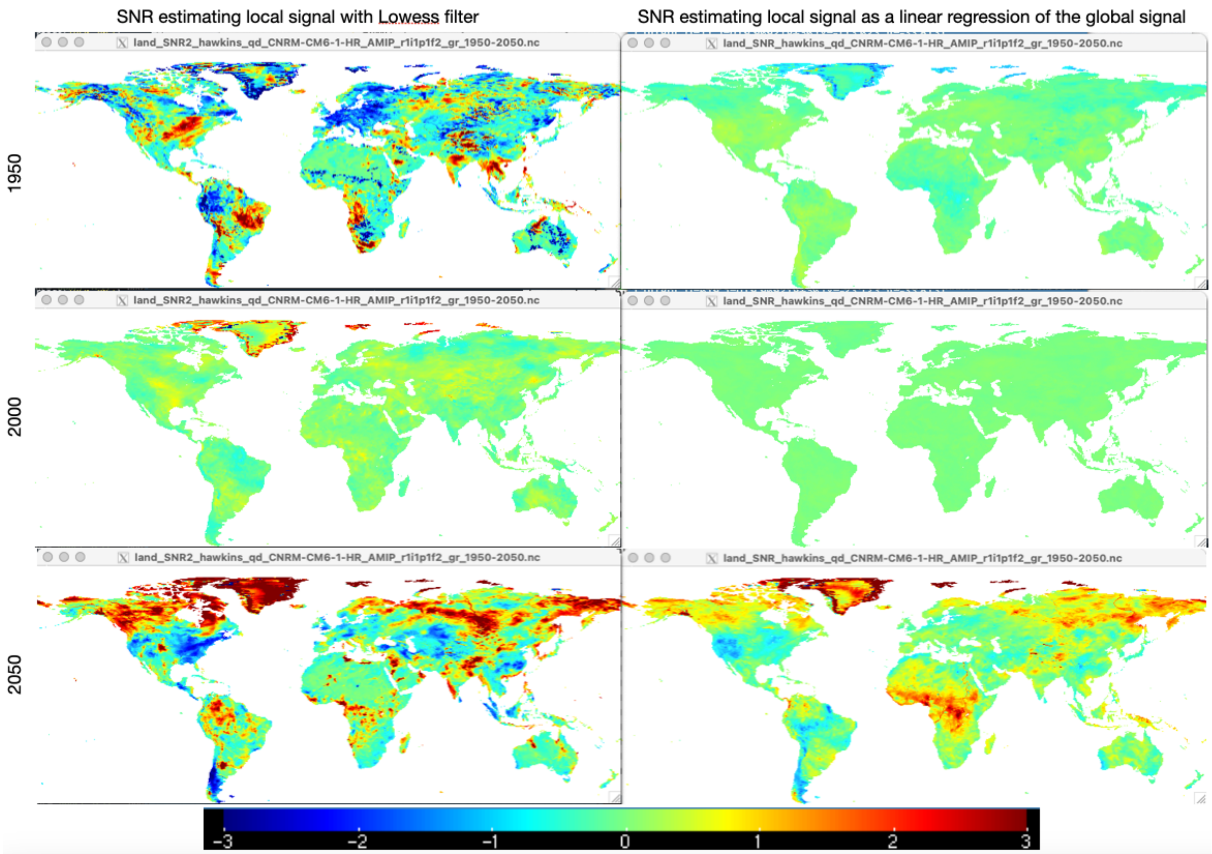

Figure R1.4. Resulting SNR for years 1950 (top), 2000 (middle), and 2050 (bottom) when local signals are estimated by filtering with the Lowess filter (left) compared to the SNR when local signal based on the linear regression of local anomalies with respect to global anomalies (right).

5. In Figure 2, the difference in runoff between the two periods in Greenland and central Australia is small while in Figure 3 (a) the percentage changes in river flow for these locations are outstanding. Is this discrepancy due to the use of the percent change as a measure, rather than absolute differences? I think it would be beneficial to explain the discrepancy.

The visual discrepancy between the small and large differences observed in Greenland and central Australia in our original Figures 2 and 3a has two reasons:

a) mathematical: differences expressed as percentage changes tend to be high on places where mean values are small, such as Greenland or Australia.

b) visual: Figure 3a shows uniform colors at the catchment scale, and such colors represent the percentage change at the river mouth, which integrates the runoff of the entire catchments. For instance, the strong values of Greenland catchments in Figure 3a are directly influenced by the high difference between FUTURE and PAST near the deltas shown in Figure 2, rather than the small differences observed upstream.

We clarify this in the revised manuscript.

6. About ToE: again, the method of signal and noise decomposition would largely determine the results of ToE. I wonder if the author could check the robustness of the ToE as well.

We agree with the reviewer, the ToE is sensitive to the signal and noise estimations. In our response to the major comment 4, we provide a comprehensive justification of the methodology we used to estimate the signal-to-noise ratio, demonstrating the robustness of our approach. We will include this clarification in the revised manuscript, addressing the sensitivity of ToE to signal-to-noise calculations.

7. Figure 7: the discussion on the results shown in Figure 7 is confusing. In my perspective, changes are described in the context of a simulation scenario so that they are not yet scalable to real world situation. However, the authors mentioned about the risk to metropolitan areas and the implications to infrastructure management. I think these discussions are relevant but somehow feel disconnected from the results of the simulation. To better connect, I think the authors should first demonstrate how the simulated river flow for PRESENT period compare to observations and form the discussion based on it. This might also help condense the content in the abstract.

We appreciate the reviewer's feedback on the discussion about simulated river flow projections and extreme events observed in the real world. Recognizing that this connection is too ambitious that connection, we opted to follow the suggestion of reviewer #2 and replace that analysis of specific cases with the assessment of groups of rivers in regions with similar trends. This modification streamlines our analysis by grouping rivers into regions with notable trends and clear consensus among model simulations—Central Africa, Arctic, South Asia, and Patagonia. This is a more concise way to better connect with the simulation results and convey the broader implications of this study. Please see our response to reviewer #2 main comment #2.

Minor comments:

1. L68: Clarification is needed regarding what “availability”

Surface and subsurface runoff is simulated by all GCMs, but these variables are not always stored. The text will be rephrased to “These HighResMIP GCMs were selected based on the availability of surface and subsurface runoff data.”

2. L86-87: Please specify the resolution of the target grid.

Thanks for noting it was unclear. We will rephrase it: “The simulations are run globally (excluding Antarctica) using the nearest-neighbour option to regrid the runoff from the original GCM resolutions to the target grid at a common resolution of 0.25°.”

L97-98: the term “expected changes in rivers” requires clarification. Rationale is needed to connect the three steps.

We will rephrase the term as “projected changes in rivers”. As indicated in our response to major comment 7 we will provide a regional analysis of the projections that better connects with the simulated projections and the evaluation of signal-to-noise ratio and time-of-emergence.