Does the HadGEM3-GC3.1 GCM overestimate land precipitation at high resolution? A constraint based on observed river discharge

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Müller, O. V., Vidale, P. L. ORCID: https://orcid.org/0000-0002-1800-8460, Vanniere, B. ORCID: https://orcid.org/0000-0001-8600-400X, Schiemann, R. ORCID: https://orcid.org/0000-0003-3095-9856 and McGuire, P. C. ORCID: https://orcid.org/0000-0001-6592-4966 (2021) Does the HadGEM3-GC3.1 GCM overestimate land precipitation at high resolution? A constraint based on observed river discharge. Journal of Hydrometeorology, 22 (8). pp. 2131-2151. ISSN 1525-755X doi: 10.1175/JHM-D-20-0290.1 Available at https://centaur.reading.ac.uk/99942/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/JHM-D-20-0290.1

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Does the HadGEM3-GC3.1 GCM Overestimate Land Precipitation at High Resolution? A Constraint Based on Observed River Discharge

OMAR V. MÜLLER,[a] PIER LUIGI VIDALE,[a] BENOÎT VANNIÈRE,[a] REINHARD SCHIEMANN,[a] AND PATRICK C. MCGUIRE[a]

[a] *Department of Meteorology, National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom*

ABSTRACT: Previous studies showed that high-resolution GCMs overestimate land precipitation when compared against observation-based data. Particularly, high-resolution HadGEM3-GC3.1 shows a significant precipitation increase in mountainous regions, where the scarcity of gauge stations increases the uncertainty of gridded observations and reanalyses. This work evaluates such precipitation uncertainties indirectly through the assessment of river discharge, considering that an increase of ~10% in land precipitation produces ~28% more runoff when the resolution is enhanced from 1° to 0.25°, and ~50% of the global runoff is produced in 27% of global land dominated by mountains. We diagnosed the river flow by routing the runoff generated by HadGEM3-GC3.1 low- and high-resolution simulations. The river flow is evaluated using a set of 344 monitored catchments distributed around the world. We also infer the global discharge by constraining the simulations with observations following a novel approach that implies bias correction in monitored rivers with two methods, and extension of the correction to the river mouth, and along the coast. Our global discharge estimate is $47.4 \pm 1.6 \times 10^3 \, \mathrm{km}^3 \, \mathrm{yr}^{-1}$, which is closer to the original high-resolution estimate ($50.5 \times 10^3 \, \mathrm{km}^3 \, \mathrm{yr}^{-1}$) than to the low-resolution ($39.6 \times 10^3 \, \mathrm{km}^3 \, \mathrm{yr}^{-1}$). The assessment suggests that high-resolution simulations perform better in mountainous regions, either because the better-defined orography favors the placement of precipitation in the correct catchment, leading to a more accurate distribution of runoff, or the orographic precipitation increases, reducing the dry runoff bias of coarse-resolution simulations. However, high-resolution slightly increases wet biases in catchments dominated by flat terrain. The improvement of model parameterizations and tuning may reduce the remaining errors in high-resolution simulations.

KEYWORDS: Rivers; Hydrologic cycle; Precipitation; Bias; General circulation models; Hydrologic models; Model evaluation/performance

1. Introduction

The hydrologic cycle is a closed system that describes the circulation of water between ocean, atmosphere, and land. The water budget states that incoming and outgoing water from a region is in balance with the change in water storage. Although simple in concept, water budgets are difficult to accurately determine in GCMs, given the considerable uncertainty in its components (Healy et al. 2007). One of the most challenging components is precipitation. Demory et al. (2014) and Roberts et al. (2018) found that while total global precipitation is remarkably resolution invariant and is in the range of observational uncertainty, its partitioning between ocean and land is strongly sensitive to GCMs resolution when it is enhanced from the conventional 1° to the 0.25° or finer resolution used nowadays. Similarly, Demory et al. (2014) reported an intensification of the hydrologic cycle in HadGEM at high resolution,

given by an increase in atmospheric moisture transport from ocean to land, which favors the occurrence of precipitation, and consequently, an increase of river discharge into the ocean. In a multimodel study, Vannière et al. (2019) found that the land precipitation is on average 10% larger at higher resolution, which is mostly explained by the enhancement of orographic precipitation.

Land precipitation increases in high-resolution models due to the emergence of resolved mesoscale processes, such as sharper atmospheric fronts, stronger temperature gradients, better defined orographic jets and their associated moisture transport, and even mesoscale convective systems (Vellinga et al. 2016). One would expect that the emergence of finer-scale processes leads to an improved representation of land precipitation. However, in the context of climate means, Roberts et al. (2018) and Vannière et al. (2019) have shown that the observations are in better agreement with low-resolution models than with high-resolution models, which produce larger land precipitation than most state-of-the-art observation-based products [e.g., GPCP in Roberts et al. (2018, their Fig. 1) and ERA-Interim in Vannière et al. (2019, their Fig. 9b)]. In this context, and as a follow-on of the previous studies, the question that motivates this paper naturally arises: are high-resolution models overly sensitive to orography, or do observations underestimate the amount of land precipitation, and in particular orographic precipitation?

On the one hand, orographic precipitation might be overly sensitive to model resolution. The extra precipitation at high
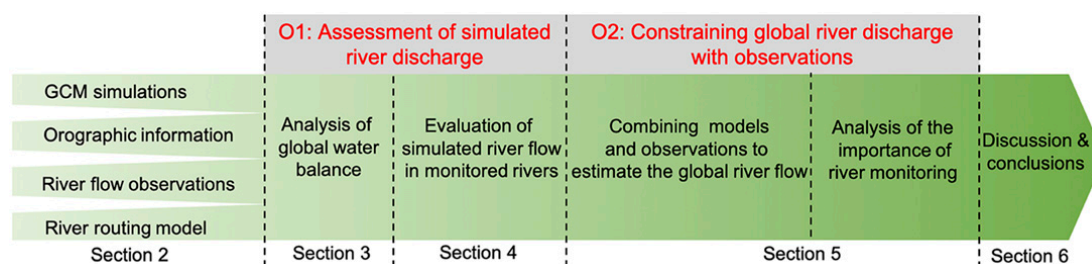
FIG. 1. Paper roadmap.

resolution was shown to be in balance with increased moisture transport from ocean to land and surface runoff, rather than with evapotranspiration which is mostly insensitive to resolution (Demory et al. 2014; Vannière et al. 2019). However, the atmospheric moisture transport cannot be measured, and thereby, can be neither used to evaluate GCMs nor to reduce uncertainties in the land water budget of reanalyses. Despite this, it was suggested that the moisture transport from ocean to land could be too high in high-resolution GCMs because of an excessive surface latent heat flux over the ocean, caused by too strong surface wind (Vannière et al. 2019).

On the other hand, land precipitation estimates based on gridded observations, reanalysis and satellite products are not free of uncertainty and errors. Adam et al. (2006) reported an underestimation of ~6% in gridded land precipitation observations, mainly given by ungauged topographically complex areas. Large errors and low correlations with the scarce observations available in mountainous regions are also shown in Harris et al. (2020) for CRU TS v4.03, a gridded observation product. Beck et al. (2020) demonstrated that many of the widely used precipitation datasets (GPCC V2015, GPCP V2.3, and MERRA-2) severely underestimate precipitation in some areas of complex orography. Beck et al. (2017) reported poor correlation between reanalysis and in situ observations in areas sensitive to subgrid processes (e.g., convection), due to the relatively coarse resolution of reanalyses, but also showed poor agreement between high-resolution remotely sensed data and ground measurements at high latitudes and altitudes, due to the known limitations of satellites in snow covered areas. In addition, Stephens et al. (2012) suggested that global precipitation is at least 10% larger than previously estimated in the light of new data from *CloudSat* satellites, and when snowfall is accounted for.

As both observed and simulated orographic precipitation are uncertain, we seek to evaluate it indirectly against another component of the land water budget, river discharge, which is directly affected by any increase of orographic precipitation and for which we possess reliable observations. Choosing river discharge for this purpose presents several advantages: (i) its measurements integrate the water balance in the whole catchment into a single time series, while precipitation observations depend on the density of gauge stations (probably scarce over mountains); (ii) rivers are strongly sensitive to changes in orographic precipitation; and (iii) river discharge compensates the atmospheric moisture transport from ocean to land. One limitation, however, is that roughly 50% of the global land is covered by nonmonitored catchments (Fekete

and Vörösmarty 2007). Moreover, the number of gauge stations is gradually declining since the 1980s (Dai et al. 2009), and satellites still do not capture narrow river channels (Pavelsky et al. 2014). Thus, the use of model simulations, starting from reanalyses and going to free-running GCMs, is inescapable in order to obtain global estimates of river discharge. Inferences based on available observations of river flow, or based on differences between precipitation and evapotranspiration, were the most common global estimates of river discharge in the past (e.g., Lvovitch 1973; Chahine 1992; Oki et al. 1995; Grabs et al. 1996; Perry et al. 1996; Dettinger and Diaz 2000; Syed et al. 2009). More recent estimates involve direct model estimations (e.g., Oki et al. 2001; Nijssen et al. 2001; McCabe and Wolock 2011; Munier et al. 2012), and the combinations of model with observations (e.g., Dai and Trenberth 2002; Fekete et al. 2002; Clark et al. 2015; Ghiggi et al. 2019; Lin et al. 2019; Harrigan et al. 2020). Our approach is to constrain model simulations with observations using geographical and orographic information of catchments to fill gaps in ungauged areas.

To answer the question raised above, the objective of this work is twofold. The first objective is to evaluate climatological precipitation biases at the catchment scale, in the low- and high-resolution (LR and HR) configurations of a state-of-the-art GCM, HadGEM3-GC3.1, against monitored river discharge. To make this evaluation possible, we calculate the GCM river discharge with a river routing model, forced by the long-term model runoff. The second objective is to estimate the global river discharge based on a bias correction of the GCM river flow. Observations will be used to correct the biases of the GCM river flow in monitored rivers, but also to infer a value of discharge in ungauged rivers using a novel methodology that considers observations and catchments' orography. Two bias correction methods (linear scaling and CDF mapping) are applied to each simulation to give robustness to the proposed methodology. By doing this research, we expect to identify the potential and limitations of high-resolution simulations, to suggest where, should there be a conscious investment in additional river discharge observations, this might help the climate modeling community to constrain even better the GCMs results, and to derive accurate estimates of the water budget.

Figure 1 summarizes the roadmap followed in this article to achieve the stated objectives. Section 2 describes the models' simulations, the observations, the orographic information, and the river routing model. Next, the paper has two main parts directly linked to the two objectives. The first part addresses

TABLE 1. GCM configurations.

Ensemble	Simulation	Simulation type	Atmosphere horizontal resolution at 50°N	Ocean resolution	Time step
LR	LRA	AMIP	N96 (~135 km)	—	20 min
	LRC	COUPLED	N96 (~135 km)	ORCA1	20 min
HR	HRA	AMIP	N512 (~25 km)	—	10 min
	HRC	COUPLED	N512 (~25 km)	ORCA025	10 min

the assessment of simulated river flow in monitored rivers (O1) and includes sections 3 and 4. Section 3 explores the role of resolution in the simulation of the hydrologic cycle, while section 4 evaluates the model's capability to simulate the river flow in monitored rivers. The second part uses the lessons learnt from the first part to optimally combine model simulations and observations to constraint the global river discharge (O2). This part is detailed in section 5, where we describe the methodology to estimate the global discharge, analyze the results, and discuss the importance of river flow observations. Finally, section 6 presents the overall discussion and concluding remarks.

2. Models and observations

a. GCM simulations and river routing model

Climate model simulations produced with the Hadley Centre Global Environment Model version 3 (HadGEM3-GC3.1, Williams et al. 2018), within the framework of the High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6 (Haarsma et al. 2016), are used to evaluate the water budget, and to force a river routing model. A full description of HadGEM3-GC3.1 simulations for CMIP6-HighResMIP is found in Roberts et al. (2019). The HadGEM3 family of models comprises a range of specific model configurations incorporating different levels of complexity but with a common physical framework. The subversion GC3.1 uses the Unified Model (UM; Cullen 1991) with the Global-Atmosphere 7.1 configuration with 85 vertical levels. The ocean is simulated with the Nucleus for European Modeling of the Ocean (NEMO; Madec and NEMO Team 2016) configuration Global-Ocean 6.0. The land surface

model is the Joint U.K. Land Environment Simulator (JULES; Best et al. 2011) with the Global-Land 7.1 configuration. JULES represents the soil column with four layers up to 3.0-m depth. A parameterization based on the Topography-Based Hydrological Model (TOPMODEL) is used to represent an extra layer beneath the soil column that simulates groundwater fluxes. Using this configuration, total runoff is the sum of saturation-excess (or surface) runoff and groundwater (or subsurface) runoff.

The GCM simulations are organized in two ensembles: low- and high-resolution (LR and HR, respectively; see Table 1). Each ensemble is composed of two simulations: atmosphere–land (AMIP) and ocean–atmosphere–land (COUPLED). The simulations are named by three letters, indicating resolution and simulation type: low- and high-resolution AMIP (LRA and HRA, respectively), low- and high-resolution COUPLED (LRC and HRC, respectively). The goal of including AMIP and COUPLED simulations is to provide members to the ensembles. Two members per ensemble is little, but the maximum possible for CMIP6-HighResMIP models. It is known that AMIP and COUPLED present different atmospheric dynamics, due to their differences in sea surface temperatures (observed versus simulated), that may alter the water budget components. For this reason, the methodology followed in this paper (analysis of water budget, river routing, bias correction, etc.) is done for each individual simulation; however, some results are presented as ensemble mean, given that the GCM is more sensitive to resolution than to the use of coupling with the ocean.

The GCM's configuration (e.g., vertical levels, parameterizations, schemes) remains identical in both resolutions, except by the necessary reduction in time step from 20 min at LR to 10 min at HR. LR uses a regular latitude/longitude grid at N96
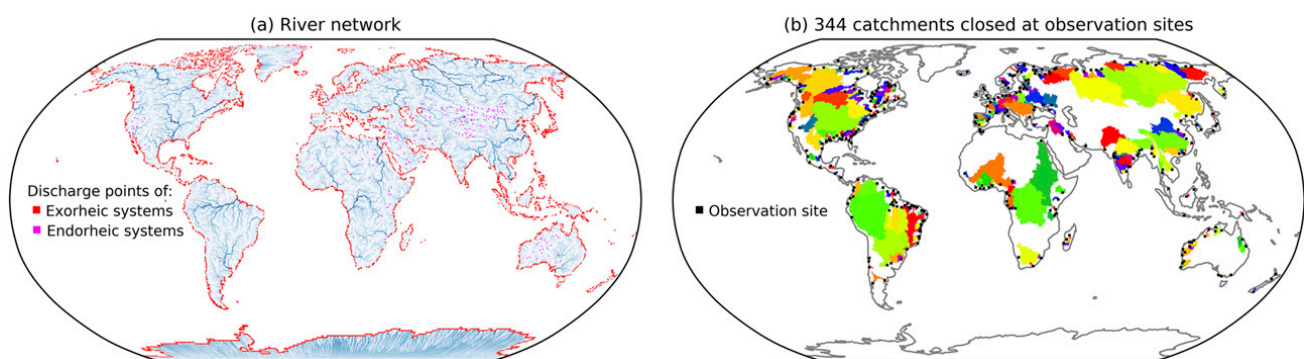

FIG. 2. (a) TRIP river network in blue (given by the sequence parameter; see its definition in section 2a), including discharge points into oceans/seas (exorheic systems) and lakes (endorheic systems). (b) Monitored rivers with black dots indicating the observation sites for river flow and colors highlighting the catchment area that contributes to each monitored point according to the TRIP river network.
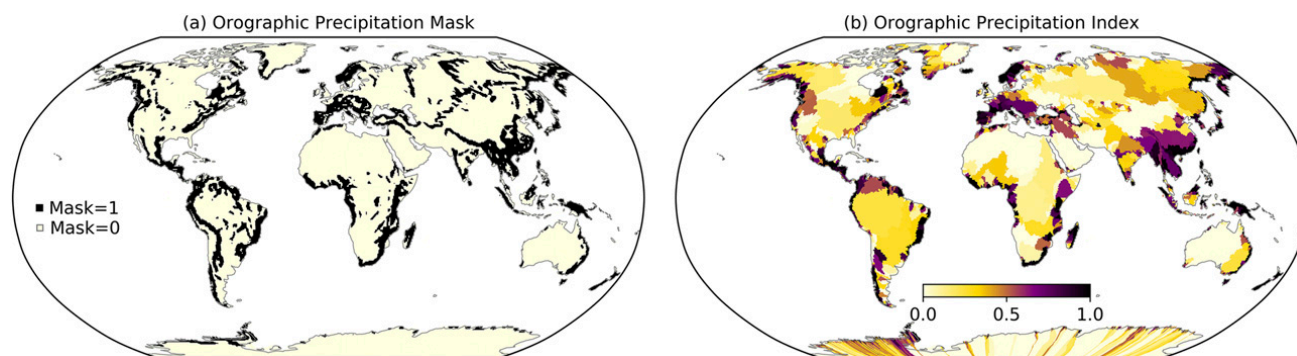
FIG. 3. (a) Binary mask, highlighting regions influenced by orographic precipitation in black, which covers 27% of global land. (b) Orographic precipitation index in TRIP catchments (1 means the catchment is dominated by orographic precipitation, 0 the contrary).

(~135 km at 50°N) in the atmosphere and 1° in the ocean for the COUPLED configuration, while HR uses N512 (~25 km at 50°N) in the atmosphere and 0.25° in the ocean component. AMIP simulations are forced by HadISST2 daily SST at 0.25° (Titchner and Rayner 2014). The period of simulation is 1950–2014. Further details are in Roberts et al. (2019).

The total (surface and subsurface) runoff produced by each GCM simulation in each land grid cell is used to force a river routing model that collects the runoff to estimate the river discharge. The model is a stand-alone implementation of the Total Runoff Integrating Pathways model (TRIP; Oki and Sud 1998). TRIP uses a simple advection method to route total runoff along a prescribed river network (Falloon et al. 2011). The advection method is explained in detail in the appendix. The river network (see Fig. 2a) is defined at 0.5° resolution based on two parameters set at each grid cell: direction and sequence. Direction indicates the flow direction when in the range [1 ... 8], or that the grid cell is a discharge point, 9 for exorheic rivers that discharge into ocean/sea, or 12 to rivers discharging into lakes. Sequence is an integer label that indicates the hierarchy of the grid cell in the river catchment, being 0 for grid cells that do not receive input from neighbors and maximum (in the catchment) at river mouth. To solve the advection/routing equations the model also requires the definition of two global fixed parameters: meandering ratio and flow velocity (further details are in the appendix). The meandering ratio m is set to 1.4 (dimensionless) while the flow velocity v is set to $0.4\,\mathrm{m\,s^{-1}}$ following the range of values used in Oki and Sud (1998), Oki et al. (1999), and Falloon et al. (2011). Note that 1) the river routing model runs at 0.5°, which implies that runoff fields are interpolated to that resolution and 2) TRIP does not consider any loss or gain of water, it just translates runoff values into river discharge allowing a comparison against observations, and thereby, a validation of runoff at catchment scale.

b. Observations

A selection of 344 near-coast river discharge observations from Dai (2017) is used to assess the river flow. The original dataset contains time series of monthly river flow for the world's largest 925 rivers collected from different sources (see Dai et al. 2009 for more details) in the period 1900–2014. The selection criteria focused on data availability, a minimum size of the catchment, and an agreement between

catchment area in model and observations. Specifically, observation sites that do not meet any of the following rules are excluded:

- Minimum of 120 observed values in the period 1950–2014.
- Minimum catchment area of 3 pixels in the river routing model.
- Minimum of 65% agreement in catchment area between the area contributing to the river discharge point in the model and the area reported in the observational dataset.
- No other gauge station located in the same model grid cell.

Figure 2b presents the selected gauge stations and the catchment areas based on the river network of our model (Fig. 2a), which cover 41.7% of the global land. The mean total observed discharge from selected rivers is $21.4 \times 10^3\,\mathrm{km^3\,yr^{-1}}$, while $1.65 \times 10^3\,\mathrm{km^3\,yr^{-1}}$ is the flow contributed by the 584 excluded observations.

c. Orographic precipitation

One of the mechanisms that is most sensitive to GCM's resolution is the enhancement of orographic lifting due to a better definition of the orography in high-resolution models. To analyze the sensitivity of the simulated global land water budget to orographic precipitation we define a global orographic precipitation mask and an orographic precipitation index per catchment. For an objective comparison across simulations, we calculated both, the mask and the index, based on ERA-Interim data (Dee et al. 2011), i.e., a reanalysis product independent from the GCM simulations.

The orographic precipitation mask is a time-invariant binary mask that separates regions with strong influence of orographic precipitation from regions dominated by nonorographic precipitation. It is calculated based on the orographic enhancement data taken from Vannière et al. (2019), who applied the diagnostic model developed by Sinclair (1994) to daily ERA-Interim data in the period 1980–2019. In the Sinclair (1994) model, the vertical orographic uplift at the surface is determined as a function of the dot product of horizontal surface wind and orographic slope. The vertical wind profile is a function of the orographic uplift and set to decrease vertically up to 200 hPa following a power law of atmospheric pressure. The condensation rate, which is function of the vertical wind, specific humidity, and relative humidity at each level, is
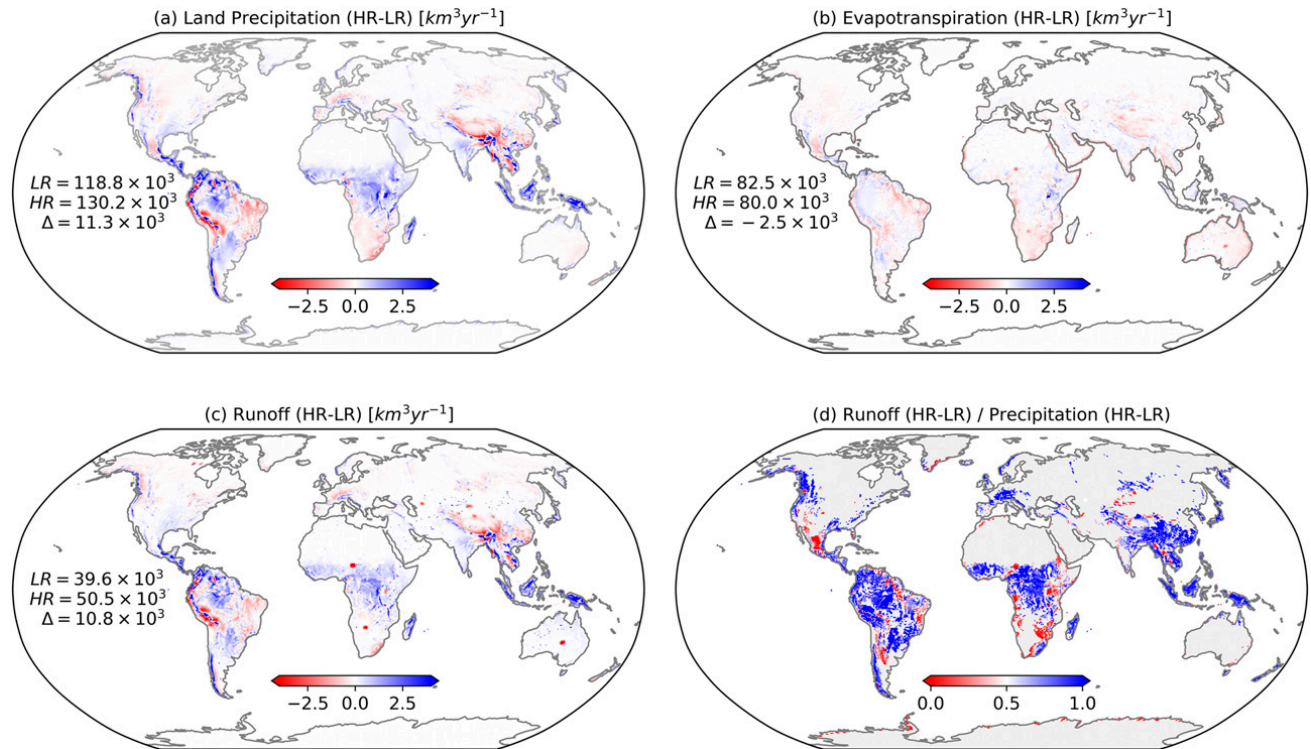
FIG. 4. Spatial difference of mean climatological water budget components between low and high resolution: (a) land precipitation, (b) evapotranspiration, and (c) runoff. The mean global value for each resolution and their difference is added as inset. Note that (a)–(c) use the same color bar allowing a direct comparison of changes in precipitation, evapotranspiration, and runoff. (d) Runoff changes between low and high resolution as fraction of land precipitation changes, which is equivalent to the fraction between (c) and (a). Gray shades indicate grid cells that are masked out because of the little changes in precipitation ($|\Delta P| < 0.4 \, \text{km}^3 \, \text{yr}^{-1}$) that may affect the diagnostic interpretation.

integrated vertically from the lifting condensation level to the top of the atmosphere, to give an estimate of orographic precipitation [i.e., R in the Eq. (4) of Sinclair 1994]. We made the further approximation that the large-scale vertical wind was zero, to retain only the orographic enhancement. Once the orographic precipitation is estimated, we calculate the time-average field, and apply a threshold of $0.3 \, \text{mm day}^{-1}$ to obtain the orographic precipitation mask. The mask, presented in Fig. 3a, clearly segments areas of high-altitude, but it differs from masks defined only based on terrain height or terrain slope (for instance the mask used in Adam et al. 2006). The main difference is that the model we used considers the horizontal wind to estimate the vertical orographic uplift, which retains the windward side of mountain chains where precipitation occurs and excludes leeward sides and dry high-altitude areas (Vannière et al. 2019). This is evident in Southeast Asia, where the mask identifies the southwest side of the Himalayas, but it does not consider the arid Tibetan Plateau. Therefore, the method yields a very accurate mask that distinguishes in detail areas prone to orographic precipitation from others.

As river catchments usually cover areas with and without orographic precipitation we define the orographic precipitation index (OPI) as the fraction of total precipitation that falls over the orographic part of the catchment c:

$$\text{OPI}_c = \frac{\sum\limits_{\forall (i,j) \in c} \overline{P}(i,j) \overline{M}(i,j)}{\sum\limits_{\forall (i,j) \in c} \overline{P}(i,j)}, \qquad (1)$$

where $\overline{P}$ is the time mean precipitation of ERA-Interim and $\overline{M}$ is the orographic precipitation mask presented in Fig. 3a. The index varies in the range [0, 1], where 0 indicates null contribution of orographic precipitation and 1 indicates the contrary. Figure 3b shows that catchments with high index values are in general small (e.g., those located in the Maritime Continent, Southeast Asia, Europe, or the Pacific coast of the Americas), while small index values ($\text{OPI}_c < 0.5$) dominate in large basins (e.g., Amazon, Congo, Yenisey, Lena) suggesting these catchments have a weak influence of orographic precipitation.

3. Global water budget and GCM resolution

The land water budget can be expressed as

$$\frac{dS}{dt} = P - E - R, \qquad (2)$$

where dS/dt is the change of water storage (soil moisture and groundwater), P is the precipitation by rain and snow, E is the
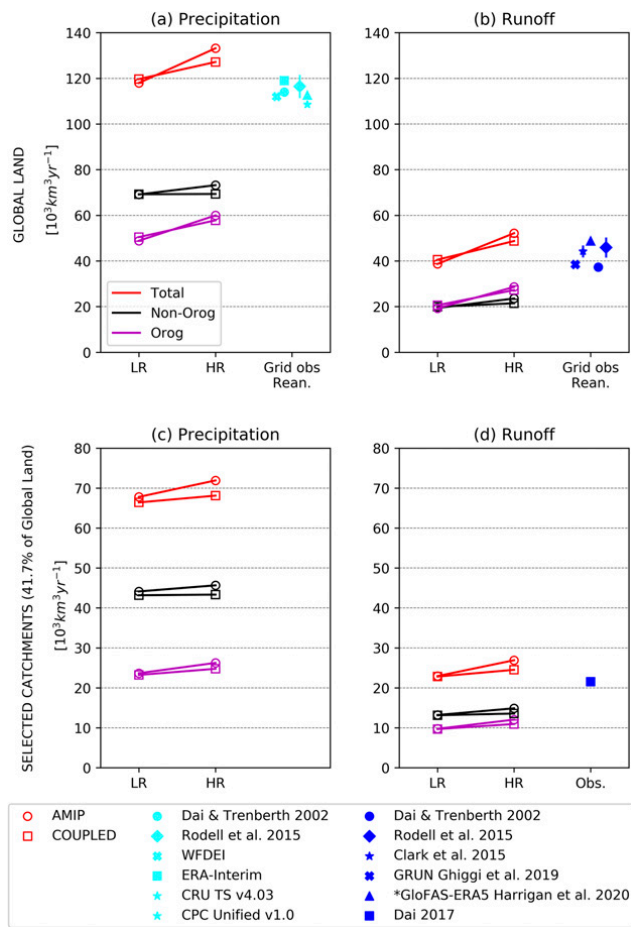
FIG. 5. Low- and high-resolution mean climatological (a),(c) precipitation and (b),(d) runoff (top) for the global land and (bottom) for selected catchments presented in Fig. 2b. The simulated variables are shown in total values, and their decomposition into orographic and nonorographic according to the mask presented in Fig. 3a. In addition, estimates based on gridded observations, reanalysis or optimized combinations between models and observations are included in cyan for precipitation and in blue for runoff. Note that the GloFAS-ERA5 value (* in the legend) is based on the period (1997–2014). Previous years were discarded given that ERA5 presents a significant imbalance in the global water budget (see section 9.2 in Hersbach et al. 2020), producing high precipitation, which in turn produces unrealistic high values of runoff.

evapotranspiration, and R is the runoff (surface and subsurface). Averaged over a long period of time, the storage term is negligible and Eq. (2) becomes

$$\overline{P} = \overline{E} + \overline{R}, \qquad (3)$$

or

$$\overline{P} = \overline{E} + \overline{Q}, \qquad (4)$$

where $\overline{P}$, $\overline{E}$, $\overline{R}$, and $\overline{Q}$ are climatological mean precipitation, evapotranspiration, runoff, and river discharge, respectively. Equations (3) and (4) are valid for individual catchments, and thereby, for the global land. Note that $\overline{R}$ and $\overline{Q}$ are conceptually equivalent, given that rivers integrate the catchment runoff

(subsurface runoff feeds the river baseflow, while surface runoff feeds the quickflow). Although there are human interventions that may alter this equivalence in the real world (e.g., evaporation from dam reservoirs), they are insignificant at the large scale, and are not represented in most climate and hydrological models.

Figure 4 shows that the enhancement of resolution affects differently each component of the climatological water budget. Precipitation (Fig. 4a) presents the largest differences in areas of high orography (e.g., the Andes and the Himalayas) and islands, in particular insular regions located between the tropics (e.g., Maritime Continent, Madagascar). At the global scale, the simulated precipitation is greater at HR, but there is a combination of both positive and negative differences in space. The differences are mainly due to the better definition of the orography, which alters the moisture transport, and thereby, precipitation patterns and timing. It explains changes over high orography, but also in flat areas sensitive to changes in atmospheric circulation such as the Sahel (Müller et al. 2021), central Africa, and La Plata basin. Focusing on high orography areas, there is an increase of precipitation in the windward side at HR, accompanied with drying in rain shadows (e.g., Mexican and Tibetan Plateaus).

Evapotranspiration presents a spatial sensitivity to resolution similar to precipitation, but remarkably weaker in magnitude (see Fig. 4b). Instead, the runoff field (Fig. 4c) shows a notable resemblance to precipitation, both in magnitude and in spatial distribution. The prevalence of precipitation impact on runoff over evapotranspiration is also evidenced in Fig. 4d. It presents the changes in runoff due to resolution as fraction of precipitation changes. A fractional value of one means that the change in precipitation affects only runoff, zero means the precipitation change impacts only evapotranspiration, 0.5 means both runoff and evapotranspiration are affected equally. The predominance of blue colors (fraction > 0.5) confirms the strong link between precipitation and runoff differences, revealing that the extra precipitation at HR mostly ends up in rivers. In summary, we found a strong effect of resolution on global land precipitation and runoff, and conversely, a weak impact on evapotranspiration (in agreement with Vannière et al. 2019). Land precipitation is on average ~10% higher at HR but produces ~28% more runoff. All these results support the premise that the assessment of river discharge offers a clear insight about precipitation amount biases at the catchment scale.

Figures 5a and 5b summarize the sensitivity of global land precipitation and runoff to resolution for AMIP and COUPLED simulations, quantifying the contribution of orographic and nonorographic regions to such sensitivity. Both AMIP and COUPLED present a clear increase of precipitation and runoff at HR, although AMIP shows larger differences. It is remarkable that between 70% and 98% of the net increase in precipitation and runoff at high resolution is given by increases in regions prone to orographic enhancement. On the other hand, regions of flat land present little sensitivity to resolution. Beyond the effect of resolution, it is also interesting to note that at
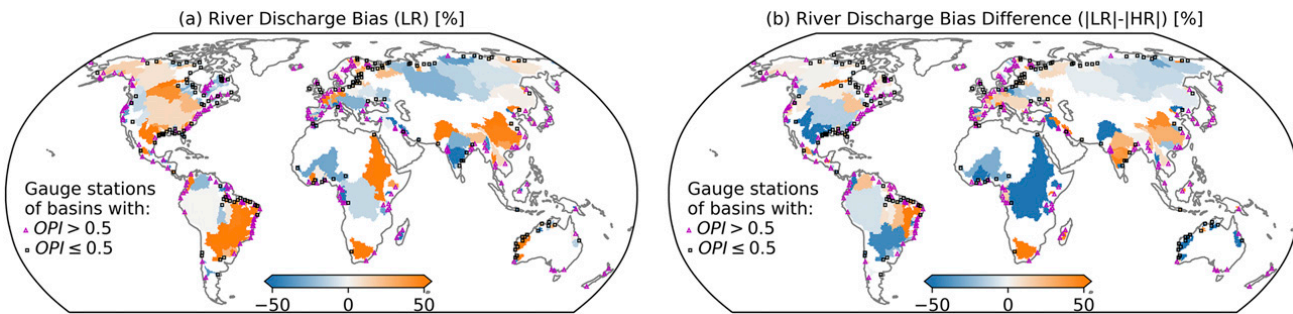
FIG. 6. Mean climatological river discharge percent bias for selected catchments: (a) LR ensemble and (b) LR minus HR ensemble (orange indicates HR reduces the bias, blue the contrary). The marks show the location of gauge station, black squares for the 156 catchments dominated by nonorographic precipitation (OPI ≤ 0.5), pink triangles for the 188 catchments with strong influence of orographic precipitation (OPI > 0.5). Antarctica is excluded due to the lack of observations.

least 40% of global land precipitation and 50% of the global runoff in HadGEM3-GC3.1 is produced in just the 27% of global land area characterized by orographic enhancement, which highlights the importance of reducing the uncertainty in such regions. Comparing against observation-based datasets (e.g., gridded observations and reanalysis) the net precipitation increase in high-resolution simulations magnifies the wet biases, mainly in areas sensitive to orographic precipitation (not shown). In particular, observation-based land precipitation estimates are in the range [108–120] × 10^3 km^3 yr^{-1}, close to the low-resolution value (119 × 10^3 km^3 yr^{-1}), but below the high-resolution estimates (130 × 10^3 km^3 yr^{-1}). Similarly, global runoff or river discharge based on observations is in the range [37–49] × 10^3 km^3 yr^{-1} in agreement with low-resolution (39.6 × 10^3 km^3 yr^{-1}), but smaller than high-resolution estimates (50.5 × 10^3 km^3 yr^{-1}). On the basis of these results, and given the uncertainty of observational products discussed in section 2, we next assess the simulated river discharge at the catchment scale, using it as a proxy to understand precipitation biases.

4. River discharge assessment in monitored rivers

As our first goal is to understand climatological biases of total precipitation as integral quantities, through the assessment of river flow, the evaluation focuses on mean absolute and percent errors. By doing so, some considerations should be taken into account. First, a correct river discharge value in large catchments can be the result of compensating biases, thereby, inferences about spatial biases can only be done in

regions with small catchments, but not in areas dominated by large catchments. Second, the temporal variability of precipitation cannot be assessed using river discharge, given the lag between precipitation events and river discharge. Third, monitored catchments are not fully representative of what is observed globally. Figures 5c and 5d show the decomposition of land precipitation and runoff into their orographic and nonorographic components for the area covered by selected rivers. The figure shows that the land covered by gauged rivers is less sensitive to resolution (and orographic enhancement) than the global land. Precipitation in monitored basins is on average ~4% higher at HR, and produces ~12% more runoff, while global land precipitation is ~10% higher at HR, and produces ~28% more runoff. This is because of 1) many of the gauged rivers correspond to large basins dominated by nonorographic precipitation such as Amazon, La Plata, Mississippi, and Congo, all of which have an orographic index below 0.4, and 2) the impossibility of monitoring the large number of small mountain rivers. Nevertheless, there are data from a number of monitored rivers in key regions such as Southeast Asia, the Alps, and the Andes that can help to constrain the precipitation uncertainty in such regions.

Although there is a good agreement between LR and observations in Fig. 5d (22.9 × 10^3 and 21.5 × 10^3 km^3 yr^{-1}, respectively), this is the result of compensating positive and negative biases (see Fig. 6a). Wet biases in North America, southeast South America, and China are combined with dry biases in Siberia and Central and West Africa. Figure 6b shows in orange that HR is able to reduce biases in small catchments with strong influence of orographic precipitation such as

TABLE 2. Number of nonorographic and orographic rivers that perform better at low or high resolution.

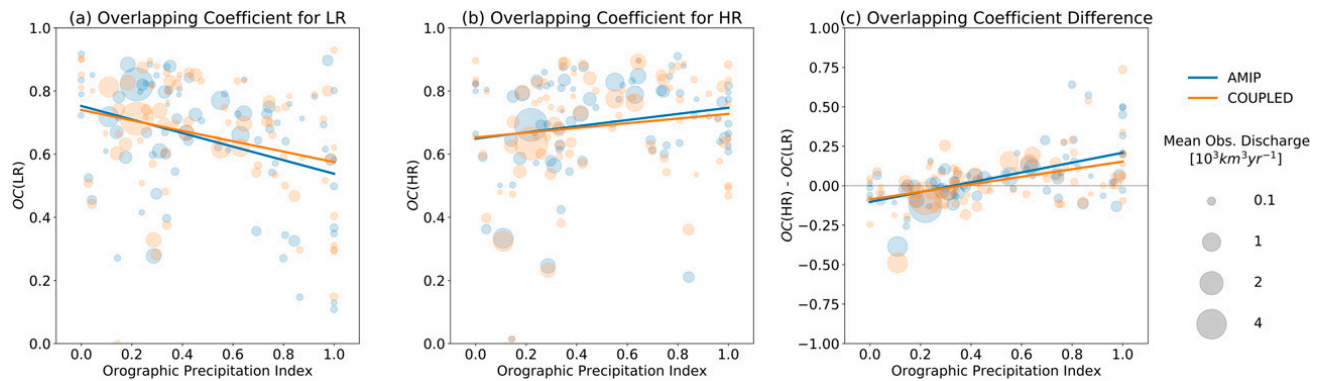|	Nonorographic (OPI ≤ 0.5)	Orographic (OPI > 0.5)	Total
No. LR performs better	85	67	152
No. HR performs better	71	121	192
Total	156	188	344
HR performs better	46%	64%	56%
Obs discharge (10^3 km^3 yr^{-1})	14.6	6.9	21.5

FIG. 7. Performance of simulated river flow measured by the overlapping coefficient as function of the orographic precipitation index of river catchments. (a) Low-resolution simulations: LRA (blue) and LRC (orange), (b) high-resolution simulations: HRA (blue) and HRC (orange), and (c) the difference between low- and high-resolution overlapping coefficients. Rivers with insignificant contribution ($Q <$ 0.03×10^3 km^3 yr^{-1}) were excluded. The circles' size is proportional to the mean observed discharge of each river. The lines plot the unweighted linear regressions.

Southeast Asia and Maritime Continent, Europe, South Africa, the Andes, the northwest Rockies and highlands in Canada, Alaska, and southeast Brazil, but it increases the biases in large catchments such as Amazon, Congo, Paraná, and Mississippi. Table 2 quantifies the number of rivers, with and without influence of orographic precipitation, which performs better at each resolution. The results indicate that HR is closer to observations in 56% of rivers, and this number rises to 64% in mountain rivers. However, monitored rivers with high orographic index only represents 32% of the total (orographic + nonorographic) observed discharge, a much smaller fraction than that estimated at the global scale, which is at least 50%.

Figure 7 complements the previous results with a normalized skill score (the overlapping coefficient OC) as a function of the catchments' orographic precipitation index. Given the probability distribution of simulated river flow and the probability distribution of observed river flow, the overlapping coefficient measures the overlapping area below the curves of the two distributions (Weitzman 1970). The discrete version can be expressed as

$$OC = \sum_{\forall i} \min\left(\frac{nmod_i}{N}, \frac{nobs_i}{N}\right), \qquad (5)$$

where $nmod_i = \text{count}(q \in i\text{th bin})$ is the histogram of the monthly model time series of length N, while $nobs_i$ is the same for the observed data. This skill score has the advantage of being normalized in the range [0–1], being zero when there is no overlap, and one when the histograms are identical. Figures 7a and 7b show OC > 0.5 for most rivers in both resolutions. However, the plots evidence that the performance of LR decrease for rivers dominated by orographic precipitation, while the score in such rivers notably improves in HR. Figure 7c presents the score difference between high and low resolution. Circles above zero in the ordinate indicate a better score for HR, while circles below zero mean the contrary. The plot suggests that the enhancement of resolution leads to a general improvement of the river flow simulation, which is

particularly evident when the topographic complexity of the catchments increases. On the other hand, catchments with low orographic precipitation index (OPI < 0.4) present, in general, slightly better scores at low resolution. The exception is Congo [second in size circle with OC(HR) − OC(LR) ≈ −0.49], where the extra precipitation at high resolution (Fig. 4a) overcompensates, in excess, for the negative bias shown at low resolution (see Fig. 6a).

Given these results, we focus the assessment on six regions prone to orographic precipitation (Fig. 8), and thereby, highly sensitive to model resolution: Southeast Asia (Figs. 8a,b), the Maritime Continent (Fig. 8c), the southern Andes (Fig. 8d), the Alps (Fig. 8e), and the Alaska Range (Fig. 8f). From this regional zoom-in, we find three deficiencies of LR that are clearly improved at HR: 1) compensation of errors, 2) high underestimation of runoff, and 3) poor spatial distribution of runoff.

1) If we aggregate catchments ahead and behind the Himalayas (Figs. 8a and 8b together) and evaluate the total river flow, LR is closer to observations than HR for the wrong reason. At LR, there is a compensation of negative biases in the rivers of India and Pakistan (ahead of the Himalayas), with positive biases in rivers located east of the Himalayas. The better representation of the orographic barrier and the Plateau in HR simulations likely improves the simulation of atmospheric monsoon circulation and moisture fluxes [as found in Schiemann et al. (2018) and Vannière et al. (2019)], producing more precipitation and runoff in India and Pakistan, and reducing rain, snow, and runoff behind the Himalayas. It notably reduces the negative biases at LR in Indian rivers (Fig. 8a), and slightly improves the performance in rivers originating in China (Fig. 8b).

2) Figure 8c presents the evaluation for rivers in the Maritime Continent, one of the regions with the greatest discrepancy between low- and high-resolution simulations. Although there are only five monitored rivers in the region, they are enough to demonstrate that the high precipitation amount at HR is much more realistic than that estimated at LR. It has been shown for earlier versions of HadGEM that the
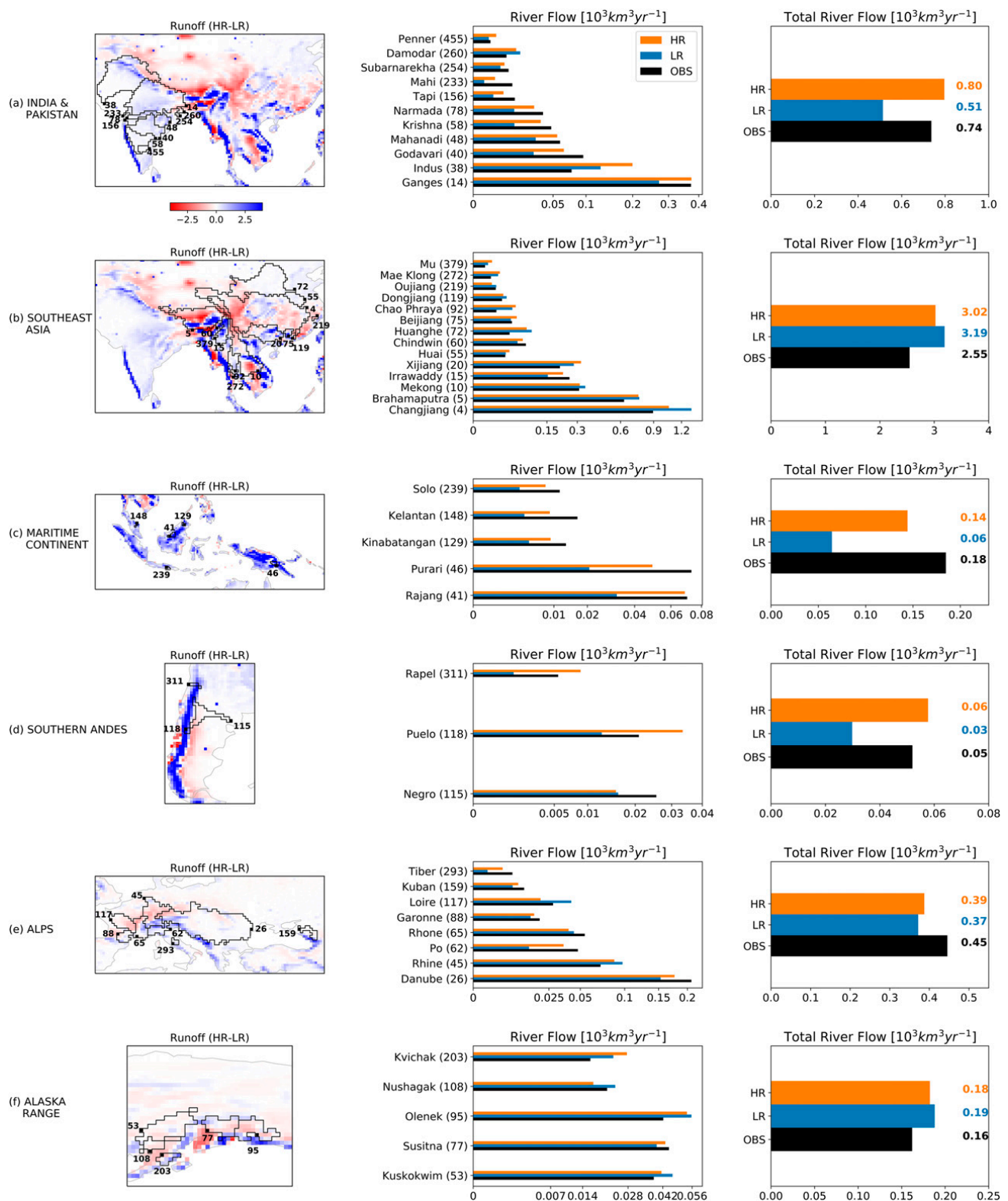
FIG. 8. (center) Evaluation of low- and high-resolution HadGEM3-GC3.1 estimates of mean river flow vs Dai (2017) observations for six regions where orographic precipitation, and thereby, (left) runoff is strongly sensitive to resolution. (right) As in the center column, but for accumulating river flow in each region. The regions are (a) India and Pakistan, (b) Southeast Asia, (c) Maritime Continent, (d) southern Andes, (e) Alps, and (f) Alaska Range. Note that the bar plots in the center column are shown in square root scale to make the small rivers visible. The rivers IDs follow the IDs defined in the observational dataset. The legends in the top panels are valid for all regions.
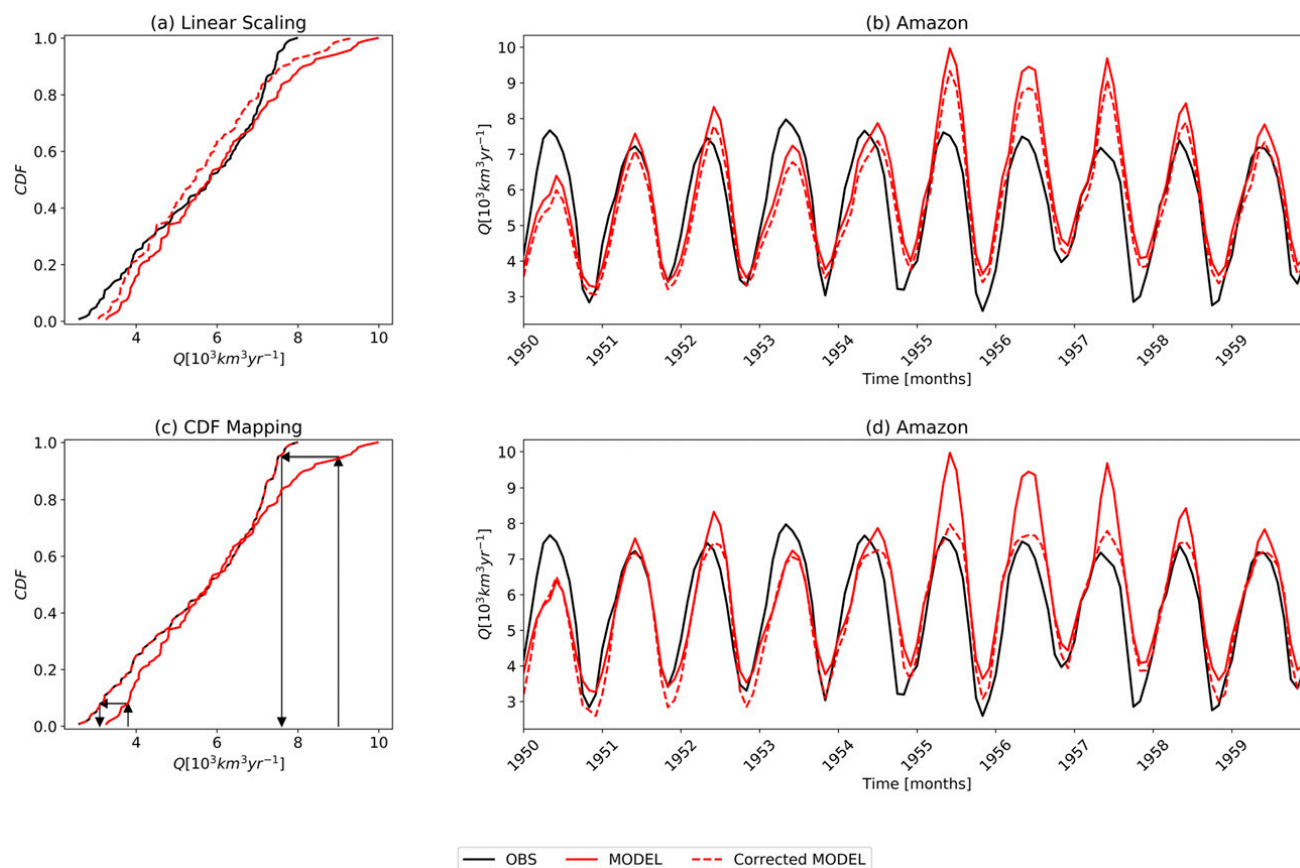
FIG. 9. Examples of bias correction methods applied to the Amazon (at Obidos) time series in the period 1950–59. (left) CDF of the observation, the model, and the model after bias correction with (a) linear scaling and (c) CDF mapping. (b),(d) As in (a) and (c), but for the Amazon time series. The arrows in (c) summarize the mapping process applied to model CDF to match observations CDF.

sensitivity to resolution in small islands is in part due to the better definition of coastlines and orography. This enhancement of surface boundary conditions constitutes an improvement in the simulated precipitation in areas like the Maritime Continent (Schiemann et al. 2014; Johnson et al. 2016). The southern part of the Andes Chain is a region exposed to westerlies that bring moisture from the Pacific, raining over the western slopes and drying over the eastern side (Viale et al. 2019), a process that is magnified at HR (see 5a). The evaluation of three small catchments shows that the river discharge generated at HR, which is almost double compared to LR, is closer to the observed river discharge (see Fig. 8d).

3) Last, in Southeast Asia, the Alps, and the Alaska Range (Figs. 8d–f), low- and high-resolution models produce a similar total amount of river flow; however, the estimates for individual catchments show, in general, a better agreement with observations at high resolution. A plausible explanation is that rain and snow are placed in the correct catchment when the resolution is enhanced.

In summary, the overall assessment over 344 monitored rivers shows that LR performs better in large catchments, while HR improves the estimation in small catchments with strong

influence of orographic precipitation. Given the strong impact of total precipitation changes on runoff (shown in section 3), the interpretation is that the improvement in river discharge at HR can be associated with a more realistic simulation of precipitation amount. First, the increased orographic precipitation helps reduce dry river discharge biases of LR simulations in many regions (e.g., Maritime Continent, southern Andes). Second, the better-defined orography favors the placement of precipitation in the correct catchment in areas of complex orography, wetting the windward side and drying the leeward side of mountain chains, leading to a more accurate simulation of river discharge of individual catchments (e.g., Southeast Asia, Alps, Alaska Range). These results give good evidence to support that the increase of precipitation in mountain or near-mountains regions at HR goes in the correct direction; however, it is not enough to draw definitive conclusions given the high fraction of unmonitored mountain catchments. For instance, the Maritime Continent, strongly affected by orographic rain, has a notable reduction of biases at HR in the five small rivers with observations, which contribute a small fraction of global runoff. However, it is highly probable that similar improvements can be found in the rest of the region, exposed to the same precipitation regime, which would have a more significant impact in the global
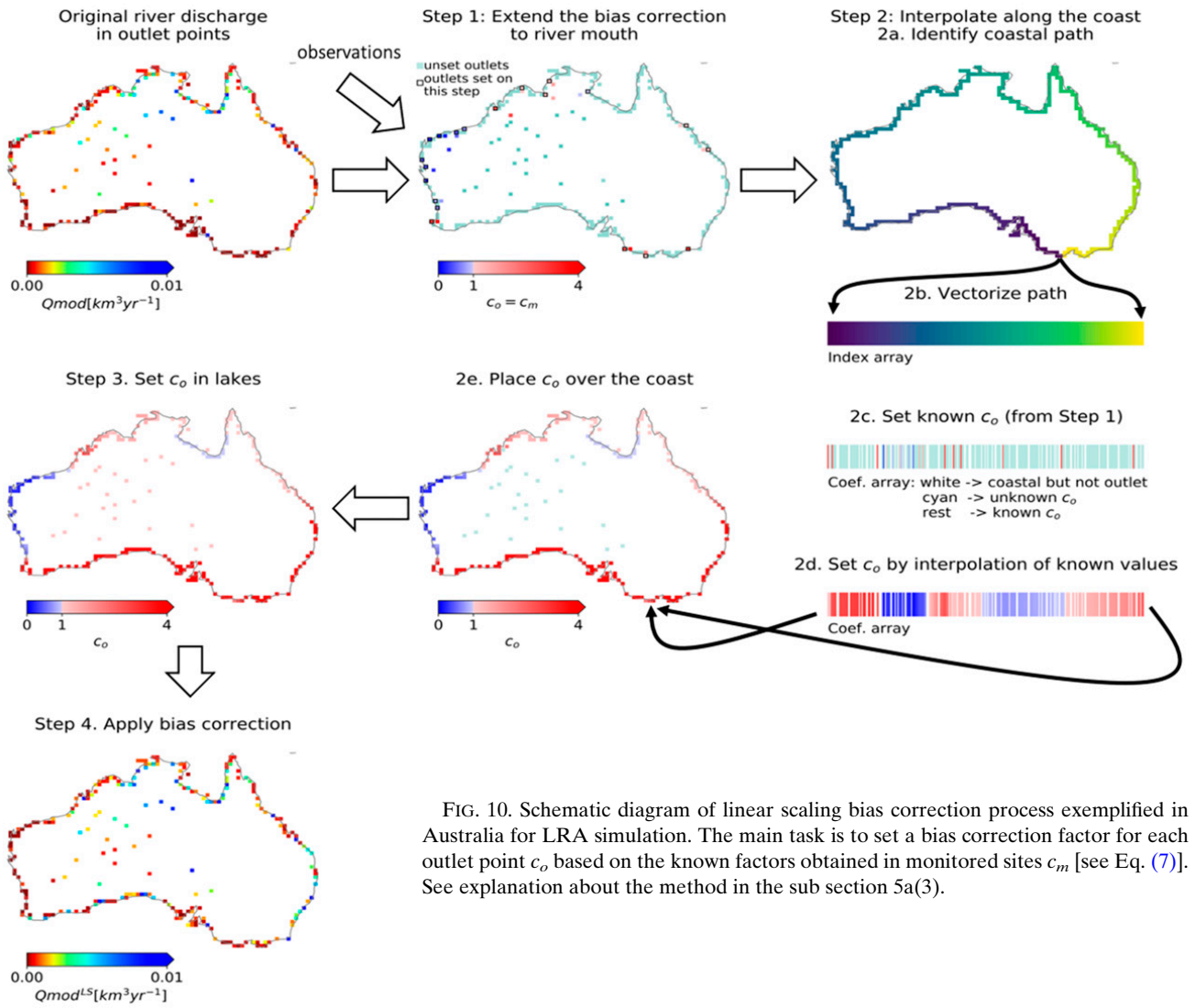
FIG. 10. Schematic diagram of linear scaling bias correction process exemplified in Australia for LRA simulation. The main task is to set a bias correction factor for each outlet point c_o based on the known factors obtained in monitored sites c_m [see Eq. (7)]. See explanation about the method in the sub section 5a(3).

discharge. Therefore, we need a method to upscale what we have learned in this section for the entire land, in order to answer the question that motivates this research: do observations underestimate the amount of orographic precipitation?

5. Global river discharge

a. Constraining global river discharge with observations

The total global discharge Q_G for each GCM is the integral of the river flow at all points that discharge into the oceans or lakes (red and pink pixels in Fig. 2a). To improve the direct GCM estimates, we propose a method to constrain the simulated values at monthly time scale with observations of the river flow available at inland points. The approach has two main steps: first, a model bias correction is applied to grid cells with observations (black pixels in Fig. 2b) and, second, the bias correction is extended to all outlet points. The following subsections present the two bias correction methods tested in this work [sections 5a(1) and 5a(2)], and a detailed description

of the overall process to constrain the global discharge with the available observations [section 5a(3)].

1) BIAS CORRECTION WITH LINEAR SCALING

Given the river flow observation time series of a monitored river Qobs and the corresponding simulated flow Qmod, the linear-scaling approach forces the simulated time series to have the same mean value as the observations:

$$\text{Qmod}_i^{\text{LS}} = \text{Qmod}_i \times c_m, \qquad (6)$$

where $\text{Qmod}_i^{\text{LS}}$ is the bias corrected value (with linear scaling) at time i,

$$c_m = \frac{\overline{\text{Qobs}}}{\overline{\text{Qmod}}} \qquad (7)$$

is the correction factor (constant) for monitored river m, and $\overline{\text{Qobs}}$ and $\overline{\text{Qmod}}$ are the long-term mean value (excluding observed missing months) of the observation and model time
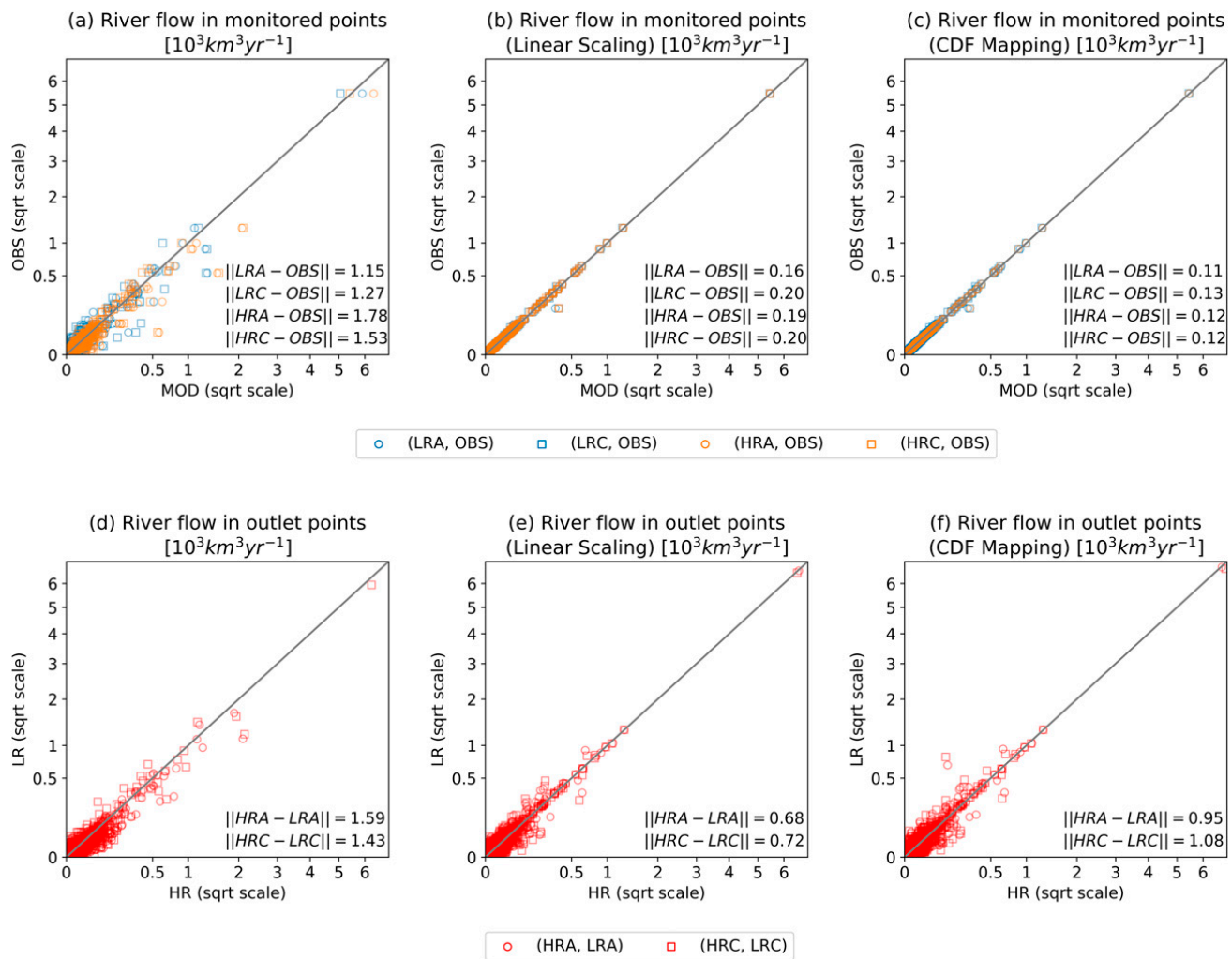
FIG. 11. Scatterplots of (top) observed mean vs model mean river discharge in monitored locations and (bottom) low- vs high-resolution mean river flow in outlet points (a),(d) before bias correction and after bias correction using (b),(e) linear scaling and (c),(f) CDF mapping. The plots are complemented with the L2-norm across all the (model, obs) pairs shown in (a)–(c), and (HR, LR) pairs in (d) and (e).

series, respectively (Lenderink et al. 2007). Note that we do not include in the c_m denominator the ratio between the drainage area informed in the observations and that used in the model, as done in Clark et al. (2015), because we directly discarded stations that do not match the reported drainage area with the catchment area defined in the model (as explained in section 2b).

Figure 9b shows an illustrative example of bias correction with linear scaling at Obidos (Amazon basin). It is evident from the time series and from the cumulative density functions (CDFs) in Fig. 9a that the model overestimates the river flow. After linear scaling, the mean value of the corrected simulations matches, by construction, the mean value of the observations ($5.6 \times 10^3 \, \mathrm{km^3 \, yr^{-1}}$).

2) BIAS CORRECTION WITH CDF MAPPING

While linear scaling accounts for the mean bias, it does not correct biases in other features of the distribution function (e.g., extremes). A more sophisticated method is CDF mapping [also called CDF matching in Reichle and Koster (2004), or CDF-t in Michelangeli et al. (2009), or quantile mapping in Maraun (2013)], which corrects the distribution function of the simulated values to agree with the observed distribution function (Teutschbein and Seibert 2012) as follows:

$$\mathrm{Qmod}_i^{\mathrm{CM}} = \mathrm{CDFobs}^{-1}[\mathrm{CDFmod}(\mathrm{Qmod}_i)], \qquad (8)$$

where $\mathrm{Qmod}_i^{\mathrm{CM}}$ is the bias corrected value (with CDF mapping) at time i, CDFobs^{-1} is the inverse function of the CDF of the observed record, and CDFmod is the CDF of the simulated time series.

The arrows in Fig. 9c help to clarify the method. The Qmod value is transformed into $\mathrm{Qmod}^{\mathrm{CM}}$ by adopting the inverse value of the observation CDF (black) at the quantile where Qmod falls in the simulated CDF (solid red). In that way, the CDF of the corrected time series (dashed red) is equal by construction to the CDF of the observations (black). By using this method, the corrected time series not only matches the mean value of the observations, but also all the statistical
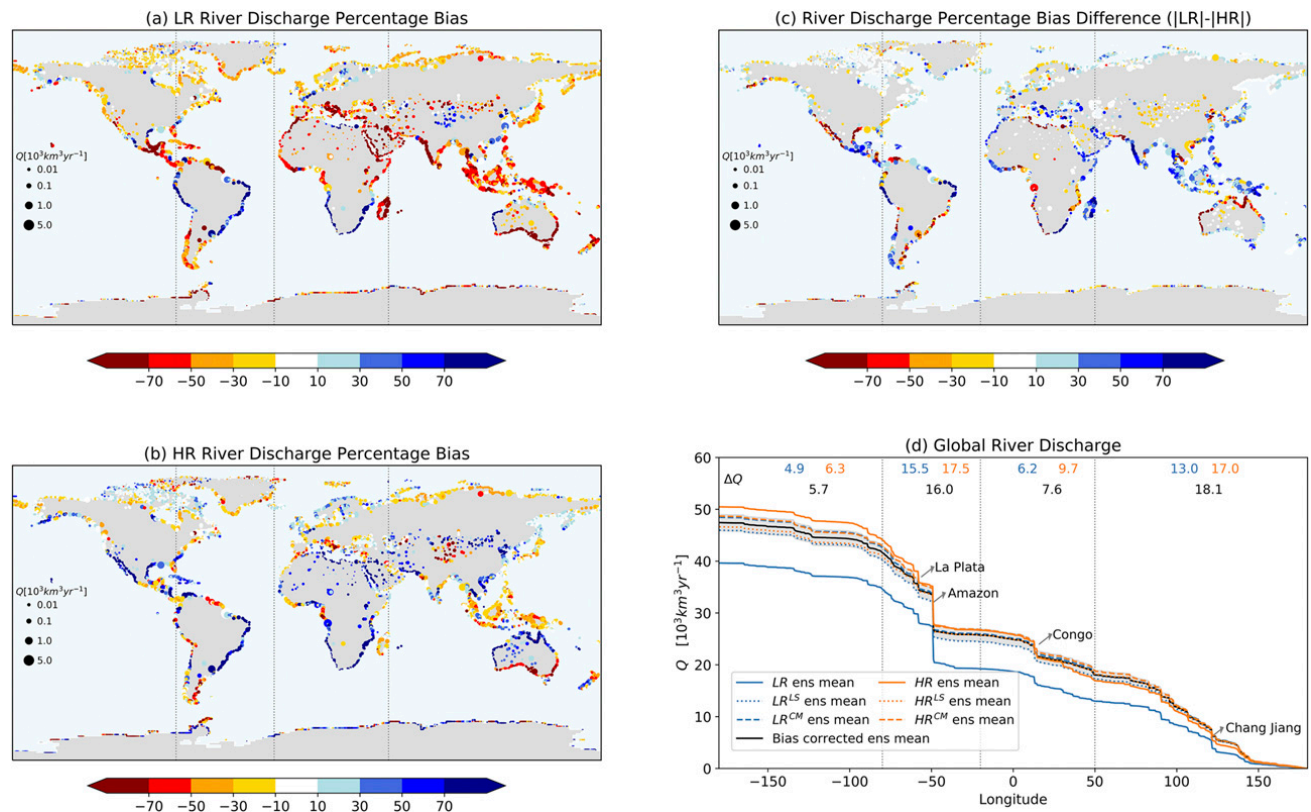
FIG. 12. River discharge percentage bias for (a) LR ensemble mean, (b) HR ensemble mean, and (c) their difference ($|\text{LR}| - |\text{HR}|$), considering the ensemble mean of all bias corrected simulations with both methods as reference. The circles size is proportional to the river flow produced at each outlet. (d) Accumulated river flow per $0.5°$ of longitude for the ensemble mean of LR and HR simulations (solid blue and orange), their respective corrected version with both methods (dotted and dashed), and the ensemble mean of all bias corrected simulations with both methods (black) with its respective uncertainty band calculated as one standard deviation of the involved simulations. The colored numbers indicate the accumulated river flow of each longitudinal segment for the original LR and HR ensembles and the bias corrected ensemble. Labels indicating the location of largest rivers are included as reference.

properties of the distribution (e.g., kurtosis, skewness). The advantages of the method appear clearly in Fig. 9d, which shows how extreme values agree in magnitude between observations and the corrected simulation, while preserving the differences in time variability.

3) FROM INLAND TO COASTAL CONSTRAINT

The simulated global river discharge is the sum of discharge at all TRIP outlet points. On the other hand, river flow observations are available near the river mouth. Therefore, a method is needed to extrapolate the correction done at the observation sites to the river mouth, and then to the rest of the outlet points. The general approach to achieve this goal is summarized in Fig. 10, just for the linear scaling method over Australia. Australia has 20 monitored rivers m, and 233 outlet points o, 22 of them belonging to endorheic systems. The overall process requires that we define a bias correction coefficient c_o for each outlet point o, and then apply the bias correction to the corresponding time series. The recipe is as follows:

- Step 1: Calculate c_m for inland points with observations by applying Eq. (7), and assign the same value $c_o = c_m$ for the corresponding outlet o located downstream of each point

with observation m. When an outlet has two or more possible m (no case in Australia), c_o is given by the weighted average of the involved c_m, where the weights are the fractional contribution of each involved river to the total observed discharge of all of them. Number of coefficients set at this step: 20.

- Step 2: A "coastal interpolation" is applied to set the coefficients along the coast. It requires (i) to trace the coastal path for indexing. The path starts at any point with a coefficient already set in step 1, and finishes at its predecessor outlet. (ii) For practical reasons, the coastal path is converted into a vector containing the position (latitude, longitude) of each point. (iii) A vector of the same length is filled with known c_o (from step 1), and with a miss when the point is not an outlet. (iv) Unknown coefficients are then calculated by linear interpolation. (v) The coefficients are placed on the map using the index array. Number of coefficients set at this step: 211.

- Step 3: There are two exceptional cases where previous steps are not enough to set all c_o. (i) Coasts without observations upstream. It occurs typically on lakes (e.g., Eyre, Chad), small islands (e.g., Hawaii, Cuba, etc.), or remote lands (e.g., Antarctica, Greenland). In these cases coefficients are

TABLE 3. Global river discharge (10^3 km^3 yr^{-1}) simulated by HadGEM3-GC3.1 at different resolutions and constrained by observations.

Simulation	Original estimate	Linear scaling	CDF mapping
LRA	38.7	45.0	46.8
LRC	40.5	46.9	50.0
HRA	52.1	46.0	47.7
HRC	48.7	47.2	49.6
Ensemble mean $\pm$ ensemble std dev	45.0 $\pm$ 5.6	46.3 $\pm$ 0.9	48.5 $\pm$ 1.3
Bias-corrected ensemble mean $\pm$ ensemble std dev		47.4 $\pm$ 1.6	

defined based on the orographic index: c_o is set as the mean c_m of all rivers with OPI falling in a close range to the index of the river under evaluation (OPI$_m \in$ OPI$_o \pm$ 0.05). (ii) Coasts with one observation upstream (usually small islands). In this case, the same c_o value is extended along the coast. Number of coefficients set at this step: 233.

- Step 4: Each outlet time series is corrected by applying Eq. (6), but using the corresponding c_o, instead of c_m.

The previous steps can also be adapted to estimate a global discharge value, but constraining the simulated time series with the CDF mapping method. The difference is that an array of coefficients is set per outlet point, instead of only one coefficient as shown for linear scaling. Let us imagine we want to extend the bias correction from a monitored point, which was corrected as in Fig. 9c, to its outlet (step 1). To perform the CDF mapping, we first need to estimate the target CDFmod^{CM} as

$$\mathrm{CDFmod}_q^{CM} = \mathrm{CDFmod}_q \times \mathrm{sc}_q, \qquad (9)$$

where sc$_q$ is the scaling coefficient for the quantile q. These coefficients are the scale factor that needs to be applied to each quantile of CDFmod to get CDFobs in the upstream monitored point, i.e., the horizontal arrow (of each quantile) in Fig. 9c. Having the target CDF defined, we proceed with the CDF mapping process given in Eq. (8), but replacing CDFobs with CDFmod^{CM}. Thus, sc is the array that can be set following similar steps to those described before, to set the coefficient c.

b. Testing the proposed methodology

Following the proposed methodology, we constrained the hydrological simulations in monitored points, and then, all outlet points, which allows to calculate a global discharge value for each simulation. Figure 11a compares observed versus predicted river flow in the 344 monitored sites for each simulation. The scatterplot shows an overall good agreement, although with positive and negative biases in most rivers, which can occasionally duplicate the observed amount. These biases are virtually eliminated with both bias correction methods (see Figs. 11b,c). The remaining biases are explained by the fact that model means are based on complete time series, while the bias correction parameters were calculated by masking out non-observed monthly values in simulated time series. Note that the Euclidean distance between simulations and observations are notably reduced from [1.15–1.78] $\times$ 10^3 km^3 yr^{-1} to ~0.19 and ~0.12 $\times$ 10^3 km^3 yr^{-1} for linear scaling and CDF mapping, respectively.

The results of the method in monitored points are promising, given that all simulations converge to observed estimates, but expected, given that the uncertainty in the bias corrected version is only produced by missing monthly values in the observations. Instead, in outlets, an extrapolation of the bias correction is needed, which notably increases the difficulty of the challenge, as it implies going from monitored points to the corresponding outlets, and extrapolating to ungauged basins along the coast. Figure 11, bottom row, shows that both methods, but mainly the linear scaling, tend to reduce the differences between low- and high-resolution simulations in the 5992 outlets. There are many dots overlapped in Figs. 11e and 11f, which make the convergence not evident visually, mainly when $Q < 0.5 \times 10^3$ km^3 yr^{-1}, but the L2-norm demonstrates that the simulated mean river flow converges to similar values in all simulations after bias correction. In particular, the Euclidean distance between low and high resolution is reduced by ~54% in average (from ~1.5 to ~0.7 $\times$ 10^3 km^3 yr^{-1}) with linear scaling, and by ~32% in average (from ~1.5 to ~1.0 $\times$ 10^3 km^3 yr^{-1}) with CDF mapping. This means that the proposed methodology remarkably reduces the uncertainty caused by resolution, in each outlet, and thereby, globally.

The global river discharge from the low-resolution ensemble mean is 39.6×10^3 km^3 yr^{-1}, but increases by 28%, up to 50.5×10^3 km^3 yr^{-1}, in the high-resolution ensemble mean. This discrepancy is notably reduced to $47.4 \pm 1.6 \times 10^3$ km^3 yr^{-1} when all the simulations are constrained by observations (see Fig. 12d and Table 3). According to this estimation, the global runoff is between the original LR and HR estimations, but closer to HR, in particular to HRC. Moreover, when bias correction is applied, both resolutions converge to a very similar global value, 46.3×10^3 km^3 yr^{-1} for linear scaling, and 48.6×10^3 km^3 yr^{-1} CDF mapping. In other words, the remaining uncertainty of the new estimate is completely independent of resolution, and it is mostly explained by the uncertainty introduced when the bias correction is extrapolated from gauged to ungauged rivers. In Fig. 12d, the numbers in color and the curves help to identify where the biases arise. LR underestimates runoff across all longitudes, while HR performs very well in Asia and slightly overestimates in the Western Hemisphere. This overestimation at HR is mostly explained by AMIP positive biases in southeast Brazil (not shown), the region where AMIP and COUPLED present the largest differences. Figures 12a and 12b present the percentage bias of LR and HR ensembles using the bias corrected ensemble mean as reference, and Fig. 12c presents their difference, with blue colors indicating an
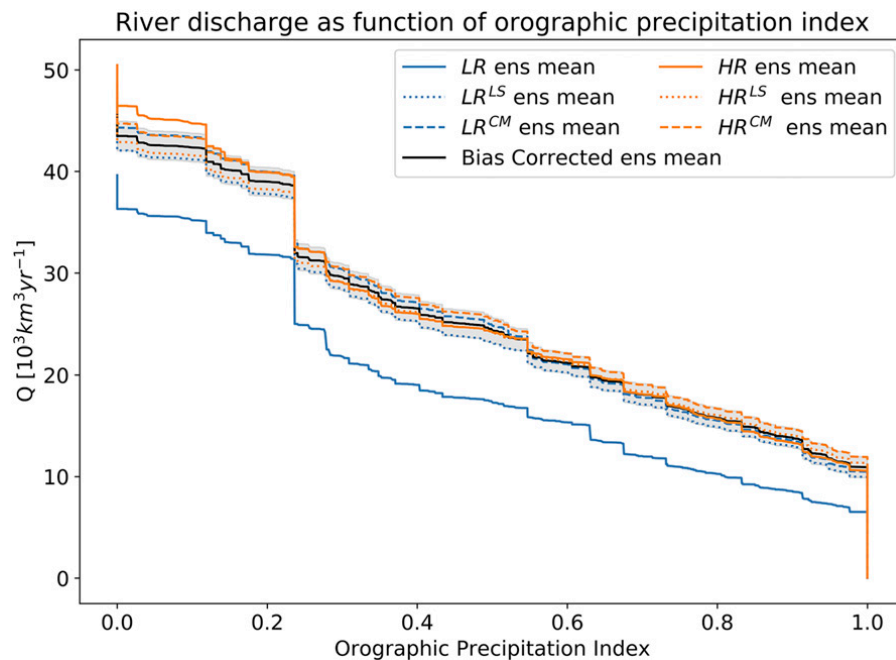
FIG. 13. Accumulated river flow discharge as function of catchments orographic precipitation index for the ensemble mean of LR simulations (solid blue), LR corrected with linear scaling (dotted blue), and LR corrected with CDF mapping (dashed blue). The same curves in orange for high resolution ensembles, and the ensemble mean of all (LR and HR) bias corrected simulations in black with its respective uncertainty band calculated as one standard deviation of the involved simulations.

improvement at HR, and red colors the contrary. The LR bias map brings out the compensation of biases, with high dry biases in Southeast Asia, the Maritime Continent, Central America, and the southern Andes compensated with high wet biases in Southern Africa and most of South America. Instead, HR notably reduces LR dry biases, although some positive biases remain. The main improvement of HR is in the intertropical belt, but also in northern Asia, Europe, and the southern Andes. Most of these improvements are attributable either to a better distribution of rain and snow (e.g., Southeast Asia, Alps, Alaska Range), or to the increase of orographic precipitation (e.g., Maritime Continent, Andes). The main weakness of HR is the overestimation of river discharge in some large catchments like La Plata or Congo.

To highlight the importance of orographic precipitation in the global generation of runoff, we plot in Fig. 13 similar curves to that showed before, but now, as function of catchments orographic precipitation index. The bias corrected simulations (gray and green bands) shows that ~50% of the global discharge is contributed by small catchments having a strong component of orographic precipitation (OPI > 0.5). Based on the bias correction done in this work, the runoff produced by this kind of precipitation is correctly simulated in high-resolution simulations, but it is underestimated by ~30% in low-resolution simulations. Conversely, the evolution of the curve in flat catchments (OPI < 0.5) for the bias corrected simulations is similar to that presented by low-resolution, but is slightly overestimated by high-resolution simulations, such as for the Congo and La Plata rivers with comparatively large

biases (see their respective jumps at OPI = 0.12 and OPI = 0.28 in Fig. 13).

c. Sensitivity of the methodology to the number of gauge stations

So far, we applied a methodology to constrain the river flow in 5992 outlets, using only 344 gauge stations where river flow is measured. The catchments of monitored rivers cover about ~40% of the global land and represent ~40% of the global river discharge. As a significant proportion of the river flow is not directly constrained by observations upstream, it is important to understand to what extent the methodology is sensitive to the lack of observations. At the same time, this will help to determine the potential benefits of including more observations in the procedure. To evaluate this, we re-estimate the global river discharge following the same methodology as before (with linear scaling and CDF mapping) but randomly discarding a percentage of the gauge stations. In this cross-validation approach, we randomly discarded five times 5%, 10%, 15%, and 20% of the gauge stations, which represent 17, 34, 52, and 69 monitored rivers, respectively. It produces a total of 160 estimations: 5 random selections × 4 discarding percentages × 4 simulations × 2 bias correction methods.

Figure 14 summarizes the result of this experiment for linear scaling in the top panel and CDF mapping in the bottom panel. Note that there is no direct relation between the percentage of gauge stations discarded and the percentage of flow that those rivers represent for the total monitored discharge (21.5 × 10^3 km^3 yr^{-1}). Next, we present the new estimates Q_G^{LS} and
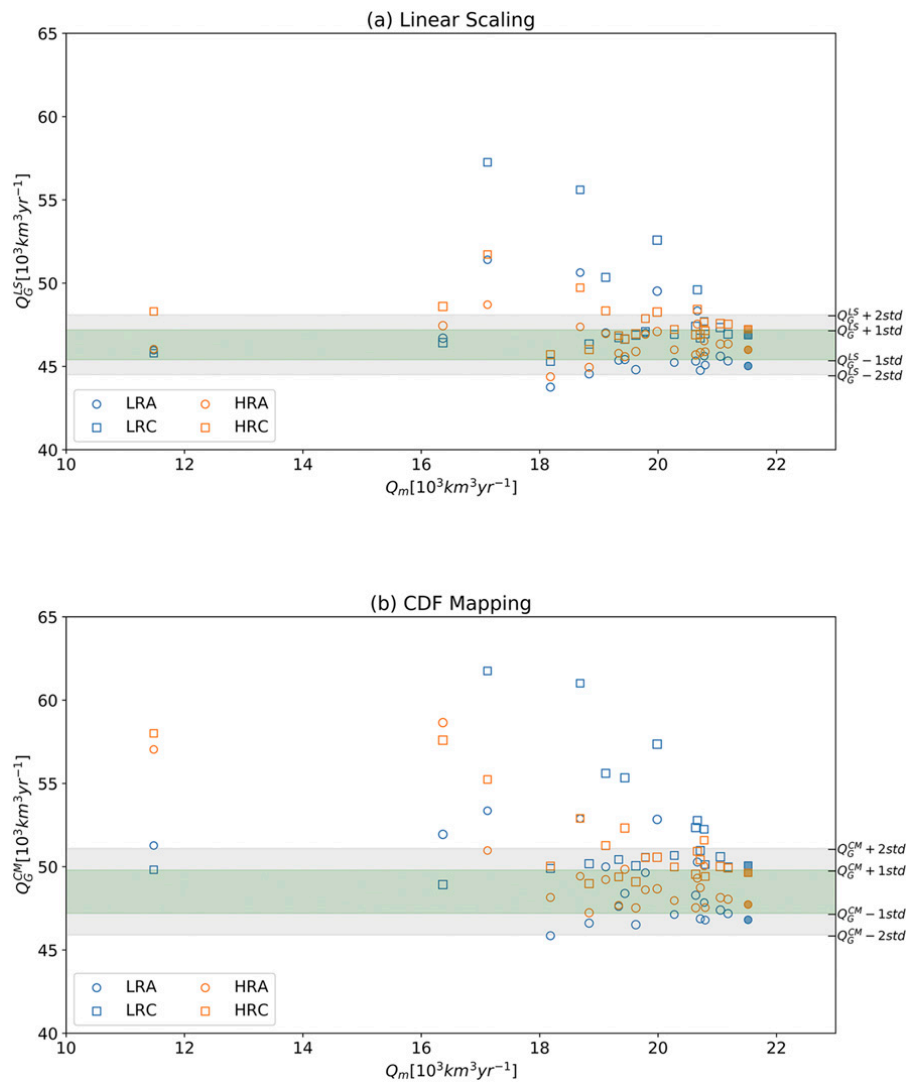
FIG. 14. Global river flow estimations based on the bias correction of each simulation by applying (a) linear scaling and (b) CDF mapping to constrain with the complete set of observation sites used in this study, and also, randomly discarding 5%, 10%, 15%, and 20% of the gauge stations. The filled markers show the original estimates using the 344 monitored rivers. The x axis shows the cumulative observed flow once some gauge stations are discarded, while the y axis shows the corresponding global estimates. The green and gray band are indicative of the range of values where the global river discharge estimate varies when all monitored rivers are used to constrain the estimate. See further details of this experiment in section 5c.

Q_G^{CM} as a function of the total monitored flow Q_m, which is more relevant than the number of gauge stations involved in each estimation. For linear scaling, the plot shows the method is robust even to discarding observation sites of rivers with high contribution of flow. About ~50% of the estimates converge in the range of $Q_G^{LS} \pm 1$ std $= 46.3 \pm 0.9 \times 10^3$ km^3 yr^{-1} (Table 3; std = standard deviation), which is obtained when all gauge stations are included, and more than 76% in the range of $Q_G^{LS} \pm 2$ std. The estimations that fall outside this range are, in general, based on low-resolution coupled (LRC) simulations. For CDF mapping, ~38% of the estimates fall in the $Q_G^{CM} \pm 1$ std $= 48.5 \pm 1.3 \times 10^3$ km^3 yr^{-1} (Table 3) range and ~73% in the $Q_G^{CM} \pm 2$ std interval. This suggests that CDF mapping is more

sensitive to the lack of observations, probably due to the fact that it corrects the extremes of each time series. Thus, a small change of extremes in a large river catchment can produce a significant change in the final integration. As for linear scaling, most outliers are related to estimations based on LRC, but also with LRA.

Beyond the differences between the two bias correction methods, some features are shared by both. In any case the minimum Q_G is 43.8×10^3 km^3 yr^{-1}, suggesting that the initial Q_G of low-resolution simulations is an underestimate. The overall method is robust even discarding up to 1.0×10^3 km^3 yr^{-1}, and it suggests that the remaining uncertainty can be reduced further if more near-mouth observations were included. On the other hand, the uncertainty increases

when rivers with high biases are discarded, e.g., Paraná (tributary of La Plata), Nile, and Indus. The lack of observations for these key rivers can exacerbate their biases when the coastal interpolation from adjacent rivers corrects the bias in the opposite direction.

6. Discussion and concluding remarks

The assessment of global precipitation patterns from GCMs is usually done by direct comparison against reanalyses, gridded observations, and/or remotely sensed data. While these products are suitable in most cases, it has been shown in previous studies that they present high uncertainty on areas (i) of low density of ground observations (e.g., in mountainous regions) (ii) with occurrence of subgrid processes (e.g., convective storms), and (iii) in high latitudes (e.g., satellites). Given that our focus is on the evaluation of precipitation sensitivity to model resolution, we proposed an indirect method of evaluation, through the assessment of river flow, a natural integrator of the water balance at the catchment scale, whose observations are particularly useful to understand the water budget in catchments with low density of in situ measurements of precipitation and/or evapotranspiration. With the final goal of advancing the understanding of precipitation biases in low and high-resolution HadGEM3-GC3.1 simulations submitted to CMIP6-HighResMIP, we have 1) assessed the river discharge in monitored rivers, and 2) extended the knowledge acquired in step 1 to ungauged rivers, to produce a global discharge estimate. The enhancement of model resolution increases the orographic precipitation due to a better definition of orographic features. It leads to higher positive biases when compared with most products based on observations, except with Stephens et al. (2012), who suggest that the lack of in situ observations over mountains produce an underestimation of orographic precipitation in reanalysis products. Combining the GCM simulations with river flow observations and orographic information, we have developed a novel methodology to constrain river discharge globally, which allows us to shed light on runoff biases, and thereby, on precipitation biases.

The analysis of the water balance sensitivity to resolution showed that HadGEM3-GC3.1 HR simulations produce ~10% more global land precipitation, which in turn, increases runoff by ~28%, but does not produce significant changes in evapotranspiration, in agreement with the findings of Demory et al. (2014) for HadGEM and Vannière et al. (2019) for HighResMIP GCMs. The main changes in precipitation occur in regions sensitive to orographic precipitation, typically over mountains, where horizontal surface fluxes prevail. This feature explains the strong and direct impact of precipitation changes on rivers, which amplify the differences observed in precipitation. Moreover, we have shown that the orographic precipitation accounts for ~40% of global land precipitation, but produces more than ~50% of global runoff. Thus, the strong sensitivity of river flow to orographic precipitation, makes it ideal to assess the potential benefits of resolution in climate models, in particular to qualitatively infer precipitation biases at the catchment scale.

The direct comparison of low- and high-resolution simulations, against the set of 344 near-coast river discharge observations,

showed that LR offers an overall better estimate of river discharge compared to HR; however, we have found common cases where the enhancement of resolution brings clear benefits. HR improves the individual catchments assessment in regions where LR compensates biases. For instance, in Southeast Asia, there is a compensation of negative biases upwind of the Himalayas, with positive biases in the lee, in LR simulations. Instead, HR simulates wetter conditions in the windward side of the Himalayas, which notably reduces the biases in Indian rivers, and drier conditions over the Tibetan Plateau, which slightly improves the simulation of river discharge in rivers that originates over the Plateau. Similarly, the assessment of catchments in the Alps and in the Alaska Range suggests that HR places the precipitation in the correct catchment, providing a more realistic spatial distribution of runoff, and thereby, a better agreement between simulated and observed river discharge in individual catchments. Another common improvement evidenced at HR, is given in rivers where orographic precipitation is notably enhanced. For instance, in the southern Andes and the Maritime Continent, there are large differences between low- and high-resolution simulations of total precipitation and runoff. The assessment suggests that the HR wetter conditions are more realistic, probably due to a better resolved coastline and orography.

The global river flow produced in catchments with orographic complexity is more than 50%; however, the monitored river flow is dominated by the contribution of large catchments (Amazon, Congo, Paraná, Nile, etc.) that lay mostly on flat terrain, which does not allow to draw definite conclusions at global scale. In other words, the set of monitored rivers is not representative (in terms of orographic influence on precipitation and runoff formation) of the set of all global rivers. To obtain a global discharge estimate, we therefore carried out a bias correction in monitored rivers, and extended the bias correction to nonmonitored rivers with a coastal interpolation. We followed this novel approach using two different bias correction methods to constrain river flow, linear scaling, and CDF mapping. The results showed that this procedure remarkably reduces the differences between the discharge estimates from low- and high-resolution models in most of the 5992 outlets, and thereby globally. Moreover, the method remains robust and independent of resolution when discarding observation sites that represent up to $1 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$, or even excluding more, if the discarded sites are not those that require strong bias correction. Comparing the bias correction techniques, CDF mapping is more sensitive to the lack of observations, probably because the correction of extremes in major rivers produce a significant change in the global discharge value. The robustness of the proposed method, as well as its independence from resolution in the estimation of global discharge estimate, makes the methodology promising for application in other CMIP6 models in future work, in particular those that follow the HighResMIP protocol. An accurate estimate of global river discharge is key to constraint the global water budget, a challenge that has been, and still is, subject of many papers (Lvovitch 1973; Trenberth et al. 2007, 2011; Rodell et al. 2015, among others).

Our best estimate of global discharge is $47.4 \pm 1.6 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$. This suggests that the real global runoff is between
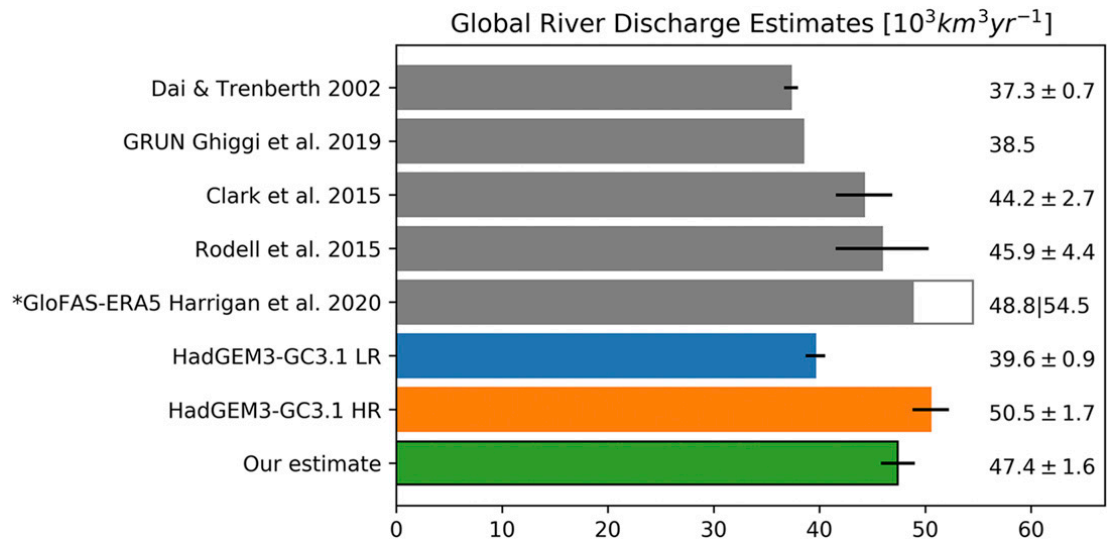
FIG. 15. Global river discharge estimated by different authors in gray and those derived in this study in colors: HadGEM3-GC3.1 LR and HR (blue and orange) and our combined estimate (green). A complete set of estimates can be found in (Clark et al. 2015). GloFAS-ERA5 (with * in the legend) $48.8 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$ is based on the period 1997–2014, while $54.5 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$ is based on the period 1979–2014. From 1979 to 1996 ERA5 presents a significant imbalance in the global water budget (see section 9.2 in Hersbach et al. 2020), which produces unrealistically high values of runoff. Therefore, we consider $48.8 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$ as a better estimate for GloFAS-ERA5.

the original LR and HR estimates (39.6 and $50.5 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$, respectively), but closer to HR. Note that the final estimation is based on eight different estimates that include variations in GCM resolution (low and high), simulation type (AMIP and COUPLED) and bias correction method (LS and CM), variations that strengthen the final result. In the light of the river discharge observations, the main advantage of high-resolution simulations is the proficiency in areas of complex orography. In these areas, the better-defined orography increases the orographic precipitation and runoff amounts, which helps reduce the high negative biases of river flow in low-resolution simulations, but also favors the placement of precipitation in the correct catchment. The highest biases of HR simulations appear in the Nile, Congo, and La Plata basins. Although excess of rain may explain part of the biases in the Nile, it is a river where models used to overestimate its flow when compared to observations (e.g., Dai and Trenberth 2002; Clark et al. 2015; Ghiggi et al. 2019) probably due to its strong intervention for pumping, and also to its exposure to high temperatures over Sahara, which facilitates the direct evaporation from the river channel, processes that are not represented in models. Beyond this particular case, an important aspect to consider in the analysis of biases at HR, is that the model tuning is done in LR models, but not in HR. This requirement of the HighResMIP protocol allows us to understand changes in model simulations only explained by the change of resolution, but at the same time, provides room for additional reduction of biases in HR simulations with model tuning. In terms of model development, preliminary results of JULES (the HadGEM3-GC3.1 land surface model) uncoupled simulations showed that the replacement of the soil hydraulic model, along with a more realistic definition of soil minerals and properties, decreases the runoff in the tropics and increases it in high latitudes. This is expected to reduce the remaining biases in high resolution simulations.

Figure 15 compares the estimates of global discharge based on low- and high-resolution simulations and the final estimate of this study with previous estimates by different authors including Dai and Trenberth (2002), Rodell et al. (2015), Clark et al. (2015), Ghiggi et al. (2019), and Harrigan et al. (2020). Our global discharge estimate is higher compared to most previous estimates with different methodologies, which are in the range $[37–46] \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$. A significant difference from those estimates is that they are based on coarse-resolution reanalysis products ($\geq 0.5°$). Our results show that the extra orographic precipitation in high resolution models, alternatively characterized as positive biases when compared to reanalysis precipitation, systematically reduces biases in simulated river flow. The global discharge based on a recent river flow reanalysis (GloFAS-ERA5; Harrigan et al. 2020) is $48.8 \times 10^3 \, \text{km}^3 \, \text{yr}^{-1}$. GloFAS-ERA5 does not correct biases: it just calibrates the hydrological model to match the seasonal variability of the observations. However, it is important to note that it is based on runoff produced by the reanalysis ERA5 at a resolution of $0.28°$, which suggests that also reanalyses tend to produce wetter conditions with the increase of resolution. Evapotranspiration processes make it impossible to directly link biases in runoff with biases in precipitation; however, given that evapotranspiration is almost insensitive to resolution in our simulations, we can infer that biases found in runoff have a strong resemblance to total precipitation biases. Thus, our results suggest that HadGEM3-GC3.1 at high resolution (~25 km) slightly overestimates land

precipitation, but notably reduces the dry biases produced at low resolution (~135 km), mainly in areas of complex orography, in agreement with the results reported by Adam et al. (2006) and Beck et al. (2020).

The developments that enabled this research offer a new capability to HadGEM3-GC3.1. TRIP is an important part of the GCM as it provides freshwater input to drive the ocean component. However, due to the lack of a better option, the routing model is conventionally run at a fixed resolution of 1°, independently of the resolution used in the GCM. The resolution discrepancy causes many estuaries to be located inland or displaced, which distorts the ocean inflow. Here, we have developed a 0.5° version of the river routing model that minimize the mismatch with the coastal points of the ocean model in high-resolution simulations. In future applications, we plan to move a step forward and develop a routing model with a river network derived from the model's orography and using the same resolution as JULES.

Last, we would like to emphasize the importance of river flow monitoring. This research has shown that river flow observations, even those measuring small catchments, are key to constrain the water balance at the catchment scale, but also at the global scale. Thus, river flow observations are relevant for climate modeling, but also for many sectors including climate change monitoring, flood monitoring, water supply, and hydropower energy. Dai et al. (2009), based on statistics of the Global Runoff Data Centre, reported that the number of stations with discharge data have continuously declined since 1979, just the year when comprehensive satellite observations began. But satellite observations do not replace river monitoring: they are complementary, and many advantages can be obtained by optimally combining different sources of observations (Lavers et al. 2019). The recovery of abandoned river flow gauge stations, as well as, monitoring more catchments, especially in places strongly affected by orographic precipitation, is crucial for the assessment of the hydrologic cycle in GCMs.

APPENDIX

TRIP Equations

The main TRIP equations are reported in Oki et al. (1999). Here, we complement that work with a detailed derivation of the equations. The river flow is calculated in all land grid cells by using a prognostic water balance model. The change in the water storage S in the river channel is defined as the difference between the sum of inflow Q_i and the outflow Q_o from the river storage in the grid cell:

$$\frac{dS}{dt} = Q_i - Q_o. \tag{A1}$$

The inflow is the sum of the runoff generated within the grid cell, and the input from upstream neighbor grid cells. The outflow is estimated as a linear function of the storage:

$$Q_o = \frac{v}{l}S, \tag{A2}$$

where v is the flow velocity, and $l = d \times m$ is the length of the river channel between grid cells, calculated as the distance d between grid cells centers multiplied by a meandering ratio m. Substituting Eq. (A2) into (A1):

$$\frac{dS}{dt} = Q_i - \frac{v}{l}S. \tag{A3}$$

Equation (A3) is a nonhomogeneous differential equation, whose solution is the sum of the solution for the related homogeneous differential equation and a solution for the nonhomogeneous equation. The homogeneous equation is given by $dS/dt = Q_o$ and the solution is

$$S(t)_h = Ae^{-\frac{v}{l}dt}. \tag{A4}$$

The nonhomogeneous equation is $dS/dt = Q_i$ and the solution is given by

$$S(t)_{nh} = \frac{l}{v}Q_i. \tag{A5}$$

Then, the solution to Eq. (A3) is

$$S(t) = Ae^{-\frac{v}{l}t} + \frac{l}{v}Q_i, \tag{A6}$$

evaluating it at $t = t_0$:

$$A = S(t_0) - \frac{l}{v}Q_i, \tag{A7}$$

then

$$S(t) = \left[S(t_0) - \frac{l}{v} Q_i \right] e^{-\frac{v}{l}t} + \frac{l}{v} Q_i, \qquad (A8)$$

$$S(t) = e^{-\frac{v}{l}t} S(t_0) + \left(1 - e^{-\frac{v}{l}t} \right) \frac{l}{v} Q_i. \qquad (A9)$$

Consequently, at time step $t + \Delta t$ the storage is estimated by

$$S(t + \Delta t) = e^{-\frac{v}{l}\Delta t} S(t) + \left(1 - e^{-\frac{v}{l}\Delta t} \right) \frac{l}{v} Q_i. \qquad (A10)$$

It means that the storage at time $t + \Delta t$ is the storage at time t that still remains in the grid cell after Δt, and the fraction of inflow that remains in the grid cell after Δt. Once the storage is known, the outflow can be calculated straightforward by solving Eq. (A1).

REFERENCES

Adam, J. C., E. A. Clark, D. P. Lettenmaier, and E. F. Wood, 2006: Correction of global precipitation products for orographic effects. *J. Climate*, **19**, 15–38, https://doi.org/10.1175/JCLI3604.1.

Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, https://doi.org/10.5194/hess-21-589-2017.

——, E. F. Wood, T. R. McVicar, M. Zambrano-Bigiarini, C. Alvarez-Garreton, O. M. Baez-Villanueva, J. Sheffield, and D. N. Karger, 2020: Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *J. Climate*, **33**, 1299–1315, https://doi.org/10.1175/JCLI-D-19-0332.1.

Best, M., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, https://doi.org/10.5194/gmd-4-677-2011.

Chahine, M. T., 1992: The hydrological cycle and its influence on climate. *Nature*, **359**, 373–380, https://doi.org/10.1038/359373a0.

Clark, E. A., J. Sheffield, M. T. H. van Vliet, B. Nijssen, and D. P. Lettenmaier, 2015: Continental runoff into the oceans (1950–2008). *J. Hydrometeor.*, **16**, 1502–1520, https://doi.org/10.1175/JHM-D-14-0183.1.

Cullen, M., 1991: The Unified Forecast/Climate Model. Seminar on Numerical Methods in Atmospheric Models, Reading, United Kingdom, ECMWF, 233–260, https://www.ecmwf.int/node/8836.

Dai, A., 2017: Dai and Trenberth global river flow and continental discharge dataset. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 28 November 2018, https://doi.org/10.5065/D6V69H1T.

——, and K. E. Trenberth, 2002: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeor.*, **3**, 660–687, https://doi.org/10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2.

——, T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *J. Climate*, **22**, 2773–2792, https://doi.org/10.1175/2008JCLI2592.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Demory, M.-E., P. L. Vidale, M. Roberts, P. Berrisford, J. Strachan, R. Schiemann, and M. Mizielinski, 2014: The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate Dyn.*, **42**, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4.

Dettinger, M. D., and H. F. Diaz, 2000: Global characteristics of stream flow seasonality and variability. *J. Hydrometeor.*, **1**, 289–310, https://doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2.

Falloon, P., R. Betts, A. Wiltshire, R. Dankers, C. Mathison, D. McNeall, P. Bates, and M. Trigg, 2011: Validation of river flows in HadGEM1 and HadCM3 with the TRIP river flow model. *J. Hydrometeor.*, **12**, 1157–1180, https://doi.org/10.1175/2011JHM1388.1.

Fekete, B. M., and C. J. Vörösmarty, 2007: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS Publ.*, **309**, 129–136.

——, ——, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles*, **16**, 1042, https://doi.org/10.1029/1999GB001254.

Ghiggi, G., V. Humphrey, S. I. Seneviratne, and L. Gudmundsson, 2019: GRUN: An observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data*, **11**, 1655–1674, https://doi.org/10.5194/essd-11-1655-2019.

Grabs, W., T. De Couet, and J. Pauler, 1996: Freshwater fluxes from continents into the world oceans based on data of the global runoff data base. Global Runoff Data Centre Rep. 10, 228 pp.

Haarsma, R., and Coauthors, 2016: High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Harrigan, S., and Coauthors, 2020: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, **12**, 2043–2060, https://doi.org/10.5194/essd-12-2043-2020.

Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, https://doi.org/10.1038/s41597-020-0453-3.

Healy, R. W., T. C. Winter, J. W. LaBaugh, and O. L. Franke, 2007: Water budgets: Foundations for effective water-resources and environmental management. USGS Circular 1308, 90 pp., https://pubs.usgs.gov/circ/2007/1308/.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Johnson, S. J., and Coauthors, 2016: The resolution sensitivity of the south Asian monsoon and Indo-Pacific in a global 0.35 AGCM. *Climate Dyn.*, **46**, 807–831, https://doi.org/10.1007/s00382-015-2614-1.

Lavers, D. A., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme, and F. Pappenberger, 2019: A vision for improving global flood forecasting. *Environ. Res. Lett.*, **14**, 121002, https://doi.org/10.1088/1748-9326/ab52b2.

Lenderink, G., A. Buishand, and W. van Deursen, 2007: Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. *Hydrol. Earth Syst. Sci.*, **11**, 1145–1159, https://doi.org/10.5194/hess-11-1145-2007.

Lin, P., and Coauthors, 2019: Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resour. Res.*, **55**, 6499–6516, https://doi.org/10.1029/2019WR025287.

Lvovitch, M. I., 1973: The global water balance. *Eos, Trans. Amer. Geophys. Union*, **54**, 28–53, https://doi.org/10.1029/EO054i001p00028.

Madec, G., and NEMO Team, 2016: NEMO ocean engine. Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace 27, 386 pp., https://www.nemo-ocean.eu/wp-content/uploads/NEMO_book.pdf.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1.

McCabe, G. J., and D. M. Wolock, 2011: Century-scale variability in global annual runoff examined using a water balance model. *Int. J. Climatol.*, **31**, 1739–1748, https://doi.org/10.1002/joc.2198.

Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, https://doi.org/10.1029/2009GL038401.

Müller, O. V., P. L. Vidale, B. Vannière, R. Schiemann, R. Senan, R. J. Haarsma, and J. H. Jungclaus, 2021: Land–atmosphere coupling sensitivity to GCMs resolution: A multimodel assessment of local and remote processes in the Sahel hot spot. *J. Climate*, **34**, 967–985, https://doi.org/10.1175/JCLI-D-20-0303.1.

Munier, S., H. Palanisamy, P. Maisongrande, A. Cazenave, and E. F. Wood, 2012: Global runoff anomalies over 1993–2009 estimated from coupled land–ocean–atmosphere water budgets and its relation with climate variability. *Hydrol. Earth Syst. Sci.*, **16**, 3647–3658, https://doi.org/10.5194/hess-16-3647-2012.

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *J. Climate*, **14**, 3307–3323, https://doi.org/10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2.

Oki, T., and Y. Sud, 1998: Design of total runoff integrating pathways (TRIP)—A global river channel network. *Earth Interact.*, **2**, 1–37, https://doi.org/10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2.

——, K. Musiake, H. Matsuyama, and K. Masuda, 1995: Global atmospheric water balance and runoff from large river basins. *Hydrol. Processes*, **9**, 655–678, https://doi.org/10.1002/hyp.3360090513.

——, T. Nishimura, and P. Dirmeyer, 1999: Assessment of annual runoff from land surface models using total runoff integrating pathways (TRIP). *J. Meteor. Soc. Japan*, **77**, 235–255, https://doi.org/10.2151/jmsj1965.77.1B_235.

——, Y. Agata, S. Kanae, T. Saruhashi, D. Yang, and K. Musiake, 2001: Global assessment of current water resources using total runoff integrating pathways. *Hydrol. Sci. J.*, **46**, 983–995, https://doi.org/10.1080/02626660109492890.

Pavelsky, T. M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. Paiva, G. H. Allen, and Z. F. Miller, 2014: Assessing the potential global extent of SWOT river discharge observations. *J. Hydrometeor.*, **519**, 1516–1525, https://doi.org/10.1016/j.jhydrol.2014.08.044.

Perry, G., P. Duffy, and N. Miller, 1996: An extended data set of river discharges for validation of general circulation models. *J. Geophys. Res.*, **101**, 21 339–21 349, https://doi.org/10.1029/96JD00932.

Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, https://doi.org/10.1029/2004GL020938.

Roberts, C., R. Senan, F. Molteni, S. Boussetta, M. Mayer, and S. Keeley, 2018: Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP. *Geosci. Model Dev.*, **11**, 3681–3712, https://doi.org/10.5194/gmd-11-3681-2018.

Roberts, M. J., and Coauthors, 2019: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, **12**, 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019.

Rodell, M., and Coauthors, 2015: The observed state of the water cycle in the early twenty-first century. *J. Climate*, **28**, 8289–8318, https://doi.org/10.1175/JCLI-D-14-00555.1.

Schiemann, R., M.-E. Demory, M. S. Mizielinski, M. J. Roberts, L. C. Shaffrey, J. Strachan, and P. L. Vidale, 2014: The sensitivity of the tropical circulation and maritime continent precipitation to climate model resolution. *Climate Dyn.*, **42**, 2455–2468, https://doi.org/10.1007/s00382-013-1997-0.

——, P. L. Vidale, L. C. Shaffrey, S. J. Johnson, M. J. Roberts, M.-E. Demory, M. S. Mizielinski, and J. Strachan, 2018: Mean and extreme precipitation over European river basins better simulated in a 25km AGCM. *Hydrol. Earth Syst. Sci.*, **22**, 3933–3950, https://doi.org/10.5194/hess-22-3933-2018.

Sinclair, M. R., 1994: A diagnostic model for estimating orographic precipitation. *J. Appl. Meteor.*, **33**, 1163–1175, https://doi.org/10.1175/1520-0450(1994)033<1163:ADMFEO>2.0.CO;2.

Stephens, G. L., and Coauthors, 2012: An update on earth's energy balance in light of the latest global observations. *Nat. Geosci.*, **5**, 691–696, https://doi.org/10.1038/ngeo1580.

Syed, T. H., J. S. Famiglietti, and D. P. Chambers, 2009: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J. Hydrometeor.*, **10**, 22–40, https://doi.org/10.1175/2008JHM993.1.

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456–457**, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052.

Titchner, H., and N. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, https://doi.org/10.1002/2013JD020316.

Trenberth, K. E., L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeor.*, **8**, 758–769, https://doi.org/10.1175/JHM600.1.

——, J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Climate*, **24**, 4907–4924, https://doi.org/10.1175/2011JCLI4171.1.

Vannière, B., and Coauthors, 2019: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. *Climate Dyn.*, **52**, 6817–6846, https://doi.org/10.1007/s00382-018-4547-y.

Vellinga, M., M. Roberts, P. L. Vidale, M. Mizielinski, M.-E. Demory, R. Schiemann, J. Strachan, and C. Bain, 2016: Sahel decadal rainfall variability and the role of model horizontal resolution. *Geophys. Res. Lett.*, **43**, 326–333, https://doi.org/10.1002/2015GL066690.

Viale, M., and Coauthors, 2019: Contrasting climates at both sides of the Andes in Argentina and Chile. *Front. Environ. Sci.*, **7**, 69, https://doi.org/10.3389/fenvs.2019.00069.

Weitzman, M. S., 1970: Measures of Overlap of Income Distributions of White and Negro Families in the United States. U.S. Bureau of the Census Tech. Paper 22, 23 pp.

Williams, K., and Coauthors, 2018: The Met Office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J. Adv. Model. Earth Syst.*, **10**, 357–380, https://doi.org/10.1002/2017MS001115.