Observational Constraints on Basin-Scale Runoff: A Request for Both Improved ESMs and Streamflow Reconstructions

Hervé Douville

Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL108824

Observational Constraints on Basin-Scale Runoff: A Request for Both Improved ESMs and Streamflow Reconstructions

H. Douville[1]

[1]Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Key Points:

- Global 21st century projections of basin-scale runoff remain highly model-dependent within the latest generation of Earth System Models
- Gridded runoff reconstructions and precipitation observations can however be used to constrain the projections through Bayesian statistics
- A model-observation mismatch is underlined regarding recent changes in the annual runoff to precipitation ratio over specific river basins

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Douville,
herve.douville@meteo.fr

Citation:

Douville, H. (2024). Observational constraints on basin-scale runoff: A request for both improved ESMs and streamflow reconstructions. *Geophysical Research Letters*, *51*, e2024GL108824. https://doi.org/10.1029/2024GL108824

Received 15 FEB 2024
Accepted 21 MAY 2024

Author Contribution:

Conceptualization: H. Douville
Data curation: H. Douville
Formal analysis: H. Douville
Investigation: H. Douville
Writing – original draft: H. Douville

Abstract Efforts to predict long-term changes in continental runoff at both global and basin scales generally remain ambiguous. Here we use a global runoff reconstruction and a Bayesian statistical method to narrow uncertainties in runoff projections from the latest generation of global climate models. Three representative tropical river basins are used to illustrate the application and showcase the potential for substantial reduction in modeling uncertainty. Yet, results are fairly sensitive to the selected reconstruction thus highlighting the need for reliable and homogeneized gridded runoff data sets or river discharge measurements. Moreover, climate models do not account for water withdrawals, whose effect on observed runoff should also be removed in order to detect and attribute the hydrological effect of climate change. Finally, and more importantly, most models fail at capturing the observed recent decrease in runoff ratio, which may highlight either model deficiencies or increasing water derivation over the selected river basins.

Plain Language Summary The response of river discharge under the effect of climate change generally remains very uncertain. Bayesian statistical tools and global runoff reconstructions, constrained by flow observations, can nevertheless be used to evaluate the capacity of climate models to simulate the historical runoff and, thus, constrain its future changes at the basin scale. Three representative tropical river basins help illustrate the method. The results are nevertheless sensitive to the choice of the runoff reconstruction, thus emphasizing the need to have good quality flow data, if possible corrected for the direct effects of water withdrawals from the rivers or the aquifers which supply them. Worryingly, the latest generation of global climate models show a systematic underestimation of the downward evolution of the ratio between runoff and precipitation, which could reflect the increasing importance of these withdrawals or the inability of models to capture the rapidly increasing land surface evapotranspiration under climate change.

1. Introduction

Both precipitation (P) and surface evaporation (E) are projected to increase, globally averaged over land, in a warming climate (Douville et al., 2021). Yet, changes in terrestrial water availability (P-E) remain highly model, season and region-dependent, and cannot be simply constrained by narrowing uncertainty in the projection of global mean surface air temperature (hereafter GSAT) (Elbaum et al., 2022). Observed trends in P-E are also very uncertain due to limited evapotranspiration in situ measurements and large inconsistencies across multiple satellite data sets (Robertson et al., 2016). Likewise, atmospheric reanalyzes do not close the water budget and even the latest ERA5 reanalysis from the European Centre for Medium-range Weather Forecasts (ECMWF) shows non-physical hydrological variations due to stepwise changes in the global observation system (Mayer et al., 2021). This lack of reliable P-E reconstructions limits our current understanding of the water cycle response to human emissions of greenhouse gases (Allan et al., 2020).

Beyond P and E, basin-scale runoff may however represent an easier component of the land surface water budget to deal with. While the detection and attribution of global changes in continental runoff changes are also limited by uncertainties in runoff reconstructions and contrasting influences from anthropogenic emissions of greenhouse gases (GHG) and aerosols, basin-scale runoff may be better documented and show stronger changes than its global counterpart (Alkama et al., 2013; Dai et al., 2009). Yet, river discharge is also sensitive to regional changes in land use and water management, which are still misrepresented or not accounted for in most global climate models (GCMs) (Abbott et al., 2019).

Runoff projections are usually based either directly on GCMs or on off-line hydrological models driven by bias-adjusted atmospheric forcings derived from these GCMs. As revealed by successive phases of the coupled model intercomparison project (CMIP, Eyring et al., 2016), such projections remain highly uncertain for three reasons:

scenario uncertainty, modeling uncertainty and internal climate variability. Yet, modeling uncertainty usually plays a dominant role, except for near-term changes and/or low-mitigation scenarios where internal variability has a stronger relative contribution (Douville et al., 2021; Lehner et al., 2020). The situation has not much evolved since at least three generations of GCMs. A recent comparison of CMIP6 and CMIP5 model performance in simulating present-day runoff over the period 1981–2005 indicates for instance that CMIP6 models have not made significant progress on the basin scale (Guo et al., 2022)

Wang et al. (2022) assessed the runoff (R) and runoff ratio (RR, as estimated from the ratio of annual R to annual P) simulated by 23 CMIP6 models during the historical period and in two emission scenarios (SSP1-2.6 and SSP5-8.5). Compared with observed river discharge data, the multi-model median historical runoff was found to have a comparable global mean magnitude (about 0.8 mm/day), and to display a similar spatial distribution over the 1994–2015 period. During the twenty-first century, the median global mean R is projected to increase while the RR would slightly decrease except for the long-term under the SSP5-8.5 high-emission scenario. The latter finding provides further support to the nonlinearity of the runoff response to global mean temperature changes (Cui et al., 2023; Zhang et al., 2018).

These findings are consistent with those derived from previous generations of GCMs. Tang and Lettenmaier (2012) showed for instance that the runoff sensitivity implied by CMIP3 simulations varied substantially across both models and river basins. Although, the runoff response along the trajectory of global mean warming was found to be approximately linear over many land areas, regional deviations from linearity were apparent especially for small global warming increments. Such nonlinearities may partly arise from the timescale-dependence of runoff sensitivities to both temperature and precipitation changes, as emphasized by Dutot and Douville (2023).

Another multi-model study found a global mean runoff increase of 2.9% per 1°C of global warming in CMIP5 against only 1.9% in CMIP3 (Zhang et al., 2014). This difference may partly arise from the early development of Earth System Models (ESMs), whose stomatal closure effect on evapotranspiration is not totally offset by an enhanced plant photosynthesis. Among the CMIP5 models, the ESMs however did not show a reduced spread compared to the other models. Similarly, no step change in model performance or model consensus was found between CMIP5 and CMIP6 (Miao et al., 2023). Globally averaged, the multi-model ensemble mean annual runoff was found to increase by +16.1% under the SSP5-8.5 CMIP6 scenario (against +10.8% under the RCP8.5 CMIP5 scenario), yet with large uncertainties dominated by modeling uncertainty (Wu et al., 2024).

In the present study, the focus is on annual changes in total (i.e., surface and subsurface) runoff, as simulated by a subset of 35 CMIP6 models and spatially averaged over a few selected river basins representative of a variety of regional climates. The main objective is to constrain the R and RR projections, using runoff reconstructions and precipitation observations. Section 2 describes the data and summarizes the employed statistical method. Section 3 illustrates the key results with further support provided in the Supplementary Information (SI). Section 4 highlights the key role of the observations and, beyond the development and evaluation of improved ESMs, advocates for more reliable naturalized streamflow reconstructions.

2. Data and Methods

The two main observed 0.5° global gridded data sets used are monthly precipitation (P) from the Global Precipitation Climatology Centre (GPCC, Schneider et al., 2022), and monthly runoff (R) from the Global RUNoff reconstruction (GRUN, Ghiggi et al., 2019). GRUN is based on a machine learning method that estimates R from antecedent temperature and P observations. The algorithm was trained with monthly observations at the scale of relatively small catchments (<2,500 km^2) and was then applied to global temperature and P data from Phase 3 of the Global Soil Wetness Project (GSWP3) covering the 1902–2014 period. Both GPCP and GRUN data have here been aggregated at the basin scale using a suitable mask, at the same 0.5° resolution, from the Total Runoff Integrated Pathway (TRIP, Oki & Sud, 1998). In addition, the 1850–2022 HadCRUT5 global mean surface temperature (hereafter GMST)—a combination of near-surface air T over land and sea ice and of sea surface temperature over the ocean—was also used as a surrogate for observed GSAT.

The bias-adjusted GSWP3 precipitation forcing used by Ghiggi et al. (2019) is slightly different from our reference GPCC product. Yet, GRUN was carefully cross-validated with monthly runoff observations from the Global Streamflow Indices and Metadata Archive (GSIM). More than 8,000 stations worldwide were selected

after screening for spurious observations. River discharge time series were also corrected for temporal changes in instrumentation, recalibration of streamflow rating curves. Interestingly, the data were also automatically adjusted to account for flow regulation (i.e., dam construction) and other human activities (i.e., irrigation), which are generally ignored in GCMs.

It should be emphasized that the machine-learning method allowed Ghiggi et al. (2019) to produce an ensemble of 50 gridded monthly runoff estimates over the 1902–2014 period. The use of different training observations has indeed the potential to generate different outcomes if the model is not able to diagnose the relationship between the response (R) and the predictors (P and temperature) consistently. The 50 runoff reconstructions were generated using a Monte Carlo approach in which the random forest algorithm was trained using a random 60% subset of the grid cells with observations. Unless specified hereafter, the GRUN reconstruction refers to the ensemble mean of the 50 realizations, but the individual members have been also used to assess how the reconstruction uncertainties may affect the range of constrained projections (cf. SI).

An alternative monthly reconstruction of P, R and RR has been also derived from the ECMWF ERA5-Land reanalysis over the 1950–2021 period (Muñoz-Sabater et al., 2021a, 2021b). This global high-resolution data set for the land component of the fifth generation of European ReAnalysis (ERA5) describes the evolution of the water and energy cycles over land in a consistent manner over the production period, and is potentially suitable for trend analyses. It was achieved through global numerical integrations of the ECMWF land surface model driven by downscaled meteorological forcings. Evaluation against independent in situ observations and satellite-based data sets has proven the added value of ERA5-Land in the description of the hydrological cycle, including an overall better agreement of river discharge estimations with available observations (Muñoz-Sabater et al., 2021a, 2021b). Other river discharge observations (GRDC) or runoff reconstructions (Hobeichi et al., 2019) could have been used, but include many missing data or may not be long or homogeneous enough for the purpose of our study.

Moving to CMIP6 models, both historical (1850–2014) and 21st century (2015–2100) simulations providing monthly outputs for precipitation and runoff have been downloaded from a data portal of the Earth System Grid Federation (https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/). The SSP5-8.5 high emission scenario was selected in order to maximize the signal-to-noise ratio. This choice enabled the use of only one realization for each model (among a total number typically ranging from 1 to 25). While considering all available members can be useful to improve the estimation of the forced model response, it has been shown in a previous study (Dutot & Douville, 2023) that it only makes a slight difference when focusing on annual precipitation or runoff aggregated over large river basins. Such a spatial aggregation has been completed after projecting the 0.5° TRIP mask on the model-dependent native grid. The simulated historical warming was also diagnosed from raw model outputs, but using near-surface air temperature only (GSAT) rather than combined with sea surface temperature over the ocean (GMST).

Regarding the statistical method, we have used the Kriging for Climate Change (hereafter KCC) toolbox developed by Ribes et al. (2021) and Qasmi and Ribes (2022a, 2022b). It is based on Bayesian statistics where a *prior* distribution, $\pi(x)$, of the forced response to anthropogenic forcings is derived from raw model outputs and constrained directly with observations (here both observed global mean surface temperature and basin-wide average runoff). In the present study, the x vector denotes the forced component of the basin-scale and water-year average of precipitation or runoff, or of the ratio between these aggregated variables, as estimated from the model outputs on their native grids. The *prior* (i.e., the unconstrained forced response) is estimated using a Generalized Additive Model (assuming the additivity of the model responses to individual forcings) and a simple Energy Budget Model (allowing us to diagnose the runoff response to volcanic eruptions; for more details, see supplementary materials from Ribes et al., 2021). For the sake of simplicity and given the limited number of independent models, this *prior* is assumed to follow a normal distribution and thus only needs an estimate of the ensemble mean and spread. Next, observations y (here the entire record of global mean surface temperature and basin-scale precipitation, runoff or runoff ratio reconstructions) are used to derive a *posterior* distribution (after constraint). We assume that observations can be described as: $y = \mathcal{H}x + \varepsilon$, where $\mathcal{H}$ is a pseudo-observation operator allowing to extract the part of x observed in y, and ε represents both internal variability and observational errors (if available). Since $\pi(x)$ and ε are supposed to follow normal distributions, the *posterior* can be easily derived using the Gaussian conditioning theorem (see supplementary materials from Ribes et al., 2021).

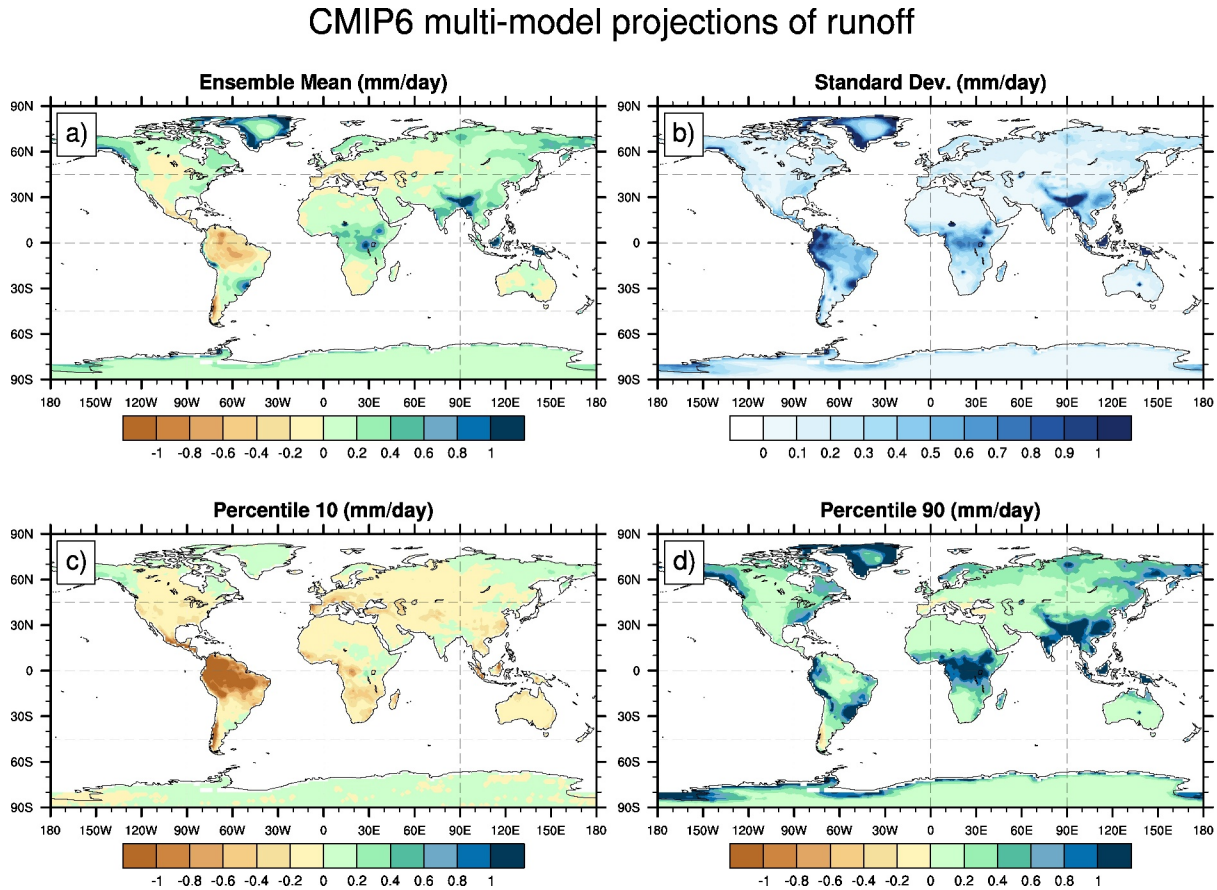

Figure 1. Projected changes in total runoff (mm/day) under the SSP5-8.5 high-emission scenario. All CMIP6 ensemble statistics are derived from the differences between the 2081–2100 and 1995–2014 water-year climatologies. To highlight the inter-model spread, we not only show (a) the multi-model ensemble mean anomalies, but also (b) their local standard deviation and (c), (d) their 10% and 90% local percentiles, respectively. Only one realization is used for each model.

The KCC method has previously been used to constrain CMIP6 runoff projections over the Arctic (Dutot & Douville, 2023). It was then also compared to an alternative statistical method assuming similar runoff sensitivities, to precipitation and temperature, at interannual and climate change timescales (Lehner et al., 2019). This assumption was shown to be incorrect and a "pseudo-observation" framework (in which independent model outputs are used as a surrogate for both historical and future climate observations) was used to demonstrate the higher skill of KCC compared to this more empirical approach. The Bayesian method also proved to be not much sensitive to the choice of the *prior* distribution (CMIP6 vs. CMIP5 models). Yet, it did not lead to a substantial narrowing of runoff uncertainties over the Arctic given the strong observed interannual variability and the limited signal-to-noise ratio in the GRUN runoff reconstructions.

3. Results

Figure 1 illustrates both the ensemble mean and inter-model spread of projected changes in total runoff at the end of the 21st century. Not surprisingly given the corresponding changes in total precipitation (Figure S1 in Supporting Information S1, the ensemble mean runoff (Figure 1a) shows an increase over the wet tropics and in the high-latitudes. In contrast, absolute runoff changes show an ensemble mean decrease over Amazonia as well as, though to a lesser extent, in some mid-latitude regions and in the subtropics. Note that relative rather than absolute changes in runoff may show quite different regional and seasonal patterns (cf. Figure 8.18 in Douville et al., 2021) given corresponding variations in baseline (1995–2014) runoff. For the sake of simplicity, the focus is here only on absolute and water-year changes and, thus, on the tropics where the largest anomalies are found but are also strongly model-dependent (cf. Figure 1b). As shown by 10th and 90th local percentiles of the model distribution

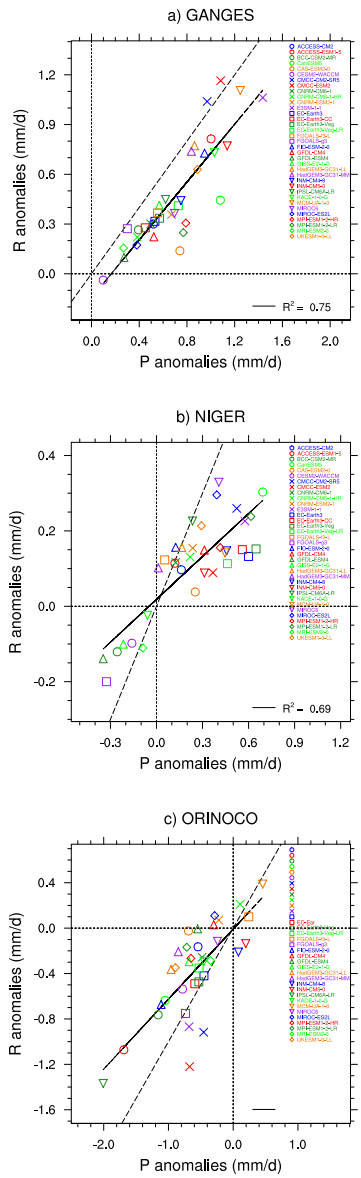

Figure 2. Scatterplots of future absolute changes in runoff (*y*-axis) versus precipitation (*x*-axis). All basin-scale changes (in mm/day) are estimated as the difference between the 2081–2100 and 1995–2014 climatologies, as averaged across three representative river basins: (a) Ganges, (b) Niger and (c) Orinoco, respectively. The black solid line refers to the linear regression fit and the black dashed line refers to the 1:1 bisector.

(Figures 1c–1d), even the sign of the runoff response remains uncertain in most regions, except the robust increase found in the northern high latitudes and a very limited fraction of the tropical monsoon regions.

Figure 2 shows scatterplots of basin-scale aggregated runoff versus precipitation climatological anomalies, again averaged along the water year that is mostly relevant for the northern high latitudes (due to the snow influence) but is also suitable for the boreal monsoon regions (with maximum precipitation from June to September). Three illustrative river basins are shown, Ganges, Niger and Orinoco. Not surprisingly, they all show a statistically significant linear relationship (as highlighted by the regression lines in Figure 2) between runoff and precipitation anomalies across the CMIP6 ensemble. Yet, this relationship is basin-dependent (as indicated by the slope of the regression line) and suggests a less-than-one runoff amplification, thereby also indicating changes in land surface evapotranspiration.

A brief look at other river basins in the northern hemisphere (Figures S3 and S4 in Supporting Information S1) suggests that the drivers of mid- and high-latitude runoff anomalies are generally more complex than in the tropics, where the concentration of annual precipitation over a few months (e.g., during the monsoon season over India and West Africa) may lead to a stronger relationship between runoff and precipitation anomalies across the CMIP6 ensemble (e.g., Chang et al., 2014). A more even distribution of monthly precipitation across the water year has also the potential to give more weight to model-dependent changes in evapotranspiration in the basin-scale water budget. The relationship between R and P anomalies is even weaker in the high latitudes where potential changes in the ratio between liquid and solid precipitation and, thus, in snowfall and snowmelt, represent another potential modeling source of uncertainty.

The regression slopes shown in Figure 2 should not be confused with changes in RR which are computed as the difference of two R/P ratios rather than as the ratio of R and P anomalies. Figure S2 in Supporting Information S1 shows similar scatterplots as in Figure 2 but for changes in RR versus precipitation. All basins show a weak positive relationship, albeit only significant over Ganges, which suggests that enhanced precipitation favors an increase in RR (i.e., soil moisture saturation or precipitation rate exceeding the infiltration capacity), whose response is however also influenced by other factors. Looking for them is beyond the scope of the present study, but several potential candidates can contribute such as land use change (as considered by some but not all CMIP6 models), changes in precipitation seasonality (Douville et al., 2021) and changes in daily precipitation intensity (Douville & John, 2021).

As explained in the method section, KCC has been used to constrain the simulated evolution of the water-year runoff at the basin scale, from 1850 to the end of the 21st century (water year 2099 extends from October 2099 to September 2100). Figures S5 to S7 in Supporting Information S1 illustrate the results obtained over three representative tropical basins — Ganges, Niger and Orinoco — using either the GMST (panel b) or the GRUN (panel c) observational constraint only. Also shown are the results for GSAT projections constrained with GMST (panel a, not basin-dependent), as well as the results for runoff projections constrained with both GMST and GRUN (panel d). Only one realization of each of the 35 CMIP6 model is used to produce the prior distribution, but all available members of the GMST (200) and GRUN (50) reconstructions are used to constrain, separately or simultaneously, the posterior distribution. For all river basins, GMST observations exert a weak constraint on the projected runoff, as revealed by the limited narrowing of the 5–95% confidence interval of the posterior versus prior distribution at the end of the 21st century. The GRUN and combined GMST-GRUN

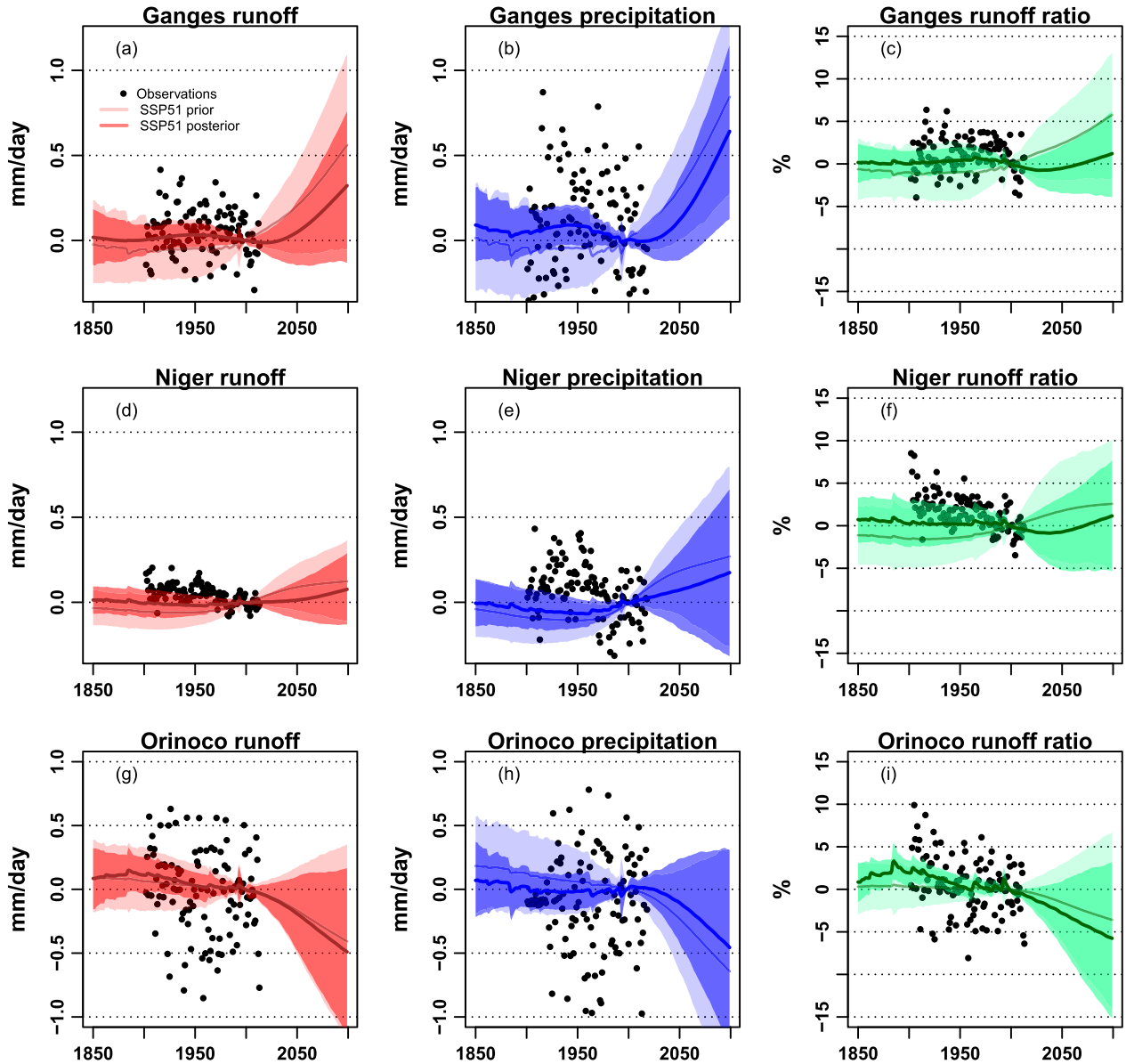

Figure 3. Constrained (i.e., posterior) versus unconstrained (i.e., prior) hydrological anomalies under the SSP5-8.5 high-emission scenario. Constrained versus unconstrained changes in precipitation (mm/day), runoff (mm/day) and runoff ratio (%) over three tropical river basins: (a), (d), (g) Ganges; (b), (e), (h) Niger; and (c), (f), (i) Orinoco. Black dots correspond to the GRUN water-year runoff anomalies. All projections are constrained by the entire record of both HadCRUT5 GSAT observations (1850–2022) and of the corresponding basin-scale averaged variable using GPCP and/or GRUN observations (1901–2014). The thick lines denote the best estimate of each distribution (i.e., the ensemble mean) while shadings denote the corresponding 5–95% confidence intervals.

constraints are however stronger, especially over Ganges where they narrow the late 21st century confidence interval by up to 30%.

The KCC method was applied successively to R, P and RR time series, using consistently both a global (GMST) and basin-scale (from GRUN, GPCC or GRUN/GPCC) observational constraint. The results are shown in Figure 3 for the same three tropical rivers. Here, only the ensemble mean GRUN reconstruction was used, in line with the single GPCC estimate for observed precipitation and, thus, also leading to a single observational estimate of RR (ratio between R and P after spatial and annual aggregation using the water year definition).

Looking first at the first column in Figures 3a–3g (in red), the constrained runoff projections are quite consistent with the results previously shown in Figure 2 and Figures S5 and S6 in Supporting Information S1. Assuming no

observational uncertainty thus only makes a marginal difference when using the GRUN runoff reconstruction. This can be simply explained by the fact that the machine learning algorithm provides robust reconstructions when using 60% of the observed river discharge measurements. The GRUN ensemble spread is thus weak compared to the runoff interannual variability (also considered in KCC, Ribes et al., 2021). Yet, it does not account for instrumental and measurement uncertainties (Ghigghi et al., 2019) and may thus represent a low estimate of total uncertainties in runoff reconstructions.

Interestingly, the three selected basins show contrasted ensemble mean runoff projections with increased values over Ganges, limited changes over Niger, and decreased values over Orinoco. Yet, the *prior* (i.e., unconstrained) ensemble distribution of the simulations, as derived from a normal distribution assumption, highlights large modeling uncertainties, including about the sign of the response. Constraining the projections with both GMST and GRUN leads to a moderate narrowing of modeling uncertainties, but can shift the whole distribution and the ensemble mean, especially over Ganges where the increase in future runoff is much less in the *posterior* compared to the *prior* distribution.

Looking at precipitation projections in Figures 3b–3h (in blue), the combined GMST-GPCC observational constraint again leads to a limited narrowing of modeling uncertainty, but a possible shift of the ensemble mean from the *prior* to the *posterior* distribution. In line with the constrained runoff projections, the constrained ensemble mean precipitation changes are lower than unconstrained anomalies over Ganges and Niger, but higher over Orinoco despite no significant change in ensemble mean runoff. This paradox may have several reasons, including the GSWP3 rather than GPCC precipitation predictor used to produce GRUN or a possible overestimated decrease in evapotranspiration that may offset the overestimated decrease in precipitation.

The results shown in Figures 3c–3i (in green) for the annual RR are probably the most interesting. The observed signal-to-noise ratio is indeed stronger, and allows KCC to narrow model uncertainties more efficiently, than for R and P separately. This is particularly clear over Ganges where the projected ensemble mean RR increase is much weaker after constraint and the 5–95% confidence interval is reduced by about one third (typically the same order of magnitude than when constraining GSAT with GMST in the top left panel of Figures S5–S7 in Supporting Information S1) at the end of the 21st century. A significant reduction of the constrained versus unconstrained RR is obtained for the three tropical rivers, regardless of the sign of the associated runoff and precipitation changes. This behavior is also found, though to a lesser extent, in the northern mid-and-high latitudes (Figures S8 and S9 in Supporting Information S1).

4. Discussion and Conclusion

There are multiple evidence that the water cycle will further intensify with continuing global warming (Douville et al., 2021). Yet, and despite a robust increase of total precipitable water by around 7%/°C (e.g., Douville, Ribes, & Bock, 2022), projected changes in P-E patterns cannot be simply interpreted as a typical "wet gets wetter, dry gets drier" response (Allan et al., 2020). Future changes in energy and/or soil-moisture limited surface evapotranspiration, but also in runoff efficiency (e.g., here defined as the RR ratio between annual runoff and annual precipitation), can contribute to more complex P-E responses than simply due to the expected enhancement of the climatological horizontal moisture transport. Regional changes in P-E may also arise from changes in large-scale atmospheric circulation and, thus, cannot be simply scaled with the increase in GSAT (Elbaum et al., 2022). In some way, water therefore remains a blind spot in climate change policies and, thus, needs more accurate quantitative assessments to build more resilient adaptation and mitigation strategies (Douville, Allan, et al., 2022).

The present study provides further evidence of large modeling uncertainty in runoff projections at the basin scale, even in a high-emission scenario where such changes are less obscured by the influence of internal climate variability (Lehner et al., 2020; Wu et al., 2024). Constraining such projections is thus urgently needed but will not rapidly arise from model development, as highlighted by the limited progress between successive CMIP generations of GCMs. The KCC Bayesian statistical method may thus represent an alternative, top-down rather than bottom-up approach, to take advantage of the available projections and observations by constraining the former by the latter and, thus, provide more reliable climate information (i.e., reduced confidence intervals) for the purpose of adaptation.

10.1029/2024GL108824

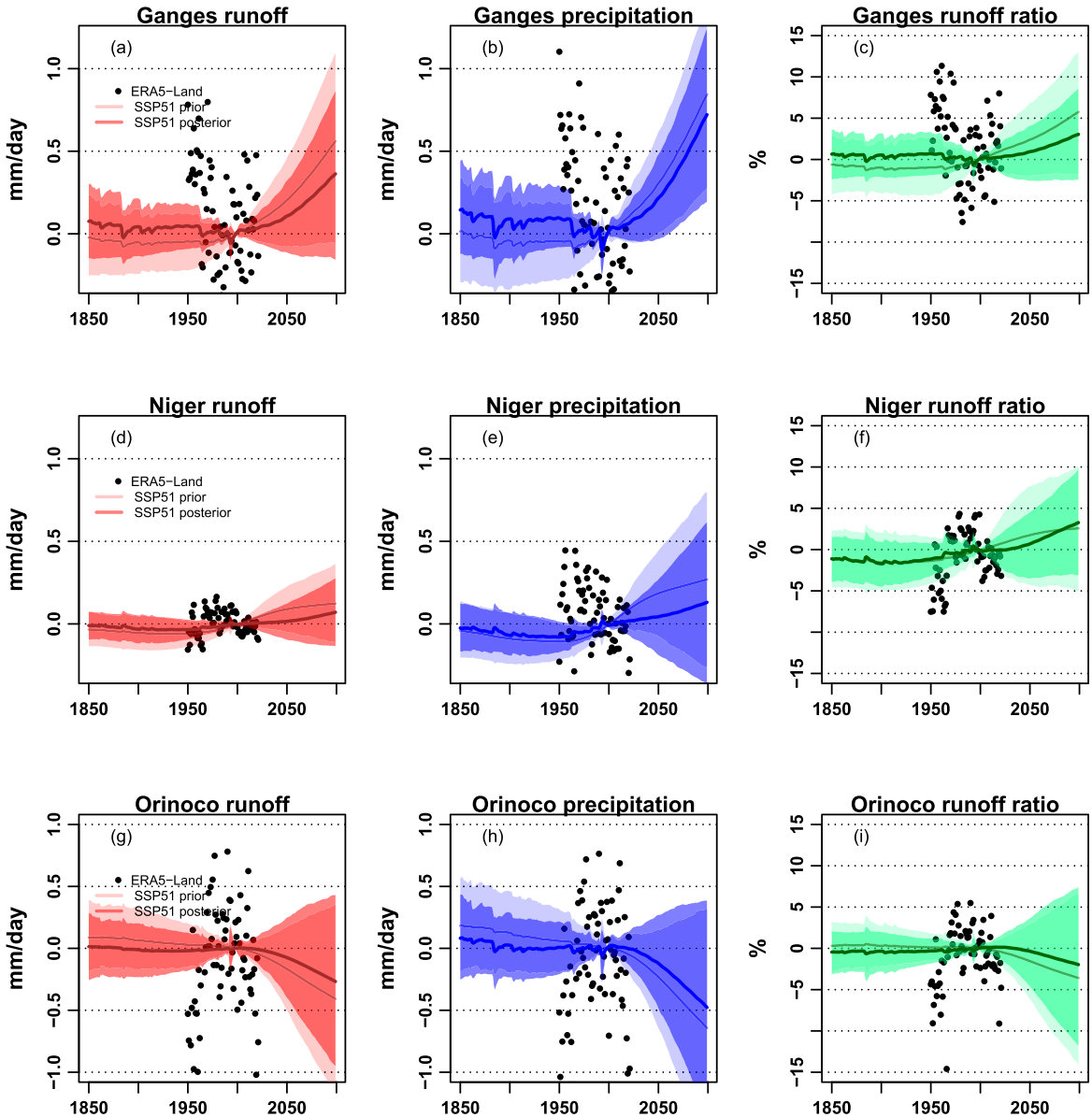

Figure 4. Same as Figure 3 but using ERA5-Land instead of the GPCC precipitation observations and the GRUN runoff reconstruction.

The results of KCC, however, may be sensitive to the quality of the observations, even more than to the choice of the prior distribution (Dutot & Douville, 2023). This obvious limitation can be exemplified by using the 1950–2022 monthly data from ERA5-Land (Muñoz-Sabater et al., 2021a, 2021b) rather than GPCC and GRUN as an alternative observational constraint for precipitation and runoff, respectively. The posterior distributions (Figure 4) are then quite different from those obtained in Section 3 (Figure 3). Such differences do not only arise from the shorter ERA5-Land record, but also from obvious discrepancies with GRUN and/or GPCC over the 1950–2014 overlapping period. In particular, ERA5-Land shows a much stronger multidecadal variability of tropical precipitation, which is not found in GPCC but has obvious consequences on the corresponding runoff.

Assessing the relative performance of ERA5-Land and GRUN in capturing the observed basin-averaged runoff (evaluated from streamflow measurements at the basin outlets) is beyond the scope of the present study. It has been however shown that the state-of-the-art ERA5 reanalysis still suffers from inhomegeneities in water budget quantities, including precipitation and evaporation, especially in the late 1990 s (Mayer et al., 2021). Since the

ERA5-Land reanalysis was itself driven by ERA5 precipitation, it is not surprising to see a strong multidecadal contrast between the late 20th and early 21st century observations in Figure 4. Our results thus suggest that ERA5-Land may not be suitable for constraining hydrological changes. They also highlights the importance of on-going efforts to provide longer and more homogeneous data sets than generally available (Ghiggi et al., 2019).

Despite strong observational uncertainties, both GRUN and ERA5-Land suggest a recent decrease in the observed RR which is not captured by most historical simulations. This mismatch can arise from multiple non-exclusive reasons: (a) internal climate variability which could be better sampled by using all available members of all available models; (b) model deficiency to capture what may be a forced runoff response to radiative anthropogenic forcings (including emissions of greenhouse gases and aerosols); (c) model inadequacy to account for water withdrawals (for irrigation or other purposes) not entirely removed from our reference GRUN reconstruction. Further investigation will be needed to discriminate between these confounding factors, potentially using larger ensembles of model simulations driven by individual rather than all external radiative forcings.

Regarding the second hypothesis, it should be however noticed that our KCC results are somehow consistent with potential model deficiencies noticed in previous studies. Yang et al. (2018) assessed runoff simulations from CMIP5 (rather than CMIP6) and, surprisingly, found that runoff is projected to increase over the majority of the globe despite a drying atmosphere (as diagnosed by the aridity index). This apparent disconnection between the trends in atmospheric drying and in continental runoff may be due to the fact that the observed recent near-surface atmospheric drying is itself underestimated by most CMIP5 and CMIP6 models (Allan & Douville, 2024; Douville & Willett, 2023; Simpson et al., 2023). This model deficiency may then explain why runoff is increasing faster than ET over most river catchments and, thus, why RR is increasing with global warming unlike in the GRUN reconstruction. Other model deficiencies can also contribute to the contrasted trends in RR between CMIP6 simulations and GRUN/GPCC observations, such as the evapotranspiration partitioning due to plant physiological effects (e.g., Yang et al., 2023) or the representation of precipitation intermittency in coarse resolution climate models (e.g., Scheff et al., 2022).

Compared to previous attempts to constrain runoff projections (e.g., Lehner et al., 2019; Yang et al., 2017), our study is based on a more recent generation (CMIP6 rather than CMIP5) of climate models, including a greater proportion of ESMs accounting for the physiological effect of atmospheric CO_2 on plant transpiration. More importantly, the KCC method was proven to be more robust than a multiple regression fitted on interannual variability given the timescale-dependent runoff sensitivity to both temperature and precipitation (Dutot & Douville, 2023). The Bayesian Model Averaging (BMA) method proposed by Yang et al. (2017) is in principle much closer to KCC, but the projected runoff was then only contrained by runoff data over the 10-year 1986 to 1995 period so that the model weights were calculated based on the capability to capture the recent climatology rather than the full 1902–2014 historical evolution as in the present study.

To sum up, we do not only need improved ESMs that may, more or less rapidly, lead to more robust hydrological projections for a given GHG scenario and/or global warming level, but also — and more urgently — more reliable multidecadal observations and reconstructions to constrain the projections and, thus, make the best possible use of available climate data to guide adaptation strategies and define relevant mitigation targets. Our preliminary results may provide further motivations for developing and improving long-term homogeneous observational products, including for instance the extension of the available VASClimO data set (1951–2000, Beck et al., 2005) as a variance adjusted version of GPCC, more suitable for trend analysis. Naturalized river flow measurements are also urgently needed and may represent a useful alternative to the introduction of more or less sophisticated irrigation schemes in current ESMs. Beyond the evaluation of the historical simulated runoff, they could also be used to adjust the observed precipitation products in data sparse regions (Beck et al., 2020). Given the ubiquituous climate change signals that are expected to emerge at the regional scale (Hawkins et al., 2020), more attention should be paid to this fundamental monitoring and related research activities. Model intercomparison projects such as CMIP are still needed for understanding, interpreting and attributing observed changes. Yet, they are not expected to predict the forced hydrological responses to anthropogenic climate change (and corresponding unforced variations) with accuracy, unless partially constrained with reliable observations and rapidly improving statistical methods (Hegerl et al., 2021).

Data Availability Statement

All model outputs from the CMIP6 historical and scenario experiments are freely accessible at https://pcmdi.llnl.gov/CMIP6/ or at https://esgf-node.llnl.gov/projects/esgf-llnl/. The GPCC, GRUN and ERA5-Land data sets are available from Schneider et al. (2022), Ghiggi et al. (2019), and Muñoz-Sabater et al. (2021a, 2021b) respectively. The KCC software from Qasmi and Ribes (2022a, 2022b) is freely accessible. All graphics have been produced using the R or NCL software freely accessible at https://cran.r-project.org/ and https://www.ncl.ucar.edu/ respectively.

Acknowledgments
Thanks are due to all CMIP6 participants and to Aurélien Ribes and Saïd Qasmi for the development of the KCC statistical package. I am also grateful to Gionata Ghiggi for the development of the GRUN runoff reconstruction, as well as to two anonymous reviewers for their useful comments on the original manuscript.

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., et al. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, *12*(7), 533–540. https://doi.org/10.1038/s41561-019-0374-y

Alkama, R., Marchand, L., Ribes, A., & Decharme, B. (2013). Detection of global runoff changes: Results from observations and CMIP5 experiments. *Hydrology and Earth System Sciences*, *17*(7), 2967–2979. https://doi.org/10.5194/hess-17-2967-201

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., et al. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, *1472*, 1–27. https://doi.org/10.1111/nyas.14337

Allan, R. P., & Douville, H. (2024). An even drier future for the arid lands. *Proceedings of the National Academy of Sciences*, *121*(2), e2320840121. https://doi.org/10.1073/pnas.2320840121

Beck, C., Grieser, J., & Rudolf, B. (2005). *A new monthly precipitation climatology for the global land areas for the period 1951 to 2000*. (published in Climate Status Report 2004 (pp. 181–190). German Weather Service.

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., et al. (2020). Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *Journal of Climate*, *33*(4), 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Chang, H., Johnson, G., Hinkley, T., & Jung, I.-W. (2014). Spatial analysis of annual runoff ratios and their variability across the contiguous U.S. *Journal of Hydrology*, *511*, 387–402. https://doi.org/10.1016/j.jhydrol.2014.01.066

Cui, T., Li, Y., Yang, L., Nan, Y., Li, K., Tudaji, M., et al. (2023). Non-monotonic changes in Asian Water Towers' streamflow at increasing warming levels. *Nature Communications*, *14*(1), 1176. https://doi.org/10.1038/s41467-023-36804-6

Dai, A., Qian, T., Trenberth, K. E., & Milliman, J. D. (2009). Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, *22*(10), 2773–2792. https://doi.org/10.1175/2008jcli2592.1

Douville, H., Allan, R. P., Arias, P. A., Betts, R. A., Caretta, M. A., Cherchi, A., et al. (2022). Water remains a blind spot in climate change policies. *PLOS Water*, *1*(12), e0000058. https://doi.org/10.1371/journal.pwat.0000058

Douville, H., & John, A. (2021). Fast adjustment versus slow SST-mediated response of daily precipitation statistics to abrupt 4xCO2. *Climate Dynamics*, *56*(3–4), 1083–1104. https://doi.org/10.1007/s00382-020-05522-w

Douville, H., Raghavan, K., Renwick, J., Allan, R. P., Arias, P. A., Barlow, M., et al. (2021). Water cycle changes. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *In climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 1055–1210). Cambridge University Press. https://doi.org/10.1017/9781009157896.010

Douville, H., Ribes, A., & Bock, O. (2022). Global warming at near-constant relative humidity further supported by recent in situ observations. *Communications Earth & Environment.*, *3*(1), 237. https://doi.org/10.1038/s43247-022-00561-z

Douville, H., & Willett, K. (2023). A drier than expected future, supported by near-surface relative humidity observations. *Science Advances*, *9*(30), eade6253. https://doi.org/10.1126/sciadv.ade6253

Dutot, E., & Douville, H. (2023). Revisiting the potential to narrow uncertainty in annual runoff projections of Arctic rivers. *Geophysical Research Letters*, *50*(16), e2023GL104039. https://doi.org/10.1029/2023GL104039

Elbaum, E., Garfinkel, C. I., Adam, O., Morin, E., Rostkier-Edelstein, D., & Dayan, U. (2022). Uncertainty in projected changes in precipitation minus evaporation: Dominant role of dynamic circulation changes and weak role for thermodynamic changes. *Geophysical Research Letters*, *49*(12), e2022GL097725. https://doi.org/10.1029/2022GL097725

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization [Dataset]. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). Grun: An observation-based global gridded runoff dataset from 1902 to 2014 [Dataset]. *Earth System Science Data*, *11*(4), 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Guo, H., Zhan, C., Ning, L., Li, Z., & Hu, S. (2022). Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the runoff. *Theoretical and Applied Climatology*, *149*(3–4), 1451–1470. https://doi.org/10.1007/s00704-022-04118-0

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., & Sutton, R. (2020). Observed emergence of the climate change signal: From the familiar to the unknown. *Geophysical Research Letters*, *47*(6), e2019GL086259. https://doi.org/10.1029/2019GL086259

Hegerl, G. C., Ballinger, A. P., Booth, B. B. B., Borchert, L. F., Brunner, L., Donat, M. G., et al. (2021). Toward consistent observational constraints in climate predictions and projections. *Frontiers in Climate*, *3*, 678109. https://doi.org/10.3389/fclim.2021.678109

Hobeichi, S., Abramowitz, G., Evans, J., & Beck, H. E. (2019). Linear optimal runoff aggregate (LORA): A global gridded synthesis runoff product (2019). *Hydrology and Earth System Sciences*, *23*(2), 851–870. https://doi.org/10.5194/hess-23-851-2019

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E., Brunner, L., et al. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamic Discuss*. https://doi.org/10.5194/esd-2019-93

Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., & Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change*, *9*(12), 926–933. https://doi.org/10.1038/s41558-019-0639-x

Mayer, J., Mayer, M., & Haimberger, L. (2021). Consistency and homogeneity of atmospheric energy, moisture, and mass budgets in ERA5. *Journal of Climate*, *34*(10), 3955–3974. https://doi.org/10.1175/jcli-d-20-0676.1

Miao, C., Wu, Y., Fan, X., & Su, J. (2023). Projections of global land runoff changes and their uncertainty characteristics during the 21st century. *Earth's Future*, *11*(4), e2022EF003286. https://doi.org/10.1029/2022EF003286

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021a). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021b). Copernicus climate change service (C3S) climate data store (CDS). [Dataset]. https://doi.org/10.24381/cds.68d2bb30

Oki, T., & Sud, Y. C. (1998). Design of total runoff integrating pathways (TRIP)—A global river channel network [Dataset]. *Earth Interactions*, *2*(1), 1–37. https://doi.org/10.1175/1087-3562(1998)002<0001:dotrip>2.3.co;2

Qasmi, S., & Ribes, A. (2022a). Reducing uncertainty in local climate projections. *Science Advances*, *8*, eabo6872. https://doi.org/10.21203/rs.3.rs-364943/v1

Qasmi, S., & Ribes, A. (2022b). Kriging for climate change [Software]. *GNU General Public License, version 3 (GPLv3)*. https://doi.org/10.5281/zenodo.5233947

Ribes, A., Qasmi, S., & Gillett, N. (2021). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), eabc0671. https://doi.org/10.1126/sciadv.abc0671

Robertson, F. R., Bosilovich, M. G., & Roberts, J. B. (2016). Reconciling land–ocean moisture transport variability in reanalyses with P-et in observationally driven land surface models. *Journal of Climate*, *29*(23), 8625–8646. https://doi.org/10.1175/jcli-d-16-0379.1

Scheff, J., Coats, S., & Laguë, M. M. (2022). Why do the global warming responses of land-surface models and climatic dryness metrics disagree? *Earth's Future*, *10*(8), e2022EF002814. https://doi.org/10.1029/2022ef002814

Schneider, U., Hänsel, S., Finger, P., Rustemeier, E., & Ziese, M. (2022). GPCC full data monthly product version 2022 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historical data [Dataset]. *Global Precipitation Climatology Centre*. https://doi.org/10.5676/DWD_GPCC/FD_M_V2022_050

Simpson, I. R., McKinnon, K. A., Kennedy, D., Lawrence, D. M., Lehner, F., & Seager, R. (2023). Observed humidity trends in dry regions contradict climate models. *PNAS*, *121*(1), e2302480120. https://doi.org/10.1073/pnas.2302480120

Tang, Q., & Lettenmaier, D. P. (2012). 21st century runoff sensitivities of major global river basins. *Geophysical Research Letters*, *39*(6), L06403. https://doi.org/10.1029/2011GL050834

Wang, A., Miao, Y., Kong, X., & Wu, H. (2022). Future changes in global runoff and runoff coefficient from CMIP6 multi-model simulation under SSP1-2.6 and SSP5-8.5 scenarios. *Earth's Future*, *10*(12), e2022EF002910. https://doi.org/10.1029/2022EF002910

Wu, Y., Miao, C., Slater, L., Fan, X., Chai, Y., & Sorooshian, S. (2024). Hydrological projections under CMIP5 and CMIP6. *Bulletin American Meteorological Society*, *105*(1), E59–E74. https://doi.org/10.1175/BAMS-D-23-0104.1

Yang, H., Zhou, F., Piao, S., Huang, M., Chen, A., Ciais, P., et al. (2017). Regional patterns of future runoff changes from Earth system models constrained by observation. *Geophysical Research Letters*, *44*(11), 5540–5549. https://doi.org/10.1002/2017GL073454

Yang, Y., Roderick, M. L., Guo, H., Miralles, D. G., Zhang, L., Fatichi, S., et al. (2023). Evapotranspiration on a greening Earth. *Nature Reviews Earth & Environment*, *4*(9), 626–641. https://doi.org/10.1038/s43017-023-00464-3

Yang, Y., Zhang, S., McVicar, T. R., Beck, H. E., Zhang, Y., & Liu, B. (2018). Disconnection between trends of atmospheric drying and continental runoff. *Water Resources Research*, *54*(7), 4700–4713. https://doi.org/10.1029/2018WR022593

Zhang, X., Tang, Q., Liu, X., Leng, G., & Di, C. (2018). Nonlinearity of runoff response to global mean temperature change over major global river basins. *Geophysical Research Letters*, *45*(12), 6109–6116. https://doi.org/10.1029/2018GL078646

Zhang, X., Tang, Q., Zhang, X., & Lettenmaier, D. P. (2014). Runoff sensitivity to global mean temperature change in the CMIP5 Models. *Geophysical Research Letters*, *41*(15), 5492–5498. https://doi.org/10.1002/2014GL060382