Title: Globally observed trends in mean and extreme river flow attributed to climate change

Authors: Lukas Gudmundsson[1,*], Julien Boulange[2], Hong X. Do[3,4,5], Simon N. Gosling[6], Manolis G. Grillakis[7], Aristeidis G. Koutroulis[8], Michael Leonard[3], Junguo Liu[9], Hannes Müller Schmied[10,11], Lamprini Papadimitriou[12,13], Yadu Pokhrel[14], Sonia I. Seneviratne[1], Yusuke Satoh[2,15], Wim Thiery[1,16], Seth Westra[3], Xuebin Zhang[17], Fang Zhao[18,19].

Affiliations:

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Zürich, Switzerland.

[2]National Institute for Environmental Studies (NIES), Tsukuba, Japan.

[3]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia.

[4]Faculty of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City, Vietnam.

[5]School for Environment and Sustainability, University of Michigan, Ann Arbor, Michigan, USA.

[6]School of Geography, University of Nottingham, Nottingham NG7 2RD, United Kingdom.

[7]Institute for Mediterranean Studies, Foundation for Research and Technology Hellas, Rethymno 74100, Greece.

[8]School of Environmental Engineering, Technical University of Crete, Chania, Greece.

[9]School of Environmental Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, 518055, China.

[10]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany.

[11]Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany.

[12]Cranfield Water Science Institute, Cranfield University, Cranfield, United Kingdom.

[13]Mott MacDonald Ltd, Cambridge, United Kingdom.

[14]Department of Civil and Environmental Engineering, Michigan State University, East Lansing, Michigan, USA.

[15]International Institute for Applied Systems Analysis, Laxenburg, Austria.

[16]Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium.

[17]Environment Canada, Toronto, Canada.

[18]Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China.

[19]Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany.

Science
AAAS

*Correspondence to: lukas.gudmundsson@env.ethz.ch.

Abstract: Anthropogenic climate change is expected to affect global river flow. Here we analyze time series of low, mean and high river flows from 7,250 observatories around the world covering 1971 - 2010. We identify spatially complex trend patterns, where some regions are drying and others are wetting consistently across low, mean and high flows. Trends computed from state-of-the-art model simulations are consistent with the observations only if radiative forcing that accounts for anthropogenic climate change is considered. Simulated effects of water and land management do not suffice to reproduce the observed trend pattern. Thus the analysis provides clear evidence for the role of externally forced climate change as a causal driver of recent trends in mean and extreme river flow at the global scale.

One Sentence Summary: Climate change impacts detected in mean and extreme river flow at the global scale.

Main Text: Among key concerns with respect to anthropogenic climate change (ACC) are impacts on the terrestrial water cycle. Earth System Models (ESM) indicate that projected ACC can influence water availability on land (*1*) and may trigger more floods (*2*) and droughts (*3*). While detection and attribution studies have shown that observed changes in atmospheric variables such as precipitation (*4, 5*) and water vapor (*6*) are consistent with model simulations that account for historical ACC, evidence for a human fingerprint on past changes in river flow and hydrological extremes is still lacking at the global scale.

Two factors have complicated the detection and attribution of changes in terrestrial water systems at the global scale. Firstly, although river flow time series are the most abundant observations of water resources and hydrological extremes, the slow mobilization of in situ observations has confined past assessments to regional and continental case studies (*7-12*) or to small collections of large river basins, with most of the records ending in the 20th century (*13-15*). To circumvent this lack of global in situ observations, researchers have used reconstructions of essential hydrological variables such as soil moisture (*16-18*) or evapotranspiration (*19*) and indicators of water availability (*20*) as the basis for climate change detection and attribution studies. Although these efforts have revealed that ACC is detectable in terrestrial water systems, they lack a direct connection to in situ observations of quantities relevant for water management.

Secondly, besides ACC, on-ground human activities such as historical water and land management (HWLM) are also altering water resources and hydrological extremes, e.g. directly through flow regulation and water abstractions, or indirectly through effects of land-cover change (*13*). For example, large-scale water withdrawal for irrigation, might induce declining trends in river flow. Likewise, reservoir expansion may lead to changes in stream flow. Several studies have shown that effects of HWLM on water resources could be equally large or might even exceed climate change impacts in some regions (*13, 21, 22*). However, ESMs, which are an important tool for attributing observed changes to human influences on the climate system, typically do not account for HWLM as a possible confounding factor (*7*).

Recent advances in mobilizing in situ river flow observations (*23, 24*) and an unprecedented multi-model ensemble that combines the ESMs' ability to account for ACC with the capacity of

high resolution Global Hydrology Models (GHM) to incorporate HWLM (*25*), allow us to tackle the challenge of attributing observed changes in river flow at the global scale. Here we consider in situ observations of daily average river flow from 7,250 gauging stations (Figure S1) that have at least 28 years with almost complete daily data in 1971 – 2010 (*26*). To balance the uneven spatial distribution of in situ observations, the analysis is constrained to predefined sub-continental regions (*27*) with at least 80 stations. Yearly time series of low (annual 10th percentile), mean (annual mean) and high (annual 90th percentile) river flows are considered, respectively representing very dry, average and very wet conditions. For each time series, trends are estimated and expressed in terms of percentage change per decade. Subsequently the median trend is computed for each region, to reduce the effects of local-scale natural variability, observational errors and spatial autocorrelation for further statistical analyses (*28*).
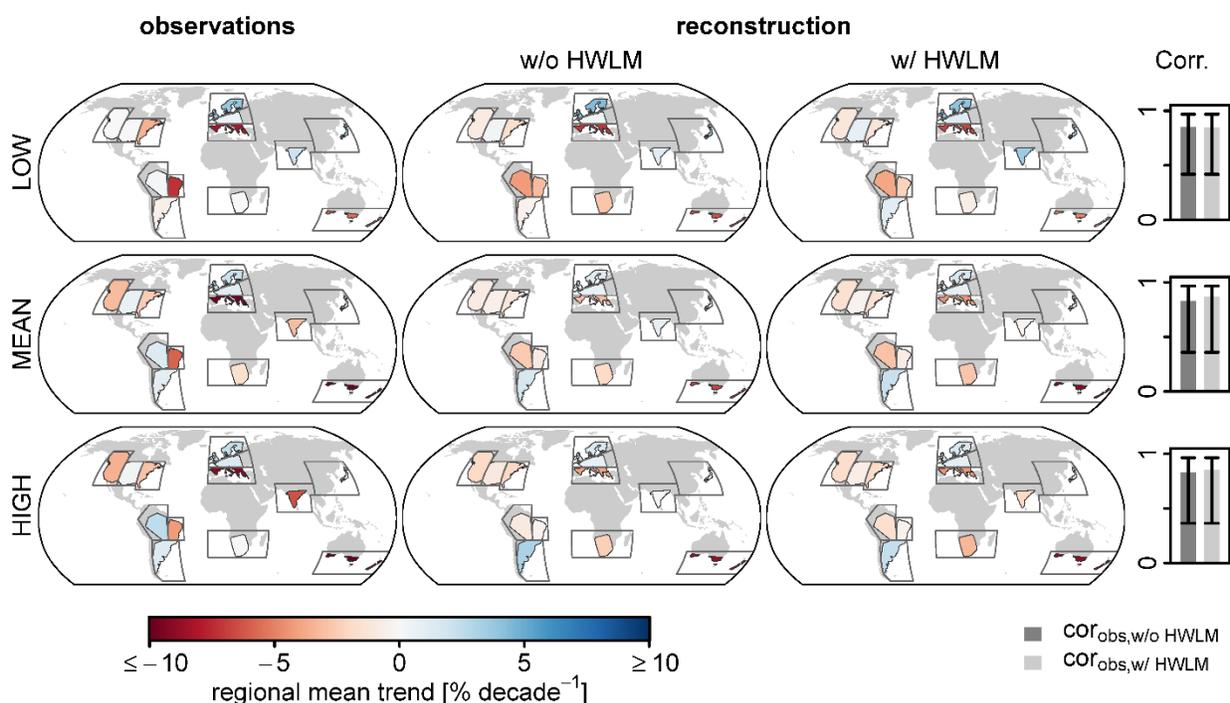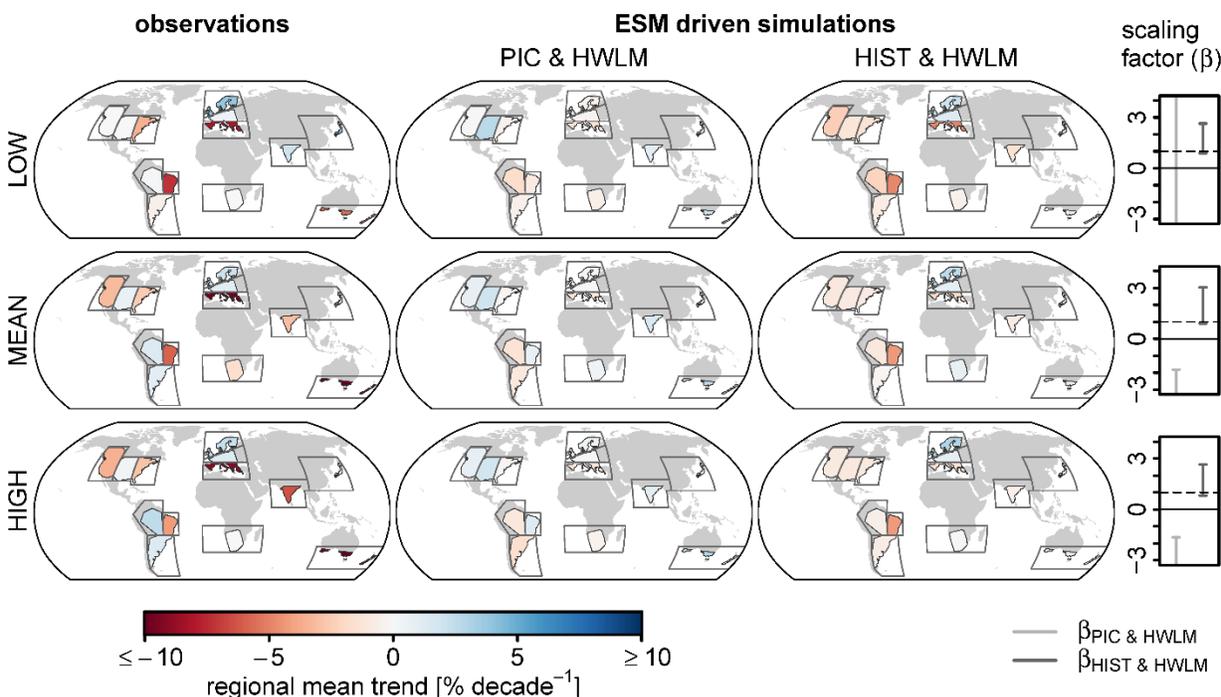

Figure 1: Comparison of observed and reconstructed regional median river flow trends (1971 – 2010). Observations: trends computed from in situ observations. Reconstruction: multi-model mean trend of global hydrology model simulations driven by observational atmospheric forcing, with (w/) and without (w/o) the effects of historical water and land management (HWLM). Hollow polygons indicate pre-defined regions used for grouping stations. Coloured polygons are defined by the convex hull around the station coordinates in the respective regions. Colours indicate the regional median trend. Note that the colour scale has been truncated to enhance the readability of the maps. Supplementary Figure S9 displays the full range of all data presented here. Corr.: Pearson correlation coefficients between the spatial patterns of observed and reconstructed regional median trends, alongside 99% confidence intervals that are based on Fisher's z-transform.

Figure 1 reveals spatially complex trend patterns in observed low, mean and high river flow. Some regions such as northeast Brazil, southern Australia, southern Africa and the Mediterranean show a drying tendency. Other regions such as north Europe tend towards wetter conditions. We also note that the level of spatial aggregation may mask sub-regional differences. The analysis also confirms previous results (*29*) indicating that the direction of change is often consistent throughout the entire flow distribution, i.e. trends in low, mean and high flows share the same sign. Reconstructions of global river flow from the Inter-Sectoral Impact Model

Intercomparison Project phase 2a (ISIMIP2a) that are based on GHMs driven with observational atmospheric data (*30*) are highly correlated with the observed trend pattern (Figure 1), although the reconstructions have a tendency to underestimate the magnitude of the observed trends. Accounting for HWLM does not improve the skill of the reconstruction in reproducing the spatial patterns of observed regional median trends, despite the fact that it can improve overall model performance (*31*).

Figure 2: Comparison of observed and simulated river flow trends (1971 – 2010). Observations: identical to Figure 1 and included here to facilitate comparison. Simulations: trends computed from model experiments that account for HWLM but consider simulated atmospheric data with either pre-industrial (PIC) or historical (HIST) radiative forcing. Note that colour scale for the trends has been truncated to enhance the readability of the maps. Supplementary Figures S9 and S10 displays the full range of all data presented here. Scaling factor: 10% to 90% confidence intervals of the scaling factors from the detection analysis. The simulated patterns are consistent with the observations if the lower bound of the confidence interval is larger than zero (solid horizontal line). The magnitude of observed and simulated change is consistent if the confidence intervals include one (dashed horizontal line). Confidence intervals exceeding the range of the ordinate are truncated to enhance readability of the plot.

While the above assessment shows that changes in the atmospheric conditions are driving observed trends in low, mean and high river flow, it remains unclear if these changes can be attributed to ACC. To tackle this question, we employ the climate change detection and attribution approach (*32*), which ingests both observations and simulation experiments that include or exclude the drivers of ACC. If (i) simulations that include the drivers of ACC are consistent with the observations and (ii) if simulations that do not include them fail to be consistent with the observations it is possible to claim attribution. Here we consider two simulation experiments from the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b) (*25*), where GHMs that account for HWLM are driven with output from ESMs which ingest different radiative forcing. The first experiment considers pre-industrial radiative forcing and is referred to as PIC&HWLM from here onwards. Since pre-industrial radiative forcing does not impose systematic trends in the simulations, the PIC&HWLM experiment allows to test the hypothesis whether the observed trend patterns can be explained by the

simulated effects of HWLM alone. The second experiment considers historical radiative forcing, which includes both anthropogenic (e.g. human greenhouse gas or aerosol emissions) as well as natural (e.g. influence of large volcanic eruptions) factors throughout the past century. This experiment is referred to as HIST&HWLM and allows to test the hypothesis whether the
5   addition of historical radiative forcing contributes to explaining the observed trend pattern.

Figure 2 compares observed river flow trends to trends from of both the PIC&HWLM and the HIST&HWLM simulations. The consistency of the observed trend pattern with either of the simulation results is tested using optimal finger printing (28, 33). To this end, the multi-model mean of the simulations (x) is regressed on the observations (y), while accounting for natural
10   variability (ϵ) as well as model and sampling errors (ν), such that $y = \beta \times (x - \nu) + \epsilon$ (33). Testing the significance of the association between the observed and simulated patterns is based on the magnitude of the scaling factor β. The simulated pattern is said to be detected in the observations if β is significantly larger than zero, i.e. if the lower end of the associated confidence intervals are above zero. Scaling factors derived from simulations with PIC&HWLM
15   are never significantly larger than zero (Figure 2). This shows that simulated effects of HWLM without accounting for historical radiative forcing cannot explain the observed trend pattern. This is in agreement with the above assessment, where accounting for HWLM did not improve the consistency of observed and reconstructed trends (Figure 1). The situation differs for the analysis of HIST&HWLM, i.e. the simulations that account for historical radiative forcing. Here
20   the scaling factors of all indices are significantly larger than zero ($p < 0.1$).

The results show that the combined effect of historical radiative forcing and HWLM is detected in observed trend patterns of low, mean and high river flow. The analysis also suggests that the magnitude of the simulated trend patterns under historical radiative forcing is consistent with the observations ($\beta \approx 1$). This in combination with the finding that accounting for HWLM does not
25   improve reconstructions (Figure 1), implies that simulated impacts of HWLM only have a minor effect on regional median river flow trends. Consequently, the results suggest that the simulated effects of historical radiative forcing on the climate system are essential for explaining the observed patterns of regional median low, mean and high river flows.

To investigate effects of the mismatch between point-scale observatories and model grid cells
30   and to assess impacts of regional sampling biases of the observations, the analysis was repeated using GHM based river flow reconstructions that allow for full spatial coverage (Figures S2-S7). Despite additional uncertainties induced model based reconstructions, the results are widely consistent with the observational assessment (Figure S8). Furthermore, a detailed inspection of observed, reconstructed and simulated trends shows that the internal variability implied by
35   HIST&HWLM simulations is comparable to observed variability (Figures S9-S12), indicating the validity of the assumption that natural climate variability can be approximated through chaotic model trajectories.

We note that as in any climate change detection and attribution exercise we cannot fully rule out that processes not captured by the models might contribute to the observed trend pattern (32).
40   For example, there remain uncertainties regarding the response of transpiration to dryness stress or in the representation of HWLM. Furthermore, while the ISIMIP2b ensemble allows for the first time to account for HWLM in a climate change detection and attribution setup, the fact that no separate simulations with either anthropogenic or historical natural forcing are available

hinders an unambiguous attribution of the observed trend pattern to anthropogenic climate change. In particular, natural changes in the radiative forcing triggered by large volcanic eruptions have been shown to affect the global hydrological cycle. However, recent research has demonstrated that the effects of such eruptions on river flow are typically confined to a few years after the eruption (*34*) and are therefore expected to only have a small influence on long-term trends.

Overall, the presented analysis lines up with the existing body of literature documenting that anthropogenic climate change is influencing the world's water cycle (*4-6, 16-20*). Possible mechanisms that drive trends in low, mean and high streamflow include large-scale shifts in precipitation (*4, 5, 17*), changes in factors that influence evapotranspiration (*6, 19, 20*) or alterations of the timing of snow accumulation and melt driven by rising temperatures (*8, 12*). Combining the evidence of these findings with the results of the presented analysis (Figure 1, Figure 2), supports the conclusion that it is likely that anthropogenic climate change is contributing to the global pattern of trends in low, mean and high river flow.

In summary, we demonstrate for the first time that the global pattern of observed changes in river flow are only captured by model simulations that account for historical radiative forcing and that simulated effects of HWLM do not significantly contribute to explaining global trend patterns of low, mean and high flows. Thus we have provided clear evidence for the role of historical radiative forcing as a causal driver of trends in mean and extreme river flow at the global scale.

Supplementary Materials:

Materials and Methods

Figures S1-S12

Tables S1-S7

References (*35-43*)

References and Notes:
1. P. Greve, L. Gudmundsson, S. I. Seneviratne, Regional scaling of annual mean precipitation and water availability with global temperature change. *Earth System Dynamics* **9**, 227-240 (2018).
2. Y. Hirabayashi *et al.*, Global flood risk under climate change. *Nature Climate Change* **3**, 816 (2013).
3. C. Prudhomme *et al.*, Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences* **111**, 3262-3267 (2014).
4. X. Zhang *et al.*, Detection of human influence on twentieth-century precipitation trends. *Nature* **448**, 461-465 (2007).
5. K. Marvel, C. Bonfils, Identifying external influences on global precipitation. *Proceedings of the National Academy of Sciences* **110**, 19301-19306 (2013).
6. K. M. Willett, N. P. Gillett, P. D. Jones, P. W. Thorne, Attribution of observed surface humidity changes to human influence. *Nature* **449**, 710-712 (2007).
7. L. Gudmundsson, S. I. Seneviratne, X. Zhang, Anthropogenic climate change detected in European renewable freshwater resources. *Nature Climate Change* **7**, 813 (2017).

8.   T. P. Barnett *et al.*, Human-Induced Changes in the Hydrology of the Western United States. *Science* **319**, 1080-1083 (2008).

9.   G. Blöschl *et al.*, Changing climate both increases and decreases European river floods. *Nature*, (2019).

10.   G. Blöschl *et al.*, Changing climate shifts timing of European floods. *Science* **357**, 588-590 (2017).

11.   C. Rosenzweig *et al.*, Attributing physical and biological impacts to anthropogenic climate change. *Nature* **453**, 353 (2008).

12.   H. G. Hidalgo *et al.*, Detection and Attribution of Streamflow Timing Changes to Climate Change in the Western United States. *Journal of Climate* **22**, 3838-3855 (2009).

13.   F. Jaramillo, G. Destouni, Local flow regulation and irrigation raise global human water consumption and footprint. *Science* **350**, 1248-1251 (2015).

14.   P. C. D. Milly, K. A. Dunne, A. V. Vecchia, Global pattern of trends in streamflow and water availability in a changing climate. *Nature* **438**, 347-350 (2005).

15.   P. Wu, R. Wood, P. Stott, Human influence on increasing Arctic river discharges. *Geophysical Research Letters* **32**, L02703 (2005).

16.   X. Gu *et al.*, Attribution of Global Soil Moisture Drying to Human Activities: A Quantitative Viewpoint. *Geophysical Research Letters* **46**, 2573-2582 (2019).

17.   K. Marvel *et al.*, Twentieth-century hydroclimate changes consistent with human influence. *Nature* **569**, 59-65 (2019).

18.   B. Mueller, X. Zhang, Causes of drying trends in northern hemispheric land areas in reconstructed soil moisture data. *Climatic Change* **134**, 255-267 (2016).

19.   H. Douville, A. Ribes, B. Decharme, R. Alkama, J. Sheffield, Anthropogenic influence on multidecadal changes in reconstructed global evapotranspiration. *Nature Climate Change* **3**, 59 (2012).

20.   R. S. Padrón *et al.*, Observed changes in dry-season water availability attributed to human-induced climate change. *Nature Geoscience* **13**, 477-481 (2020).

21.   C. J. Vörösmarty, P. Green, J. Salisbury, R. B. Lammers, Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science* **289**, 284-288 (2000).

22.   I. Haddeland *et al.*, Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences* **111**, 3251-3256 (2014).

23.   H. X. Do, L. Gudmundsson, M. Leonard, S. Westra, The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data* **10**, 765-785 (2018).

24.   L. Gudmundsson, H. X. Do, M. Leonard, S. Westra, The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. *Earth System Science Data* **10**, 787-804 (2018).

25.   K. Frieler *et al.*, Assessing the impacts of 1.5° C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development* **10**, 4321-4345 (2017).

26.   Materials and methods are available as supplementary materials at the Science website

27.   S. I. Seneviratne *et al.,* C. B. Field *et al.*, Eds. (Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2012), pp. 109-230.

28.   M. R. Allen, P. A. Stott, Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Climate Dynamics* **21**, 477-491 (2003).

29. L. Gudmundsson, M. Leonard, H. X. Do, S. Westra, S. I. Seneviratne, Observed Trends in Global Indicators of Mean and Extreme Streamflow. *Geophys. Res. Lett.* **46**, 756-766 (2019).

30. S. Gosling *et al.*, ISIMIP2a Simulation Data from Water (global) Sector (V. 1.1). *GFZ Data Services*, (2019).

31. T. I. E. Veldkamp *et al.*, Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. *Environmental Research Letters* **13**, 055008 (2018).

32. N. L. Bindoff *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* T. F. Stocker *et al.*, Eds. (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013), pp. 867-952.

33. A. Hannart, A. Ribes, P. Naveau, Optimal fingerprinting under multiple sources of uncertainty. *Geophysical Research Letters* **41**, 1261-1268 (2014).

34. C. E. Iles, G. C. Hegerl, Systematic change in global patterns of streamflow following volcanic eruptions. *Nature Geosci* **8**, 838-842 (2015).

35. H. Müller Schmied *et al.*, Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrol. Earth Syst. Sci.* **20**, 2877-2898 (2016).

36. K. E. Taylor, R. J. Stouffer, G. A. Meehl, An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society* **93**, 485-498 (2011).

37. B. Kirtman *et al.*, in *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* C. Intergovernmental Panel on Climate, Ed. (Cambridge University Press, Cambridge, 2014), pp. 953-1028.

38. M. Hauser *et al.*, Methods and model dependency of extreme event attribution: The 2015 European drought. *Earth's Future*, (2017).

39. P. K. Sen, Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association* **63**, 1379-1389 (1968).

40. M. R. Allen, S. F. B. Tett, Checking for model consistency in optimal fingerprinting. *Climate Dynamics* **15**, 419-434 (1999).

41. B. Efron, R. J. Tibshirani, *An introduction to the bootstrap*. Monographs on statistics and applied probability (Chapman and Hall/CRC, New York, 1993), vol. 57, pp. 436.

42. G. P. Weedon *et al.*, Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. *Journal of Hydrometeorology* **12**, 823-848 (2011).

43. G. P. Weedon *et al.*, The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research* **50**, 7505-7514 (2014).

Science

AAAS

Supplementary Materials for

Globally observed trends in mean and extreme river flow attributed to climate change

Authors: Lukas Gudmundsson, Julien Boulange, Hong X. Do, Simon N. Gosling, Manolis G. Grillakis, Aristeidis G. Koutroulis, Michael Leonard, Junguo Liu, Hannes Müller Schmied, Lamprini Papadimitriou, Yadu Pokhrel, Sonia I. Seneviratne, Yusuke Satoh, Wim Thiery, Seth Westra, Xuebin Zhang, Fang Zhao

Correspondence to: lukas.gudmundsson@env.ethz.ch

This PDF file includes:

Materials and Methods
Figs. S1 to S12
Tables S1 to S7

Materials and Methods

River Flow Observations

For this study, we consider river flow observations with global coverage that are held in the Global Streamflow Indices and Metadata (GSIM) archive (23,24). The GSIM archive consists of a wide range of time series indices at monthly, seasonal and yearly resolution, which were computed from daily average values after automated quality control (e.g. outlier detection). Values flagged by the quality control procedure were set to missing prior to computation. See ref. (24) for details.

Here we focus on the following indices, with values that were calculated for each calendar year separately:
- Low flows: Annual 10th percentile time series (LOW)
- Mean flow: Annual mean time series (MEAN)
- High flow: Annual 90th percentile time series (HIGH).

For the presented analysis, annual index values that are computed using less than 350 daily values were discarded, as the low- and high flow indices are very sensitive to missing data. Time series with at least 70% (i.e. 28) non-missing years in 1971-2010 were selected.

Although the GSIM archive constitutes the most comprehensive publicly available global collection of in situ river flow observations, there remain large regional differences in station density. To reduce the effects of regional sampling biases we follow a previous approach (29) and constrain the assessment to sub-continental regions defined in the Special Report on Extremes of the Intergovernmental Panel on Climate Change (IPCC) (27), later referred to as SREX-regions. Only regions with at least 80 stations were considered to ensure a robust estimation of regional scale statistics (see section "Trend analysis" for details). Thus, stations in regions with less than 80 river flow observatories were discarded. This procedure resulted in a final selection of 13 SREX regions with a total of 7250 river flow observatories. Figure S1 shows the spatial distribution of the selected stations alongside the definition of the selected SREX regions.

Note that some of the catchments may extend beyond the pre-defined region boundaries and that some of the gauging stations will be connected through river networks, which implies statistical dependence among the observations. However, since the analysis compares observed trends to simulated river discharge masked to match observational coverage (see section "Pre-processing of Modelled Data"), possible dependencies are not biasing the analysis. To further investigate this issue a supplementary detection and attribution analysis is conducted. For this observed river flow is replaced by model based reconstructions extracted at those grid-cells that contain gauging stations and is masked to match observed patterns of messiness. The results are generally consistent with the analysis using observations (Figures S2, S3).

In addition we note that even within SREX regions there is significant heterogeneity in terms of spatial sampling. For example, several SREX regions (e.g. EAS, MED, SSA, EAS) have uneven spatial sampling which is likely to bias regional average statistics; this should be considered when interpreting region-averaged trend results. Note also that supplementary assessments (Figure S4-S7) that overcome regional sampling biases by substituting observations

by reconstructed river flow confirm the findings presented in the main body of the article despite the additional uncertainty implied by model based river flow reconstructions.

Model based reconstructions of river flow (ISIMIP2a)

Model-based reconstructions of river flow from the ISIMIP2a project (30) are considered to investigate if Global Hydrology Models (GHMs) can capture observed trends if driven with observational atmospheric data, and to check the stability of the analysis in light of regional sampling biases (Figures S2-S8). These consist of GHMs that are driven with up to three observation-based reconstructions of atmospheric variables that are available for the 1971-2010 time window. Thus, the temporal variability of these reconstructions is expected to closely resemble observed river flow. For a general description of the GHM-based river flow reconstructions, including the different atmospheric data, please refer to ref. (35) and Table S1. Two configurations for GHM based river flow reconstructions are available (the terms "nosoc" and "varsoc" are according to the ISIMIP2a simulation protocol):

- "nosoc" simulations (see supplementary Table S1), which do not include effects of water and land management. These are referred to as **reconstructions without (w/o) HWLM** (Historical Water and Land Management) in the main text.
- "varsoc" simulations (see supplementary Table S2), which account for time-varying water and land management. These are referred to as **reconstructions with (w/) HWLM** in the main text.

Water and land management factors considered by the GHMs used for river discharge reconstructions are summarized in Table S3. Further details on the individual GHMs can be found on https://www.isimip.org/impactmodels/.

River flow simulations forced with climate model output (ISIMIP2b)

To investigate if effects of anthropogenic climate change constitute a causal driver of observed trends in river flow, GHM simulations from ISIMIP2b are considered (25). In ISIMIP2b, GHMs are driven with atmospheric variables that were simulated by Earth System Models (ESMs) contributing to the CMIP5 archive (36). ESMs are used to simulate the response of the atmosphere to different radiative forcing (i.e. the amount of sunlight absorbed by the Earth minus the energy radiated back to space) that can be altered by human greenhouse gas and aerosol emissions. For a detailed description of the experimental setup please refer to ref. (25). Within ISIMIP2b a range of model experiments were conducted, each having different configurations of radiative forcing (driving the ESMs) or water and land management (implemented in the GHMs). Here the following experiments are considered:

PIC&1860soc: Note that these simulations are not mentioned in the main text but used for statistical inference as described in the Materials and Methods section of this supporting information document. These are also referred to as pre-industrial control simulations since the radiative forcing for the ESMs is set to pre-industrial conditions (PIC) which yield simulations with small human impacts on the Earth system. Water and land management in the GHMs is set to levels corresponding to 1860 (1860soc, corresponding to the "1860soc" setup from ISIMIP2b). These simulations are run over 200 year long time windows. All available PIC&1860soc simulations were split into non-overlapping 40 year segments, resulting in a total

of 154 segments. Supplementary Table S4 lists the number of segments for each available ESM-GHM combination for the PIC&1860soc experiment.

PIC&HWLM: ESM simulations that account for pre-industrial radiative forcing (PIC) are used to drive GHMs that account for historical water and land management (HWLM). These simulations hence do not account for anthropogenic climate change. For 1971-2005, the simulations are forced with time-varying historical estimates of HWLM, corresponding to the "histsoc" setup from the ISIMIP2b simulation protocol. For 2006-2010, HWLM is kept at the constant level of 2005 (i.e. the "2005soc" setup from the ISIMIP simulation protocol). See supplementary Table S5 for an overview on the ESMs and GHMs contributing to the PIC&HWLM simulations.

HIST&HWLM: ESM simulations that account for historical radiative forcing (HIST) are used to drive GHMs that account for historical water and land management (HWLM). Here HWLM is identical to HWLM in the PIC&HWLM experiments. For 1971-2005, historical radiative forcing is used as input for the climate models that are used to drive the GHMs. For 2006-2010 radiative forcing from the RCP6.0 scenario is used, as this and other emission scenarios closely resemble the observed radiative forcing for this time period (37,38). See supplementary Table S6 for an overview on the ESMs and GHMs contributing to the HIST&HWLM simulations.

Water and land management factors considered by the GHMs used for river discharge reconstructions are summarized in Table S7. Further details on the individual GHMs can be found on https://www.isimip.org/impactmodels/.

Pre-processing of Modelled Data

For the analysis the river discharge field (dis) that provides routed river discharge [m^3 s^{-1}] is considered, since this simulated variable is structurally closest to observed river flow [m^3 s^{-1}]. In the following time series extracted from the river discharge fields are interchangeably used with the term daily river flow data to highlight the structural similarity with observations.

First yearly resolution time series indices representing low, mean and high flows were computed from daily river flow data from each simulation. Subsequently, the multi-model mean of each index was computed for all experiments of the ISIMIP2a and the ISIMIP2b ensemble, except for the PIC&1860soc experiments. For all analysis involving both observations and GHM results, the model output was pre-processed as follows: (1) only grid-cells that contain observations were selected; and (2) for each selected grid-cell, years with no observations were set to missing.

Trend analysis

Trends are estimated using the Sen-slope estimator (39), which is robust against outliers, and expressed here in units of percentage change per decade following a previous assessment (29). Expressing the trends in percentage change per decade, reduces effects of different catchment areas and alleviates possible model biases. For the observations, the trend was computed for each station. For the simulations, trends were computed for each grid-cell. Subsequently, the regional median trend was computed for each considered SREX region.

Finally for the model based reconstructions as well as the PIC&HWLM and HIST&HWLM simulations the ensemble mean regional median trend was calculated.

Detection and attribution methodology

   Here we employ a variation of the optimal fingerprinting method that explicitly accounts for natural variability and uncertainty of multi-model means that was introduced by (33). Note also that compared to ref. (33), a simplified notation is used here, as the analysis is conducted in a setup that only accounts for one model experiment at a time. The method is based on an error in variable regression model, $y = x^*\beta + \epsilon$. Here y is a vector of n observations, which corresponds to the observed regional median trends for each SREX region. For the remainder of this paper n represents the number of SREX regions included in the analysis (13 for the majority of the results with the exception of Figures S6 and S7 for which $n = 26$). The regression coefficient β is referred to as the scaling factor. The noise term, $\epsilon \sim N(0, \Sigma)$, is a vector of length n used to represent natural climate variability, with Σ being a covariance matrix with dimension $n \times n$ that accounts for spatial dependence between observations. In the context of this assessment ϵ represents regional median trends that could occur by chance because of natural variability. The vector x^* of length n represents the expected response of the Earth system to an external forcing and is considered to be unknown. Instead, x^* is estimated using the ensemble mean of either the PIC&HWLM or the HIST&HWLM simulations. This estimate is written as $x = x^* + v$, where $v \sim N(0, \Omega)$ is a vector of length n and denotes the sampling error with covariance matrix Ω (dimension $n \times n$). Both y and x are assumed to be centered, which is achieved here by removing their arithmetic means prior to the analysis.

   Inference is conducted using a maximum likelihood approach with the log-likelihood $l(\beta, x^*|y, x) = -0.5(y - x^*\beta)'\Sigma^{-1}(y - x^*\beta) - 0.5(x - x^*)\Omega^{-1}(x - x^*)$. For algorithmic details on deriving the maximum likelihood estimates of β and a profile-likelihood method for estimating the corresponding confidence intervals the reader is referred to ref. (33). The covariance matrices (Σ and Ω) are estimated prior to likelihood maximisation (see below), rather than jointly, which is a common approach in detection and attribution research. The matrix Σ is estimated using pre-industrial control simulations, i.e. segments of the PIC&1860soc simulations that provide estimates of the natural variability of the Earth system without strong human influence. For this, the trend-analysis is first repeated for each of the 40-year segments. This results in a sample of vectors describing the mean trends in each SREX region as they could occur due to natural variability. This sample of vectors is then subsequently organised in a matrix that is used as the basis for computing Σ, the co-variance of naturally occurring trends between all SREX regions. Estimates of Σ are required for computing β as well as the associated confidence intervals. As noted previously (40), using the same approximation of Σ for both tasks can yield biased results. To address this two independent estimates, Σ_1 and Σ_2, are computed following previous research (4,7). To this end the PIC&1860soc simulations are randomly split into two equally sized samples. These are then used to compute the co-variance matrices Σ_1 and Σ_2. Subsequently Σ_1 is used to derive the best estimate of β and Σ_2 is used to estimate the associated confidence intervals. Note that this procedure introduces additional sampling uncertainty that can affect the results of the detection analysis. To account for this sampling uncertainty we repeat the computation with 10000 random splits, yielding 10000 sample of β values and associated confidence intervals. For the lower bound of the 10% - 90% confidence interval, we report the 10th percentile of the sample values. For the upper bound the 90th percentile is used. The median is used to indicate the best estimate of β.

For this study **Ω**, which characterizes uncertainty in regional median trends that are due to model errors and sampling variability, is estimated for the multi model means of both the PIC&HWLM and the HIST&HWLM simulations using a bootstrap (41) approach:

1. For a simulation experiment with N ensemble members.
2. Compute the time series index of interest (i.e. LOW, MEAN, HIGH)
3. Compute regional median trends
4. Repeat $B = 2000$ times:
    1. Draw a bootstrap sample by randomly drawing with replacement N ensemble members.
    2. Compute the mean of the bootstrap sample.
5. This results in B realizations of regional median trends that vary because of sampling uncertainty.
6. Use these bootstrapped regional median trends to compute the covariance matrix **Ω**, representing the uncertainty of the multi model mean trends as well as the spatial dependence of this uncertainty.
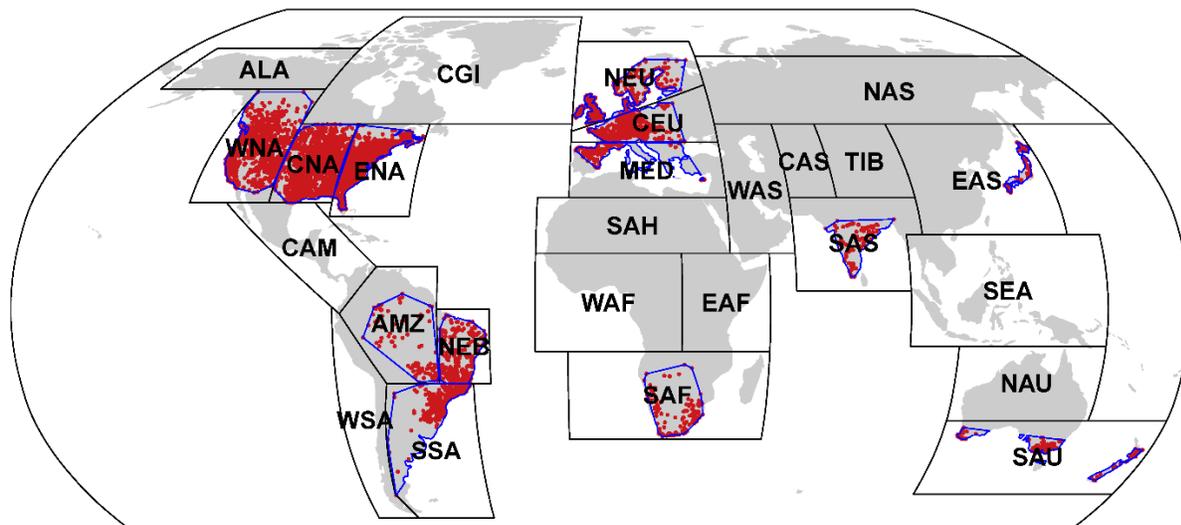
ALA: Alaska/N.W. Canada
AMZ: Amazon (100)
CAM: Central America/Mexico
CAS: Central Asia
CEU: Central Europe (822)
CGI: Canada/Greenland/Iceland
CNA: Central North America (1515)
EAF: East Africa
EAS: East Asia (172)
ENA: East North America (1531)
MED: South Europe/Mediterranean (161)
NAS: North Asia
NAU: North Australia

NEB: North−East Brazil (278)
NEU: North Europe (416)
SAF: Southern Africa (155)
SAH: Sahara
SAS: South Asia (86)
SAU: South Australia/New Zealand (210)
SEA: Southeast Asia
SSA: Southeastern South America (402)
TIB: Tibetan Plateau
WAF: West Africa
WAS: West Asia
WNA: West North America (1402)
WSA: West Coast South America

Fig. S1: Spatial distribution of river flow stations (red dots) held in the GSIM archive that meet the data selection criteria. The black polygons show the sub-continental scale SREX regions, which are used as the spatial support for subsequent analysis. The number of stations in each SREX region is indicated in parentheses (only for those regions that have sufficient stations). Blue polygons are convex hulls around all stations that fall within a SREX region and are used for mapping station-level data.
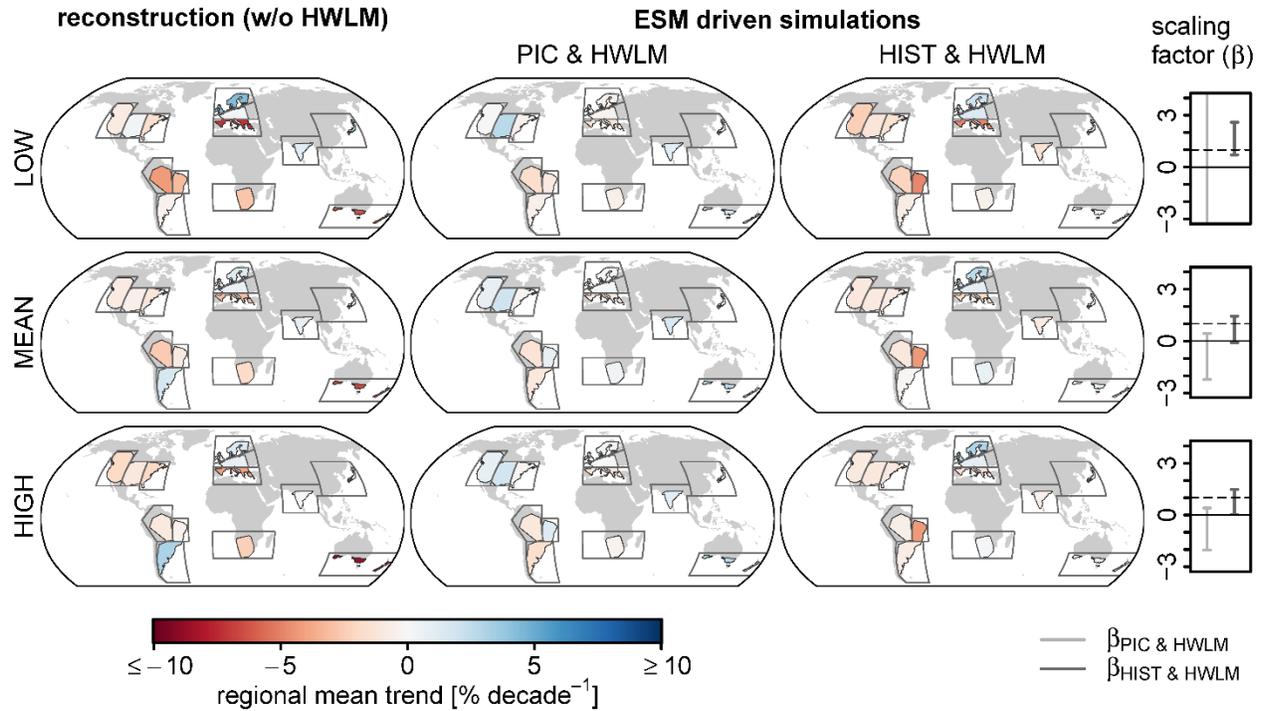
Fig. S2: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions with out (w/o) HWLM. Simulations were masked to match observed coverage (i.e. only grid-cells containing gauging stations were selected and at each site only years with observations were considered). Note that the first column (reconstruction) is identical to the corresponding column of Figure 1 in the main text. Note also that the second and third column (ESM driven simulations) are identical to the corresponding columns of Figure 2 in the main text.
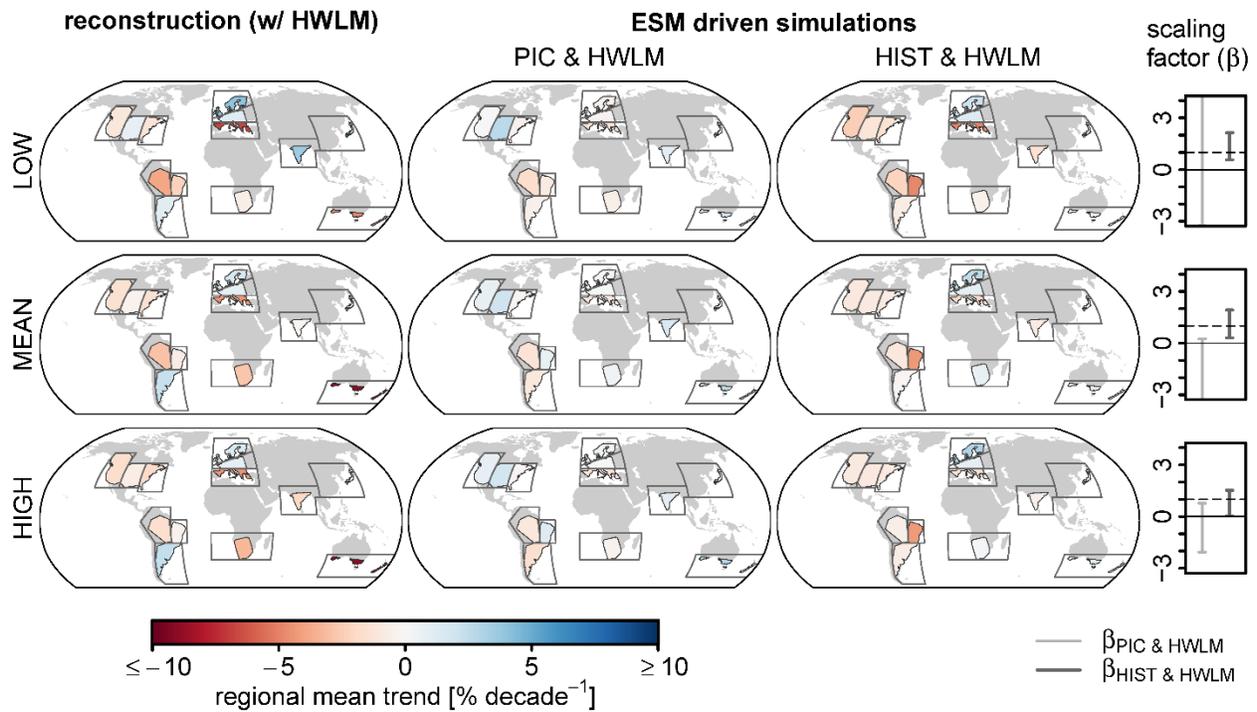
Fig. S3: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions with (w/) HWLM. Simulations were masked to match observed coverage (i.e. only grid-cells containing gauging stations were selected and at each site only years with observations were considered). Note that the first column (reconstruction) is identical to the corresponding column of Figure 1 in the main text. Note also that the second and third column (ESM driven simulations) are identical to the corresponding columns of Figure 2 in the main text.
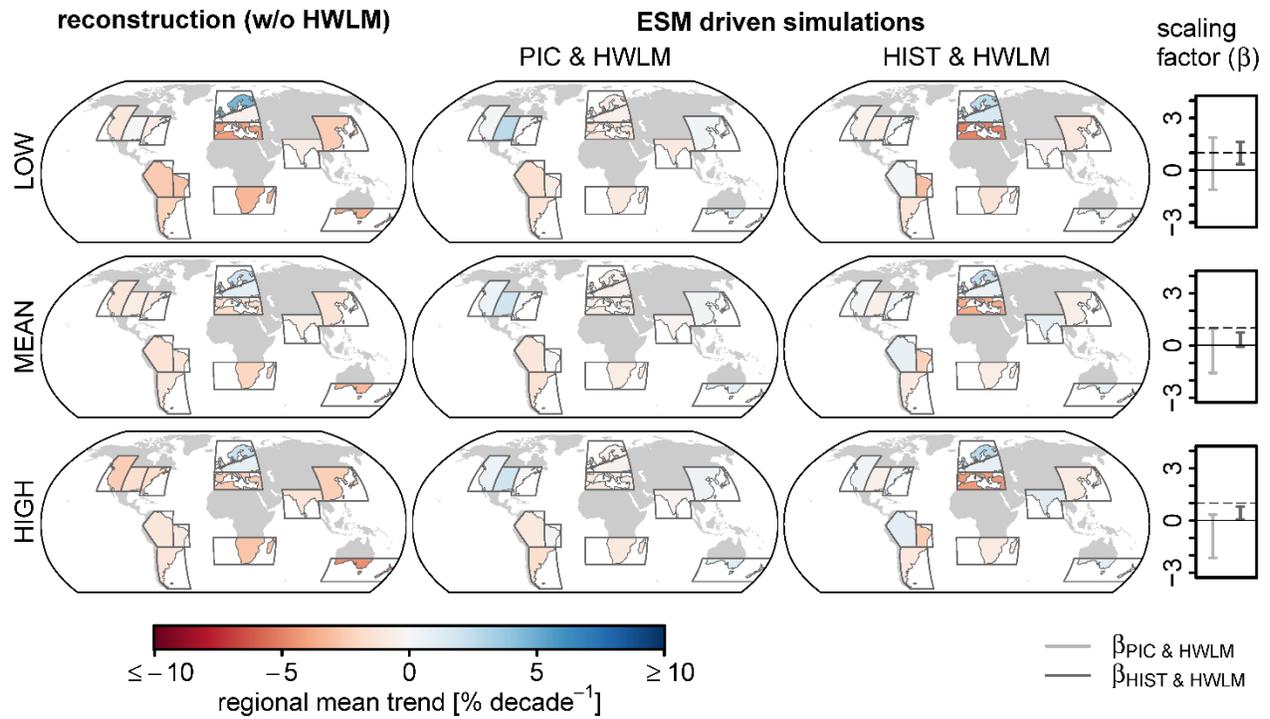
Fig. S4: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions without (w/o) HWLM. In contrast to Figure S2 all grid-cells within the SREX regions that have sufficient observations were selected to assess the influence of spatial sampling biases within the regions.

Fig. S5: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions with (w/) HWLM. In contrast to Figure S3 all grid-cells within the SREX regions that have sufficient observations were selected to assess the influence of spatial sampling biases within the regions on the analysis.
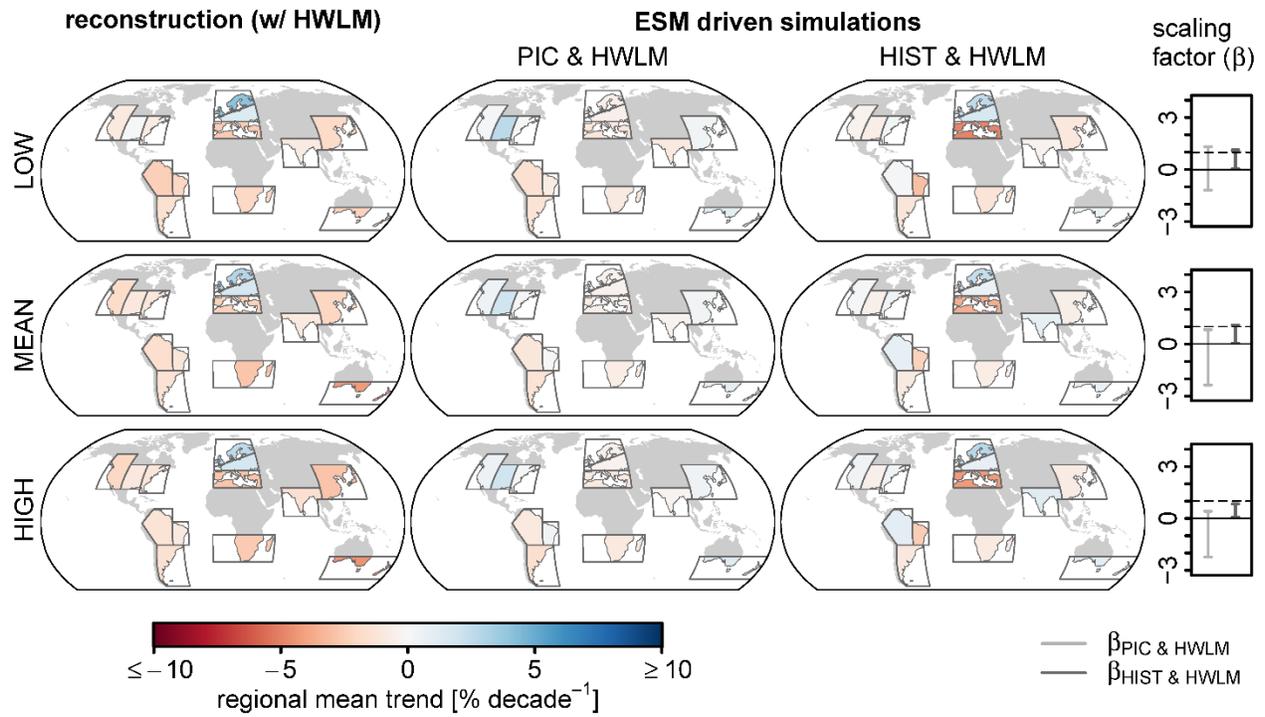
Fig. S6: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions that without (w/o) HWLM. In contrast to Figure S4 all SREX regions were considered to assess the influence of global-scale sampling biases.
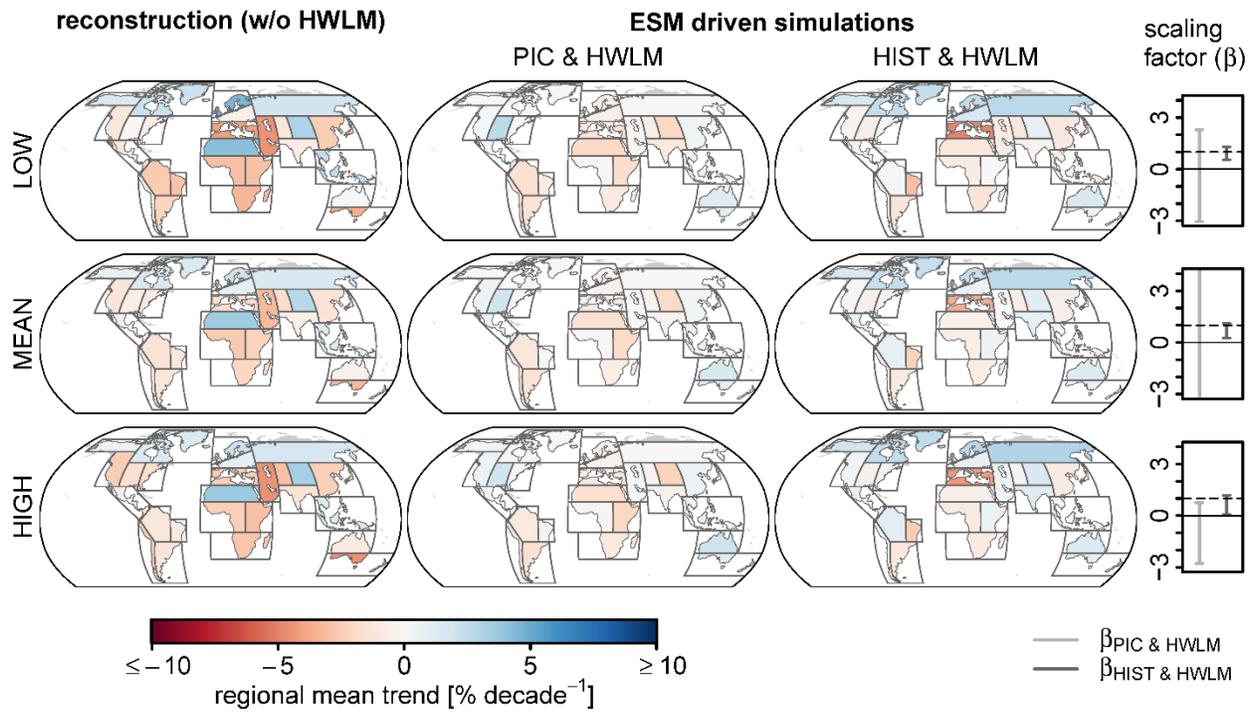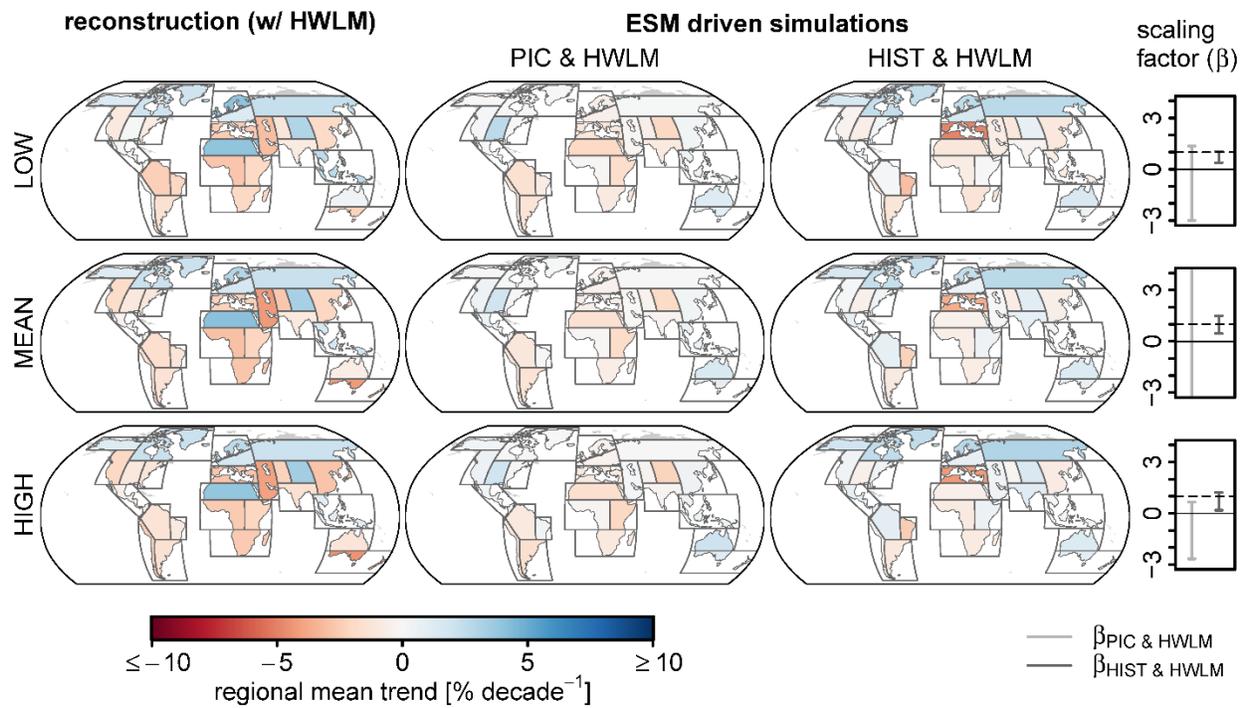
Fig. S7: Similar to Figure 2 in the main text, but with observations replaced by GHM based river river flow reconstructions with (w/) HWLM. In contrast to Figure S5 all SREX regions were considered to assess the influence of global-scale sampling biases.
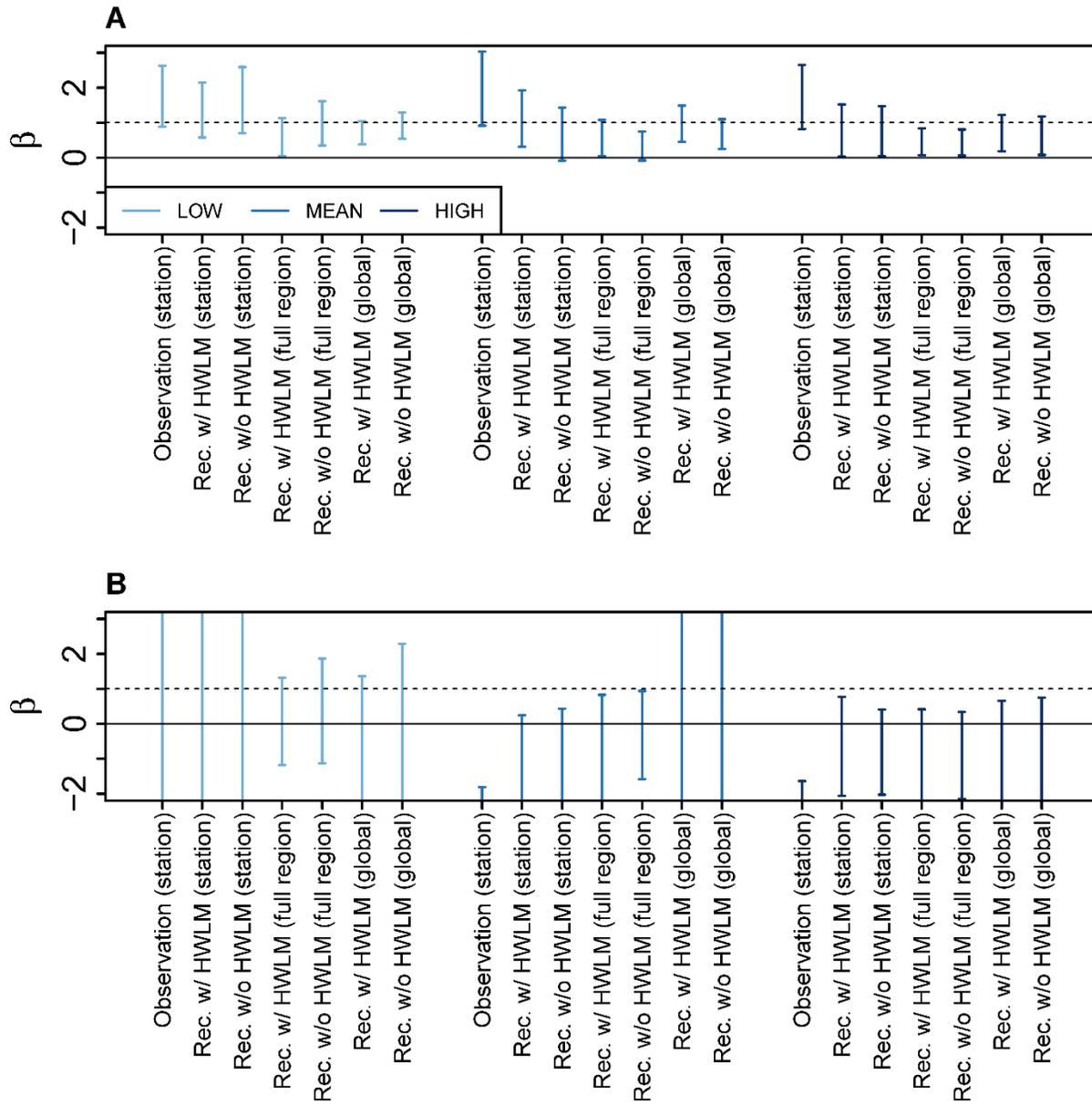
Fig. S8: Summary of the 10% - 90% confidence intervals of scaling factors (β) shown in Figure 2 (main text) and supplementary Figures S2 – S7. A: Scaling factors derived from HIST&HWLM simulations. B: Scaling factors derived from PIC&HWLM simulations. Text in parenthesis indicate the spatial domain of the analysis. (station): The analysis was conducted at locations with gauging stations (Figure 2, S2, S3). (region): The analysis was conducted using all grid-cells within those regions that have sufficient stations (Figure S4, S5). (global): The analysis was conducted using the grid-cells that fall in all possible regions (Figure S6, S7).
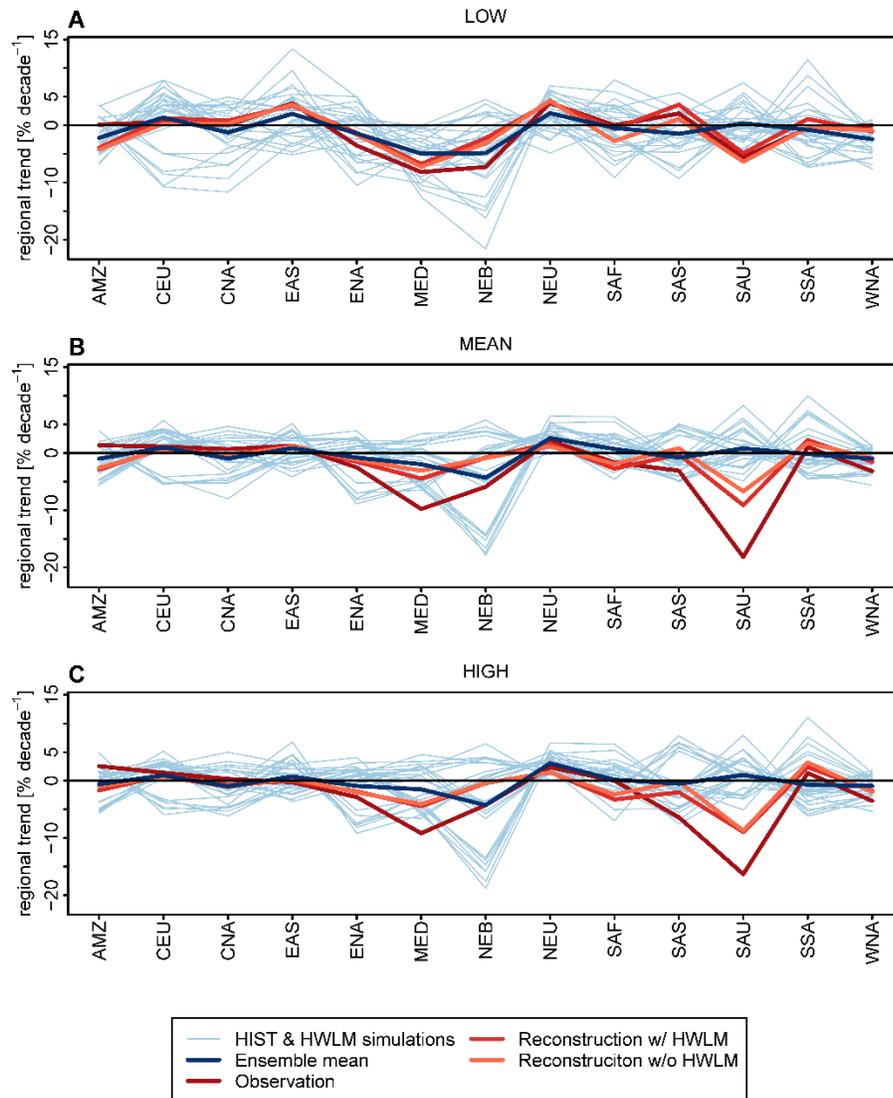
Fig. S9: Comparing observed, reconstructed and simulated spatial variability of regional median trends of annual low (A), mean (B) and high (C) river flow. Regional median trends are computed using data at the locations covered by observations. Overall the observed and reconstructed trends are enveloped by the individual HIST&HWLM simulations, indicating consistency between observed and simulated variability.

Fig. S10: Similar to Fig. S9, but comparing and reconstructed trends to PIC&HWLM simulations.

Fig. S11: Similar to Fig. S9, but comparing reconstructed and simulated regional median trends derived from all grid cells in all SREX regions. Overall, the reconstructed trends are enveloped by the HIST&HWLM simulations, indicating consistency of reconstructed and simulated variability.

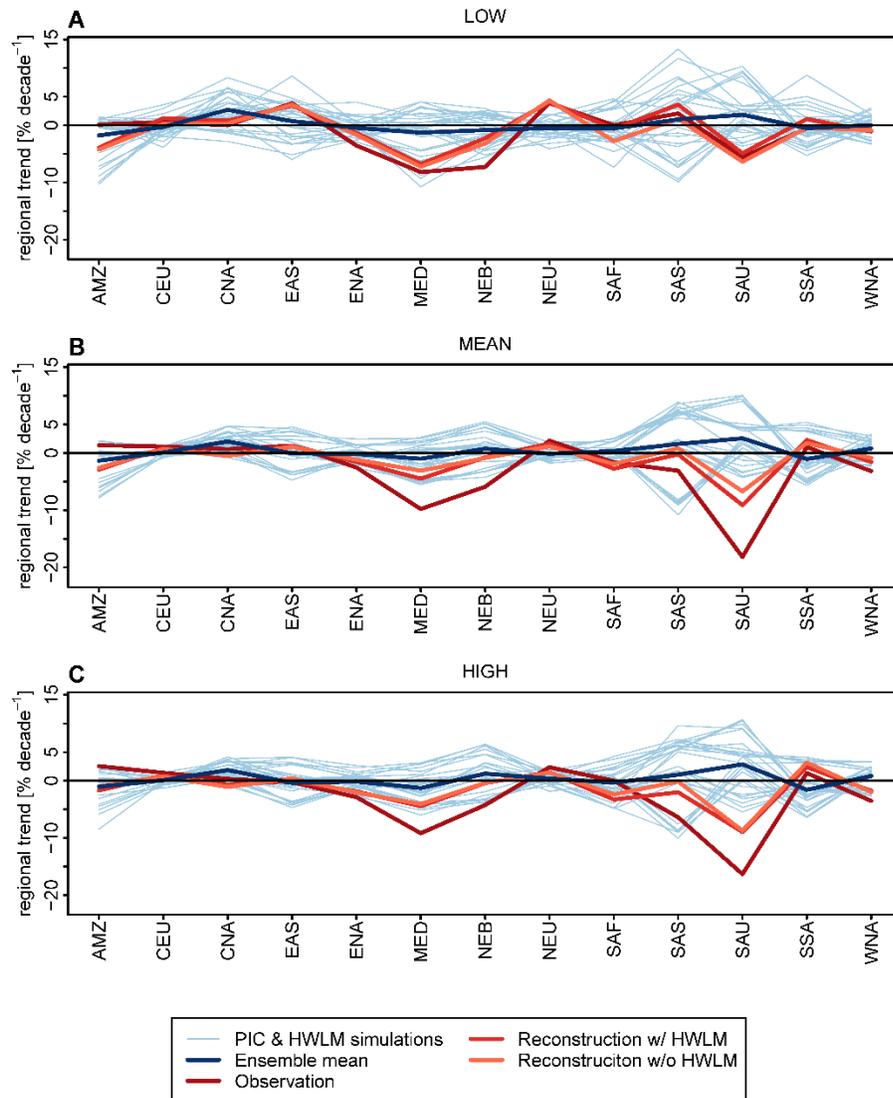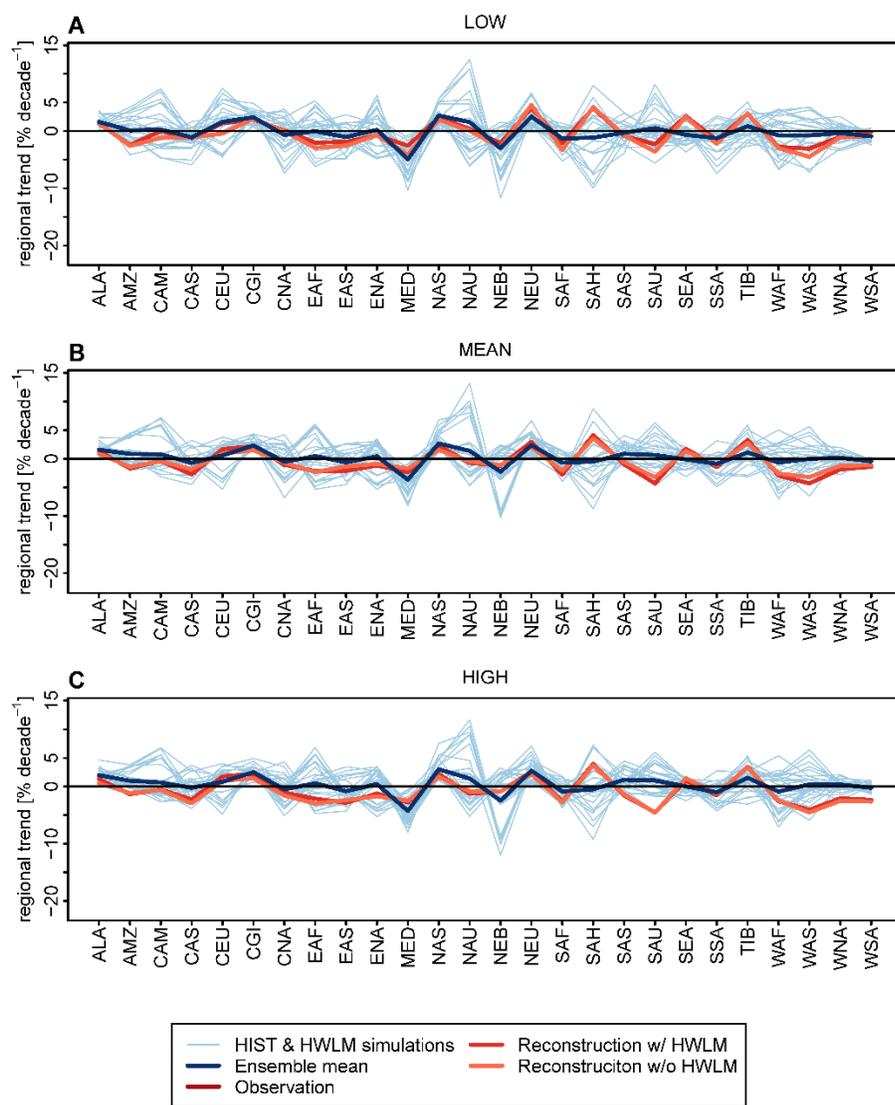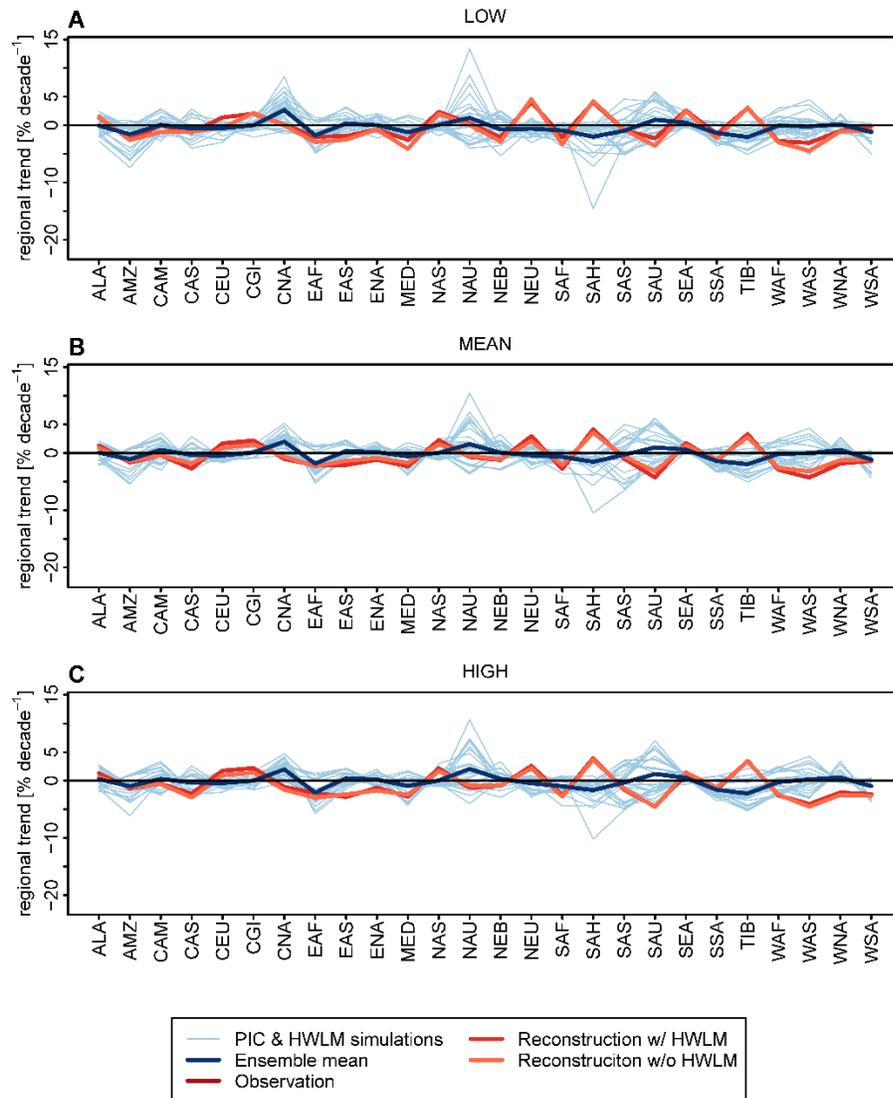**Fig. S12:** Similar to Fig. S10, but comparing reconstructed and simulated regional median trends derived from all grid cells in all SREX regions.

Table S1: Available GHM (rows) – atmospheric forcing (columns) combinations for the ISIMIP2a simulations in the "nosoc" configuration (i.e. w/o HWLM) **that excludes impacts of water and land management. x denotes availability.**

	GSWP3[a]	Princeton[b]	WFDEI[3]
CLM4.0	x	x	x
DBH	x	x	x
H08	x	x	x
LPJmL	x	x	x
MATSIRO	x	x	x
MPI-HM	x	x	x
Orchidee	x	x	x
PCR-GLOBWB	x	x	x
WaterGAP2	x	x	x

[a]Data from the Global Soil Wetness Project Phase 3 (GSWP3), based on the reanalysis data set 20CR and using the bias targets GPCC, GPCP, CPC-Unified, CRU and SRB. The data set covers the period 1901-2010. (see also http://hydro.iis.u-tokyo.ac.jp/GSWP3, "GSWP3" in ref. (35), referred there as GSWP3).

[b]Data from the Terrestrial Hydrology Group at Princeton University, based on the reanalysis data set NCEP/NCAR Reanalysis I and using the bias targets CRU, SRB, TRMM, GPCP & WMO and validated against GSWP2. The data set covers the period 1901‑2012. (see also http://hydrology.princeton.edu/data.pgf.php, "PGFv2.1" in ref. (35), text partly from https://www.isimip.org, referred there as PGMFD v2.1).

[c]Data based on the reanalysis data sets ERA-40 (for WATCH, (42)) and ERA-Interim (for WFDEI, (43)) and using the bias correction target GPCC for monthly rainfall and snowfall sums. The data set covers the period 1901‑2012, where the data for 1901-1978 are taken from WATCH, and from 1979 onwards from WFDEI. ("WFD_WFDEI" in ref. (35), referred there as WATCH-WFDEI).

Table S2: Available GHM (rows) – atmospheric forcing (columns) combinations for the ISIMIP2a simulations in the "varsoc" configuration (i.e. **with (**w/**) HWLM) that accounts for impacts of water and land management. x denotes availability. See Table S1 for an explanation of the different forcing data**.

	GSWP3	Princeton	WFDEI
DBH	x	x	x
H08	x	x	x
LPJmL	x	x	x
MATSIRO	x	x	x
PCR-GLOBWB	x	x	x
WaterGAP2	x	x	x

Table S3: Factors of historical water and land management considered by global hydrology models that were used to derive river flow reconstructions (ISIMIP2a). Further details on the models can be found at https://www.isimip.org/impactmodels/.

	Water use sectors	Reservoir management	Land use change
DBH	irrigation	yes	Yes (irrigation areas)
H08	Irrigation, domestic, industry	yes	Yes (irrigation areas)
LPJmL	Irrigation, domestic, industry, livestock	yes	Yes (irrigation areas)
MATSIRO	Irrigation, domestic, industrial	yes	Yes (irrigation areas)
PRC-GLOBWB	Irrigation, domestic, industrial, livestock	yes	Yes (irrigation areas)
WaterGAP2	Irrigation, domestic, industrial, livestock	yes	Yes (irrigation areas)

Table S4: Number of non-overlapping 40-year long segments in all GHM (rows) – ESM (column) combinations available for the PIC&1860soc experiments.

	GDFL-ESM2M	HadGEM2-ES	IPSL-CM5A-LR	MIROC5
CLM4.5	5	5	5	5
CWatM	5	5	5	5
H08	5	5	5	5
LPJmL	5	5	5	5
MATSIRO	5	4	5	5
MPI-HM	5	0	5	5
PCR-GLOBWB	5	5	5	5
WaterGAP2	5	5	5	5

Table S5: Available GHM (rows) – ESM **(columns) combinations for the simulation experiments** PIC&HWLM. x denotes availability.

	GDFL-ESM2M	HadGEM2-ES	IPSL-CM5A-LR	MIROC5
CWatM	x	x	x	x
H08	x	x	x	x
LPJmL	x	x	x	x
MATSIRO	x	x	x	x
MPI-HM	x		x	x
PCR-GLOBWB	x	x	x	x
WaterGAP2	x	x	x	x

Table S6: Available GHM (rows) – ESM (columns) combinations for the simulation experiments HIST&HWLM. x denotes availability.

	GDFL-ESM2M	HadGEM2-ES	IPSL-CM5A-LR	MIROC5
CWatM	x	x	x	x
H08	x	x	x	x
LPJmL	x	x	x	x
MATSIRO	x	x	x	x
MPI-HM	x		x	x
PCR-GLOBWB	x	x	x	x
WaterGAP2	x	x	x	x

Table S7: Factors of historical water and land management considered by global hydrology models that were used in conjunction with climate model output (ISIMIP2b). Further details on the models can be found at https://www.isimip.org/impactmodels/.

Model	Water use sectors	Reservoir management	Land use change
CLM4.5	irrigation	no	Yes
CWatM	Irrigation, domestic, industrial	yes	yes
H08	Irrigation, domestic, industrial	yes	yes
LPJmL	Irrigation, domestic, industry	yes	Yes (irrigation areas)
MATSIRO	Irrigation, domestic, industrial	yes	Yes (irrigation areas)
MPI-HM	irrigation	no	yes
PCR-GLOBWB	Irrigation, domestic, industrial, livestock	yes	Yes (irrigation areas)
WaterGAP2	Irrigation, domestic, industrial, livestock	yes	Yes (irrigation areas)