

## Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances\*

CHRISTOF LORENZ AND HARALD KUNSTMANN

*Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany*

BALAJI DEVARAJU, MOHAMMAD J. TOURIAN, AND NICO SNEEUW

*Institute of Geodesy, University of Stuttgart, Stuttgart, Germany*

JOHANNES RIEGGER

*Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany*

(Manuscript received 13 September 2013, in final form 28 April 2014)

### ABSTRACT

The performance of hydrological and hydrometeorological water-balance-based methods to estimate monthly runoff is analyzed. Such an analysis also allows for the examination of the closure of water budgets at different spatial (continental and catchment) and temporal (monthly, seasonal, and annual) scales. For this analysis, different combinations of gridded observations [Global Precipitation Climatology Centre (GPCC), Global Precipitation Climatology Project (GPCP), Climate Prediction Center (CPC), Climatic Research Unit (CRU), and University of Delaware (DEL)], atmospheric reanalysis models [Interim ECMWF Re-Analysis (ERA-Interim), Climate Forecast System Reanalysis (CFSR), and Modern-Era Retrospective Analysis for Research and Applications (MERRA)], partially model-based datasets [Global Land Surface Evaporation: The Amsterdam Methodology (GLEAM), Moderate Resolution Imaging Spectroradiometer (MODIS) Global Evapotranspiration Project (MOD16), and FLUXNET Multi-Tree Ensemble (FLUXNET MTE)], and Gravity Recovery and Climate Experiment (GRACE) satellite-derived water storage changes are employed. The derived ensemble of hydrological and hydrometeorological budget-based runoff estimates, together with results from different land surface hydrological models [Global Land Data Assimilation System (GLDAS) and the land-only version of MERRA (MERRA-Land)] and a simple predictor based on the precipitation–runoff ratio, is compared with observed monthly in situ runoff for 96 catchments of different sizes and climatic conditions worldwide. Despite significant shortcomings of the budget-based methods over many catchments, the evaluation allows for the demarcation of areas with consistently reasonable runoff estimates. Good agreement was particularly observed when runoff followed a dominant annual cycle like the Amazon. This holds true also for catchments with an area far below the spatial resolution of GRACE, like the Rhine. Over catchments with low or nearly constant runoff, the budget-based approaches do not provide realistic runoff estimates because of significant biases in the input datasets. In general, no specific data combination could be identified that consistently performed over all catchments. Thus, the performance over a specific single catchment cannot be extrapolated to other regions. Only in few cases do specific dataset combinations provide reasonable water budget closure; in most cases, significant imbalances remain for all the applied datasets.

---

 Denotes Open Access content.

---

\* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JHM-D-13-0157.s1>.

---

Corresponding author address: Christof Lorenz, Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany.  
E-mail: christof.lorenz@kit.edu

### 1. Introduction

With water, the most essential natural resource for sustaining life, coming under threat (Vörösmarty et al. 2010), it becomes imperative to understand the hydrological cycle and monitor its constituents. This urgent need, however, has not been able to stem the steady decline in the number of precipitation and runoff gauges (Fig. 1), which still provide the only direct observations of terrestrial water cycle components. This is a setback both for our understanding of the evolution of the

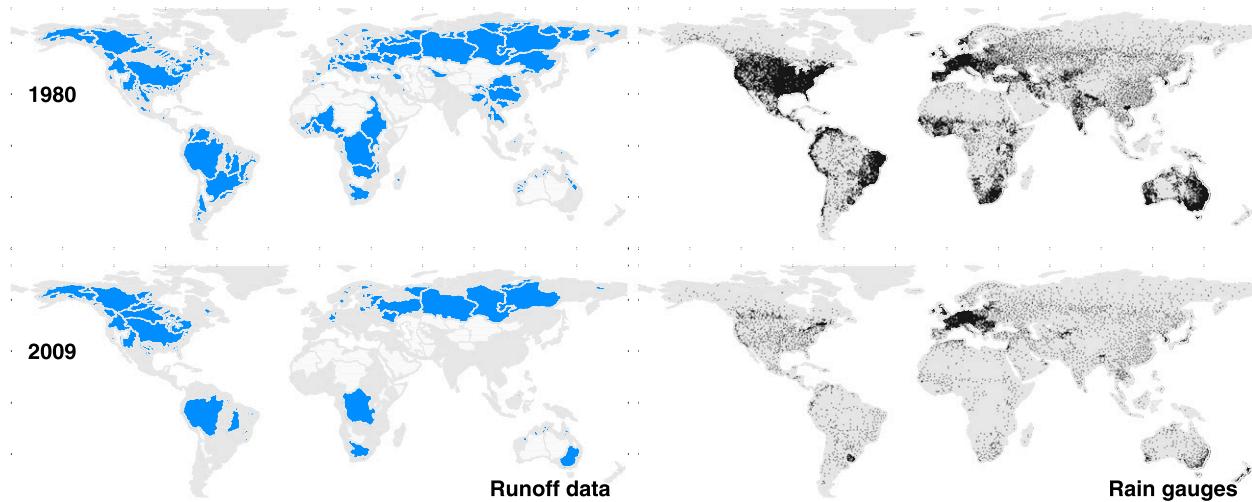


FIG. 1. The maps show the decline in the number of (left) runoff and (right) rain gauges. The catchments whose runoff measurements are publicly available (blue), that do not discharge into the oceans or other open water bodies (white), and those that are not gauged or whose runoff measurements are not publicly available (gray) are shown in the left column. In the right column, the individual dots represent one or more rain gauges in a  $0.5^\circ \times 0.5^\circ$  grid cell.

hydrological cycle and for the calibration and validation of hydrological and atmospheric models and ex situ measurements of those components. The situation is further aggravated by the lack of reliable data for evapotranspiration and water storage changes, which, together with precipitation and runoff, form the major components of large-scale hydrological cycle.

Over the years, in the absence of measurements of the hydrological cycle components, it has been normal practice to resort to sophisticated hydrological and atmospheric reanalysis models to study the global water balance (e.g., Trenberth et al. 2007). However, the hydrological or land surface models suffer from the steady changes in the spatial distribution and number of precipitation or runoff gauges. For such models, observations are needed both as an input and also as a validation measure. Therefore, there is a desperate need for new and alternative methods.

Spaceborne measurements seem to provide the necessary respite by a plethora of satellites providing nearly all the components of the global hydrological cycle: precipitation from the Tropical Rainfall Measuring Mission (TRMM; Kummerow et al. 2000), water storage changes from the Gravity Recovery and Climate Experiment (GRACE; Schmidt et al. 2008), runoff from satellite altimetry and synthetic aperture radar (Alsdorf et al. 2007), soil moisture from the Soil Moisture and Ocean Salinity (SMOS) satellite (Kerr et al. 2010) or (beginning in late 2014) the Soil Moisture Active Passive (SMAP) satellite (Entekhabi et al. 2010), and a host of variables concerning the hydrological cycle from *Aqua* (Parkinson 2003). The spaceborne observations

bring with them the advantages of global coverage and homogeneous accuracy, but with limited spatial and temporal resolutions. Hrachowitz et al. (2013) acknowledge these advancements and provide examples of spaceborne measurements becoming indispensable to hydrological research, especially in data-poor regions like West Africa, Asia, and the vast open oceans.

In the context of runoff, Alsdorf et al. (2007) provide a broad overview of the types of spaceborne runoff measurements and their relative merits and challenges. Recently, but not for the first time, Tourian et al. (2013) demonstrated the potential and accuracy of runoff derived from satellite altimetry. However, an application of this method to catchments still depends on the existence of contemporaneous or previous in situ runoff measurements. Therefore, our study focuses on independent methods using mass estimates of the water storage in river systems from GRACE, while observed runoff is used as a validation measure only.

The GRACE satellite mission by design provides indirect estimates of water storage changes (Tapley et al. 2004), which has tremendously revolutionized global- and continental-scale hydrological studies. This is corroborated by the wealth of literature dealing with a range of applications: estimation of basin- to continental-scale water storage (Crowley et al. 2006; Schmidt et al. 2008) or evapotranspiration (Rodell et al. 2004a), regional weather analysis (Seitz et al. 2008), large-scale moisture flux divergences (Swenson and Wahr 2006); groundwater depletion in India (Rodell et al. 2009), and polar ice mass loss monitoring (Schrama and Wouters 2011).

Nevertheless, the GRACE data come with their own problems, the most persistent being the filtering problem (e.g., Werth et al. 2009). The aftereffects of the filtering, leakage, and poor spatial resolution limit the utility of GRACE data only to catchments with an area of  $\sim 200\,000 \text{ km}^2$  or more (Longuevergne et al. 2010). Despite these challenges, our previous studies (Rieger et al. 2012; Fersch et al. 2012) and current improvements in GRACE data processing (Dahle et al. 2013; Bettadpur 2012) provide enough grounds to pursue a GRACE-based approach for global-scale runoff estimation.

The use of GRACE water storage changes entails the use of the water balance equations, wherein runoff is computed from a combination of precipitation and evapotranspiration in addition to water storage changes. Such water budget-based methods have already been applied by Sheffield et al. (2009), who used GRACE and other remotely sensed data for computing streamflow from the Mississippi basin. Syed et al. (2005) used an atmospheric water balance for estimating discharge over the Amazon and Mississippi basins and, in the following years, total freshwater discharge from the pan-Arctic region (Syed et al. 2007) or continents and large river basins (Syed et al. 2009). Other methods of estimating runoff from GRACE are also being proposed (e.g., Rieger and Tourian 2014), but they are still in a very nascent state of development. For a long time, evapotranspiration has been elusive and has been the thorn in the flesh in global hydrological studies (e.g., Trenberth et al. 2007; Schlosser and Houser 2007). Lately, attempts have been made to provide global evapotranspiration estimates apart from what has been available from various hydrological and atmospheric reanalysis models (Salvucci and Gentile 2013; Miralles et al. 2011; Jung et al. 2009; Mu et al. 2007). Precipitation and runoff datasets are the most mature of all the hydrological datasets, but there are significant differences between the different precipitation datasets (Lorenz and Kunstmann 2012). With an accuracy of 10%–20%, observed runoff is the most accurately measured component of the hydrological cycle (Fekete et al. 2012), but suffers from variations in the time period of observation, substantial delays in data access, and large declines in the monitoring capacity (Fekete et al. 2002; Fekete and Vörösmarty 2007; Shiklomanov et al. 2002; Vörösmarty et al. 2001).

To analyze the performance of runoff estimations independent of gauge-based observations, we thus set out to derive large-scale runoff from catchments using water-balance-based methods. Therefore, we use the most recent versions of widely used data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences. By combining all these datasets, we end up with a total of 90 possible runoff

estimates from a terrestrial water balance and 15 from an atmospheric–terrestrial water balance for each catchment. These estimates are then validated against observed runoff data from the Global Runoff Data Centre (GRDC; GRDC 2013) over 96 river basins. The performance of these estimates is also compared with the performances of sophisticated hydrological (land surface) models and an empirical method involving the precipitation–runoff ( $P-R$ ) ratio. This allows us to:

- evaluate the potential of runoff estimation from water-balance-based methods, land surface models, and an empirical predictor;
- analyze the quality of the different contributing datasets; and
- evaluate the current status of the closure of catchment-scale water budgets.

Until now, only a few studies existed where runoff from such a large number of river basins was evaluated. Some examples of these are Zaitchik et al. (2010), which focuses on the validation of Global Land Data Assimilation System (GLDAS) over 66 larger river basins, or Zhou et al. (2012), where runoff estimates from different land surface models are compared against mean annual runoff observations over 150 catchments. A validation of runoff from global climate models using the Total Runoff Integrating Pathways (TRIP) river flow model is presented in Falloon et al. (2011). In most cases, such studies focus on hydrological models, while we estimate runoff from simple budget equations where the most recent gridded observation datasets, land surface, and atmospheric reanalysis models serve as inputs.

Our study is outlined as follows: in section 2, we elaborate on the water-balance-based methods and the empirical method for estimating runoff from precipitation; in section 3, the datasets that we have used in the different approaches are detailed; in section 4, we investigate the utility of the runoff estimates from water-balance-based methods as an alternative to observed runoff and also exploit those estimates to study the global water balance closure; and finally, we draw our conclusions in section 5.

## 2. Methods

The primary concern in this study is to estimate runoff globally from the landmasses, and we have chosen to use the water balance equations as the means to our end. Water balance equations define the hydrological cycle in quantitative terms by taking into account the amount of water entering and leaving a river basin within, for example, 1 month. This budgeting can be approached in

several different ways, and here we have chosen both large-scale terrestrial and atmospheric–terrestrial water balance equations. Our choice was driven mainly by the availability of water storage change data from the GRACE satellite mission. All our calculations are performed on a monthly basis in order to be as consistent as possible with the monthly runoff observations as provided from the GRDC.

#### a. Hydrological budget approach

At continental scales, the water entering a river basin is described by precipitation  $P$ , the water leaving it is described by evapotranspiration  $ET_a$  and runoff  $R$ , and whatever remains is the water storage change  $dS/dt$ . Mathematically, this can be written as

$$P - ET_a - R = \frac{dS}{dt}, \quad (1)$$

which is the large-scale terrestrial water balance equation, and it immediately allows us to compute the river runoff as

$$R = P - ET_a - \frac{dS}{dt} \equiv R_{\text{hyd}}. \quad (2)$$

Since this method uses only hydrological variables to arrive at runoff estimates, it will henceforth be called the hydrological approach. This approach does not require runoff routing as it is taken care of by the water storage changes, and it comes with the added benefit that no matter which component we estimate, the closure of the water balance can be analyzed. The major pitfall of such an approach is that the accuracy of the runoff estimates will be only as good as the least accurate dataset. Because of the law of error propagation, errors in the individual water cycle components might thus add up and result in very uncertain estimates of runoff. Further, since at any point there is a need for three different datasets to study one of the variables of interest, consistency in terms of spatial and temporal resolution are of paramount importance. On the other hand, biases in the individual input datasets might get canceled out as runoff is computed as a residual.

#### b. Hydrometeorological budget approach

The number of inputs in (1) can be reduced to two by linking the terrestrial water balance with the atmospheric–terrestrial water balance equation (Roads et al. 1994)

$$\frac{dW}{dt} + \nabla \cdot \mathbf{Q} = ET_a - P, \quad (3)$$

where  $W$  denotes the total column water content in the atmosphere and  $\nabla \cdot \mathbf{Q}$  is the net balance of moisture flux

(i.e., moisture flux divergence). The moisture flux is defined as

$$\mathbf{Q} = \frac{1}{g} \int_{p=0}^{p_{\text{sfc}}} \mathbf{v}_h(p) q(p) dp, \quad (4)$$

with air pressure  $p$  (Pa), the gravitational acceleration  $g$  ( $\text{m s}^{-2}$ ), the horizontal wind vector  $\mathbf{v}_h$ , and the specific humidity  $q$  ( $\text{kg kg}^{-1}$ ; e.g., Fersch et al. 2012). According to Peixoto and Oort (1992), changes in the atmospheric water content  $dW/dt$  can be neglected on monthly or longer time scales, which gives us the result

$$\nabla \cdot \mathbf{Q} \approx ET_a - P. \quad (5)$$

By combining (1) and (5), runoff can be derived from the atmospheric–terrestrial water balance

$$R_{\text{atm}} = -\nabla \cdot \mathbf{Q} - \frac{dS}{dt}. \quad (6)$$

We will call this method the hydrometeorological approach as it takes both hydrological and atmospheric datasets as inputs. This method also inherits the advantages and disadvantages of the hydrological approach.

#### c. Runoff–precipitation ratio

Since there is a strong relationship between runoff and precipitation, runoff can be estimated directly from precipitation by scaling it with an empirically derived scaling factor. This factor is the ratio between runoff and precipitation, which is computed for each calendar month (January–December) individually. For each catchment, the runoff value of the  $j$ th calendar month in the  $i$ th year  $R_{ij}$  is divided by the corresponding precipitation value  $P_{ij}$ :

$$\kappa_{ij} = \frac{R_{ij}}{P_{ij}}. \quad (7)$$

Then, an average  $\bar{\kappa}_j$  of the ratios  $\kappa_{ij}$  for each calendar month  $j$  is computed over the available  $N$  number of years:

$$\bar{\kappa}_j = \frac{1}{N} \sum_{i=1}^N \kappa_{ij}. \quad (8)$$

The runoff is then computed for each month by scaling precipitation  $P_{ij}$  with  $\bar{\kappa}_{ij}$ :

$$R_{ij} = \bar{\kappa}_j P_{ij}. \quad (9)$$

This empirical model depends on previous runoff data, and the past and current precipitation data, which makes it more of a predictor model rather than an estimator.

The computation of  $\bar{\kappa}_j$  for individual calendar months takes care of the mean annual cycle, and by scaling them with the concurrent precipitation data, the associated climatic variations are also imbued into the runoff estimates.

Due in part to the simplicity and empirical nature of the model, its success is entailed to the dynamic behavior of the river system whose runoff is estimated. Runoff of a river system more prone to vagaries in the mean annual hydrological cycle, or in which there is a lot of anthropogenic intervention, might not be estimated via this model. Furthermore, evapotranspiration and water storage do not appear explicitly in these formulas, but they certainly affect the relationship between precipitation and runoff (i.e., their impacts are “hidden” within the ratios). It is thus essential that both evapotranspiration and water storage changes follow a stable annual cycle, which does not change significantly from year to year. That being said, estimating runoff from catchments with, for example, a trend in one of these two variables might lead to large uncertainties.

To evaluate the performance of this empirical model as a predictor (i.e., for ungauged regions or runoff observations with large data gaps), the values of  $\bar{\kappa}_j$  are computed from runoff and precipitation data only until the end of 2002. From 2003, runoff is estimated by multiplying current precipitation data with the mean monthly scaling factor  $\bar{\kappa}_j$  of the corresponding month  $j$ . For analyzing the reliability of this approach, different precipitation datasets (see section 3) have been used for both the estimation of  $\bar{\kappa}_j$  and the prediction of runoff.

#### *d. Water budget imbalance*

The GRACE-based methods that we use here are based on the water balance equations, which allow us to compute the imbalance in the water budget:

$$\delta_{\text{hyd}} = R - R_{\text{hyd}} = R - P + \text{ET}_a + \frac{dS}{dt} \quad \text{and} \quad (10a)$$

$$\delta_{\text{atm}} = R - R_{\text{atm}} = R + \nabla \cdot \mathbf{Q} + \frac{dS}{dt}, \quad (10b)$$

where  $\delta_{\text{hyd}}$  and  $\delta_{\text{atm}}$  are the water budget imbalances of the hydrological and atmospheric water balance equations, respectively.

### 3. Data

The water balance equations (2) and (6) and the empirical model (9) use all the components of the hydrological cycle. We have ensured that all the datasets that we have used for our analysis here are observation based. In cases where direct observations are not possible, for

example, moisture flux divergence, only partially modeled datasets are used. In addition to this, we have also taken care to use only datasets that are publicly available. An overview of the different data sources of precipitation, evapotranspiration, water storage changes, moisture flux divergences, and runoff is given in Table 1.

For the comparison of the runoff estimates with the observations, we have remapped all the input fields to the same spatial resolution ( $0.5^\circ \times 0.5^\circ$ ) using a first-order conservative interpolation (Jones 1999). From these fields, area-weighted averages were computed over the study regions shown in Fig. 2. To be temporally consistent with water storage changes from GRACE, Landerer et al. (2010) proposed to smooth the time series of the other water cycle variables according to

$$\tilde{F}_i = \frac{1}{4}F_{i-1} + \frac{1}{2}F_i + \frac{1}{4}F_{i+1}, \quad (11)$$

where the  $F$  variables are time series of precipitation, evapotranspiration, moisture flux divergences, and observed and modeled runoff. The indices  $i$ ,  $i - 1$ , and  $i + 1$  refer to the previous, current, and following month, respectively. We have applied this filter to all of our estimates.

#### *a. Runoff*

To attain maximum data coverage, both in space and time, data are collected from different sources: the GRDC; the Arctic Rapid Integrated Monitoring System (ArcticRIMS) project; Water Survey Canada; the U.S. Geological Survey (USGS); the U.S. Army Corps of Engineers (USACE); the Environmental Research Observatory (ORE) Geodynamical, Hydrological, and Biogeochemical Control of Erosion/Alteration and Material Transport in the Amazon Basin (HYBAM) project; the Department of Water, Land and Biodiversity Conservation, Government of Australia; and the Department of Water Affairs and Forestry, Republic of South Africa.

The GRACE-based approaches (i.e.,  $R_{\text{hyd}}$  and  $R_{\text{atm}}$ ) are essentially global methods capable of providing runoff estimates of all catchments without the necessity of any runoff observations. To validate these estimates, we need contemporaneous (2003–10) observed runoff data for all the catchments. Because of data gaps both in space and time, we choose catchments that have a minimum of five years of observed runoff data within the period 2003–10 so that we can perform a meaningful statistical assessment. Applying this restriction, we are left with 96 catchments out of a total of 403 catchments to perform our investigation. An overview over the different study regions, the respective river gauges, the area of the basins, and the mean annual discharge is given in Table 2 and Fig. 2. The number of available

TABLE 1. Summary of the observation- and model-based datasets containing precipitation ( $P$ ), actual evapotranspiration ( $ET_a$ ), vertically integrated moisture flux divergences ( $\nabla \cdot \mathbf{Q}$ ), water storage changes ( $dM/dt$ ), and runoff ( $R_{\text{obs}}$  and  $R_{\text{mod}}$ ).

Variable	Dataset	Version	Resolution		Time period
			Spatial	Temporal	
$P$	GPCC	6.0	$0.5^\circ \times 0.5^\circ$	1 month	1901–2010
	GPCP	2.2	$2.5^\circ \times 2.5^\circ$	1 month	1979–present
	CPC	1.0	$0.25^\circ \times 0.25^\circ$	1 month	1979–present
	CRU	3.1	$0.5^\circ \times 0.5^\circ$	1 month	1901–2009
	DEL	2.01	$0.5^\circ \times 0.5^\circ$	1 month	1900–2008
$ET_a$	ERA-Interim	—	$0.75^\circ \times 0.75^\circ$	1 month, 1 day, 6 h	1979–present
	GLDAS	Noah 3.3	$1.0^\circ \times 1.0^\circ$	1 month, 3 h	1948–present
	GLEAM	v1B	$0.25^\circ \times 0.25^\circ$	1 day	1984–2007
	MOD16	—	$0.5^\circ \times 0.5^\circ$	1 yr, 1 month, 8 days	2000–10
	FLUXNET MTE	—	$0.5^\circ \times 0.5^\circ$	1 month	1982–2011
	MERRA-Land	1.0	$1/3^\circ \times 2/3^\circ$	1 month, 1 day, 1 h	1980–present
$\nabla \cdot \mathbf{Q}$	ERA-Interim	—	$0.75^\circ \times 0.75^\circ$	1 month, 1 day, 6 h	1979–present
	MERRA	—	$1/3^\circ \times 2/3^\circ$	1 month, 1 day, 6 h, 3 h, 1 h	1979–present
	CFSR	—	$0.5^\circ \times 0.5^\circ$	1 month, 1 day, 6 h	1979–2011
$dM/dt$	GRACE GFZ	R5	—	1 month	2003–present
	GRACE CSR	R5	—	1 month	2003–present
$R_{\text{obs}}$	GRDC	—	—	—	—
$R_{\text{mod}}$	GLDAS	VIC	$1.0^\circ \times 1.0^\circ$	1 month, 3 h	1979–present
	GLDAS	CLM 2.0	$1.0^\circ \times 1.0^\circ$	1 month, 3 h	1979–present
	GLDAS	Noah 2.7	$0.25^\circ \times 0.25^\circ$	1 month, 3 h	2000–present
	GLDAS	Noah 3.3	$1.0^\circ \times 1.0^\circ$	1 month, 3 h	1948–present
	GLDAS	Mosaic	$1.0^\circ \times 1.0^\circ$	1 month, 3 h	1979–present
	MERRA-Land	1.0	$1/3^\circ \times 2/3^\circ$	1 month, 1 day, 1 h	1980–present

runoff observations is shown in Fig. 3, where the white and gray boxes indicate available and missing data, respectively.

Spatially, the collection of catchments covers a wide range of climatic and geographic conditions. This includes some tropical catchments with the highest discharge worldwide (e.g., Amazon and Congo) and also the vast high-latitude basins (e.g., Lena, Ob, and Yana)

with typically low discharge values during the winter months due to frozen conditions and peak values between June and August during and after the melting season. Also, the areal extent of the selected catchments varies from  $4.7 \times 10^6$  to  $2.1 \times 10^4 \text{ km}^2$ , where we have deliberately included the small catchments ( $< 200\,000 \text{ km}^2$ ) as well. This allows the quantification of the area dependence of GRACE-based estimates. To better characterize

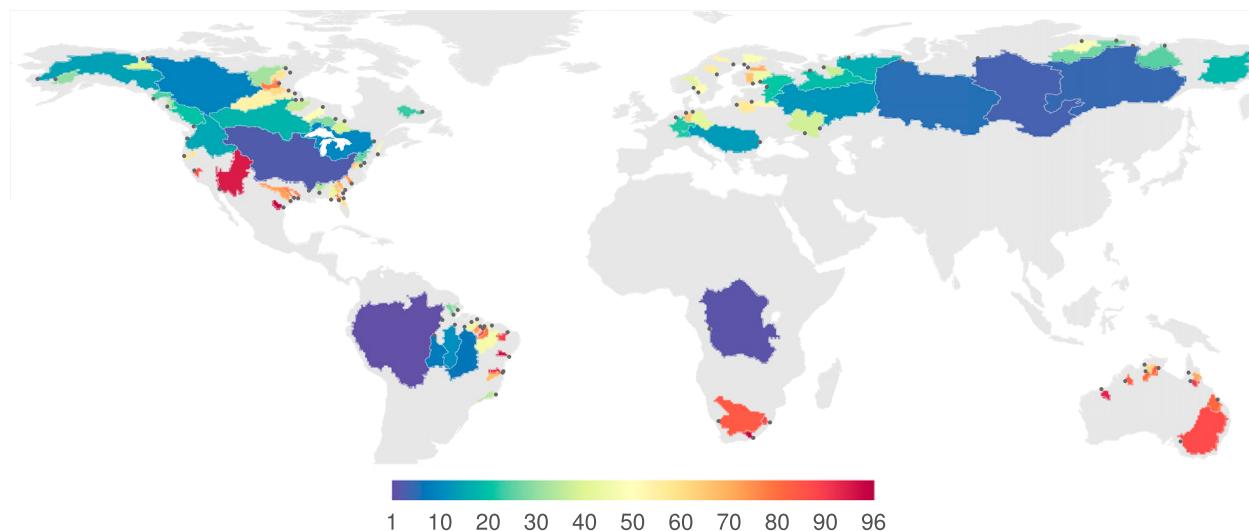


FIG. 2. Location of the basins given in Table 2, where the dot in each basin is located at the respective river gauge. As the catchments are ordered with respect to their mean annual discharge, the bluish (reddish) basins further indicate high (low) discharge.

the study regions, Figs. 4 and 5 show the annual cycle of precipitation, evapotranspiration, runoff, water storage changes, and moisture flux convergences for all the basins.

### b. Precipitation

Precipitation is traditionally measured using various types of rain gauges, such as the nonrecording cylindrical container, the recording weighing gauge, floating gauges, and tipping-bucket-type gauges. In the recent past, precipitation was also measured with weather radar and weather satellites. Generally, because of the adequate amount of sensors, global gridded precipitation data are available from different data providers. In this study, we used precipitation data from the following:

- Global Precipitation Climatology Centre (GPCC; Schneider et al. 2014),
- Global Precipitation Climatology Project (GPCP; Adler et al. 2003),
- Climate Prediction Center (CPC) Unified precipitation (Chen et al. 2008),
- Climatic Research Unit (CRU; Harris et al. 2013), and
- University of Delaware (DEL; Matsuura and Willmott 2012).

From these datasets, GPCC, CPC, CRU, and DEL are derived from gauge observations only. The largest number of stations is included in the GPCC product, while CPC archives a higher station density, for example, over North America. However, because of the very nature of such data products, the precipitation estimates can be highly uncertain, especially in regions with only a few or no stations. Figure 1 clearly shows that large parts over, for example, South America, Africa, or Siberia are completely unobserved. Moreover, the number and distribution of rain gauges varies significantly over time, which introduces a further source of uncertainty. On the other hand, observations are provided in a relatively high spatial resolution and can be assumed to be of high quality over data-rich regions such as Europe or large parts of North America. For the spatially lower resolved GPCP product, data from different rainfall sensors have been merged in order to provide a more consistent data source for global (land and ocean) precipitation. Currently, the GPCP includes microwave, infrared, and gauge-based observations of precipitation (e.g., Huffman et al. 2009).

### c. Evapotranspiration

A wide variety of model approaches for estimating evapotranspiration exist, which range from simple empirical ones to complex ones including radiative energy

balance. Many researches have focused on remote sensing approaches, in which the observation of surface temperature and vegetation indices are used for estimation of evapotranspiration. In this study, we have used data from the following:

- Global Land Surface Evaporation: The Amsterdam Methodology (GLEAM; Miralles et al. 2011),
- GLDAS, version 2 (GLDAS2; Rodell et al. 2004b),
- Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim; Berrisford et al. 2009; Dee et al. 2011),
- the land-only version of the Modern-Era Retrospective Analysis for Research and Applications (MERRA-Land; Reichle et al. 2011),
- FLUXNET Multi-Tree Ensemble (FLUXNET MTE; Jung et al. 2009, 2010), and
- Moderate Resolution Imaging Spectroradiometer (MODIS) Global Evapotranspiration Project (MOD16; Mu et al. 2007, 2011).

GLEAM, FLUXNET MTE, and MOD16 are evapotranspiration-only products, that is, they should, in principle, provide the most reliable estimates of global evapotranspiration. GLEAM is based on multisatellite observations and combines a wide range of remotely sensed observations within a Priestley–Taylor-based framework (Miralles et al. 2011). In contrast to this multisatellite approach, MOD16 is based on the Penman–Monteith method and is derived by combining remote sensing data from MODIS and global meteorological data (Mu et al. 2007). The dataset is available in a very high spatial resolution of 1 km. For this study, we use the coarser  $0.5^\circ \times 0.5^\circ$  version, which is also provided by the data center. The FLUXNET MTE dataset provides global evapotranspiration by empirical upscaling of eddy-covariance measurements from the network (Jung et al. 2009). We also use the evapotranspiration from GLDAS2 (i.e., the Noah3.3 realization) and MERRA-Land. A description of these datasets is given in section 3f.

Finally, ERA-Interim is a reanalysis model where a global atmospheric model is assimilated with numerous observations in order to provide reliable estimates of a wide variety of hydrometeorological and hydrological variables (see section 3e). A comparison of some of these datasets is presented in, for example, Mueller et al. (2011).

To get an impression of the level of agreement between the different precipitation and evapotranspiration datasets, the top row in Fig. 6 shows scatterplots between the mean annual and seasonal precipitation from GPCC and the other datasets. For the bottom row, evapotranspiration from MOD16 served as reference.

TABLE 2. River basins that were used in this study, including the respective river gauge, its location, the area of the basin, and the mean annual discharge as provided from GRDC. The abbreviation trib. stands for tributary, and d/s stands for downstream.

	River	Station	Lat	Lon	Area ( $\text{km}^2$ )	Discharge ( $\text{m}^3 \text{s}^{-1}$ )
1	Amazonas	Obidos	1.95°S	55.51°W	4 672 876	176 498
2	Congo	Kinshasa	4.3°S	15.3°E	3 615 546	40 108
3	Mississippi	Vicksburg	32.32°N	90.91°W	2 938 538	20 793
4	Yenisei	Igarka	67.48°N	86.5°E	2 454 961	20 030
5	Lena	Kyusyr	70.7°N	127.65°E	2 417 932	19 261
6	Ob	Salekhard	66.57°N	66.53°E	2 926 321	12 483
7	Tocantins	Tucurui	3.76°S	49.65°W	752 993	10 291
8	Rio Tapajos	Fortaleza	6.05°S	57.64°W	366 843	9 773
9	Mackenzie	Arctic Red River	67.46°N	133.75°W	1 666 073	9 707
10	St. Lawrence	Lasalle	45.42°N	73.62°W	943 769	8 106
11	Rio Xingu	Altamira	3.22°S	52.21°W	445 289	7 982
12	Volga	Volgograd Power Plant	48.81°N	44.59°E	1 345 070	7 778
13	Danube	Ceatal Izmail	45.22°N	28.72°E	771 277	7 091
14	Yukon	Pilot Station, Alaska	61.93°N	162.88°W	819 635	6 403
15	Columbia	Beaver Army Terminal	46.18°N	123.18°W	662 542	6 102
16	Pechora	Oksino	67.6°N	52.2°E	304 670	5 139
17	Nelson	Long Spruce Generating Station	56.4°N	94.37°W	1 126 480	3 822
18	Kolyma	Kolymskaya	68.73°N	158.72°E	421 802	3 546
19	Severnaya Dvina (north)	Ust-Pinega	64.15°N	41.92°E	330 709	3 453
20	Fraser	Hope	49.38°N	121.45°W	228 874	2 533
21	Neva	Novosaratovka	59.84°N	30.53°E	225 651	2 473
22	Rhine	Rees	51.75°N	6.4°E	169 706	2 096
23	Churchill, Fleuve (laboratory)	Above Muskrat Falls	53.25°N	60.79°W	86 658	1 736
24	Stinkine	Near Wrangell	56.71°N	132.13°W	52 894	1 562
25	Yana	Ubileynaya	70.75°N	136.08°E	220 949	1 300
26	Susquehanna	Conowingo	39.66°N	76.18°W	69 694	1 288
27	Olenek	7.5 km d/s of mouth of Pur	72.12°N	123.22°E	199 723	1 255
28	Rio Jari	Sao Francisco	0.57°S	52.57°W	46 355	1 208
29	Albany	Near Hat Island	51.33°N	83.84°W	97 926	1 154
30	Kuskokwim	Crooked Creek	61.87°N	158.1°W	82 074	1 145
31	Rio Araguari	Porto Platon	0.71°N	51.44°W	21 630	1 050
32	Thelon	Below outlet of Schultz Lake	64.78°N	97.05°W	171 346	912
33	Skeena	Usk	54.63°N	128.43°W	31 555	886
34	Tombigbee	Coffeeville	31.76°N	88.13°W	51 622	854
35	Paraiba do Sul	Campos–Ponte Municipal	21.75°S	41.3°W	60 146	728
36	Hayes (trib. Hudson Bay)	Below Goods River	56.43°N	92.79°W	97 802	707
37	Mezen	Malonisogorskaya	65.03°N	45.62°E	54 125	693
38	Don	Razdorskaya	47.54°N	40.65°E	378 180	669
39	Moose (trib. Hudson Bay)	Above Moose River	50.74°N	81.46°W	91 157	661
40	Peel (trib. Mackenzie)	Above Fort McPherson	67.25°N	134.88°W	63 040	661
41	Glama	Langnes	59.61°N	11.12°E	38 534	658
42	Elbe	Neu Darchau	53.23°N	10.89°E	134 037	645
43	Apalachicola	Near Sumatra	29.95°N	85.02°W	57 546	633
44	Rio Capim	Badajos	2.51°S	47.77°W	37 004	615
45	Vuoksi	Tainionkoski	61.22°N	28.78°E	66 393	612
46	Rio Parnaiba	Luzilandia	3.45°S	42.37°W	297 049	606
47	Connecticut	Thompsonville	41.99°N	72.61°W	24 596	590
48	Anabar	Saskylakh	71.98°N	114.06°E	79 786	578
49	Kemijoki	Isohaara	65.78°N	24.55°E	53 763	543
50	Winisk	Below Ashweig River trib.	54.52°N	87.23°W	84 526	535
51	Goeta Aelv	Vargoens KRV	58.36°N	12.38°E	48 265	530
52	Daugava	Daugavpils	55.88°N	26.53°E	61 576	501
53	Lule	Bodens KRV	65.81°N	21.67°E	22 885	497
54	Churchill	Above Red Head Rapids	58.12°N	94.63°W	299 391	489
55	Rio Gurupi	Alto Bonito	1.8°S	46.32°W	33 933	473
56	Angermanaelven	Solleftea KRV	63.17°N	17.27°E	33 376	467
57	Nemunas–Neman	Smalininkai	55.08°N	22.58°E	89 806	458

TABLE 2. (*Continued*)

	River	Station	Lat	Lon	Area (km <sup>2</sup> )	Discharge (m <sup>3</sup> s <sup>-1</sup> )
58	Klamath	Near Klamath	41.51°N	124.0°W	36 896	451
59	Kazan	Above Kazan Falls	63.65°N	95.85°W	40 945	450
60	St. Johns	Jacksonville	30.32°N	81.67°W	27 007	427
61	Seal	Below Great Island	58.89°N	96.28°W	49 662	397
62	Potomac	Near Washington, D.C.	38.95°N	77.13°W	31 151	380
63	Daly	Mount Nancar	13.83°S	130.74°E	44 883	357
64	Altamaha	Doctortown	31.66°N	81.83°W	33 706	346
65	Weser	Intschede	52.96°N	9.13°E	36 293	311
66	Kymi	Anjala	60.7°N	26.82°E	33 195	281
67	Rio Jequitinhonha	Itapebi	15.95°S	39.52°W	65 067	274
68	Bolshoy Anyuy	Konstantinovo	68.15°N	161.17°E	50 025	272
69	Savannah	Near Clyo	32.53°N	81.27°W	23 055	268
70	Oulujoki	Merikoski	65.02°N	25.52°E	26 517	259
71	Mitchell	Koolatah	15.95°S	142.38°E	44 381	247
72	Brazos	Richmond	29.58°N	95.76°W	106 914	243
73	Rio Pindaré	Pindaré-Mirim	3.66°S	45.46°W	33 895	237
74	Pee Dee	PeeDee	34.2°N	79.55°W	22 686	236
75	Suwannee	Near Bell	29.79°N	82.92°W	21 279	233
76	Trinity (Texas)	Romayor	30.43°N	94.85°W	44 365	225
77	Thlewiaza	Above outlet Sealhole Lake	60.79°N	98.78°W	53 705	219
78	Rio Itapécuru	Cantanhede	3.63°S	44.38°W	58 451	213
79	Sabine	Near Ruliff	30.3°N	93.74°W	23 626	195
80	Victoria	Coolibah Homestead	15.56°S	130.96°E	44 380	191
81	Fitzroy (west Australia)	The Gap	23.09°S	150.11°E	126 986	190
82	Roper	Red Rock	14.7°S	134.42°E	50 687	188
83	Fitzroy (east Australia)	Fitzroy Crossing	18.21°S	125.58°E	44 147	144
84	Orange	Vioolsdrif	28.76°S	17.72°E	828 475	104
85	Rio Mearim	Bacabal	4.22°S	44.77°W	27 688	102
86	San Joaquin	Near Vernalis	37.68°N	121.27°W	29 459	97
87	Murray	Overland Corner	34.17°S	140.28°E	1 022 767	88
88	Tugela	Mandini	29.14°S	31.39°E	29 848	76
89	Rio Jaguaripe	Peixe Gordo	5.23°S	38.2°W	55 288	56
90	Gilbert	Rockfields	18.2°S	142.88°E	32 248	53
91	Rio Prado	Mascote	15.56°S	39.31°W	38 733	52
92	De Gray	Coolenar Pool	20.31°S	119.25°E	51 695	33
93	Colorado (Pacific)	At Yuma	32.73°N	114.63°W	636 508	29
94	Rio Itapicuru	Usina Altamira	11.74°S	37.8°W	36 421	22
95	Nueces	Calallen	27.88°N	97.63°W	43 371	16
96	Groot-Vis	Matolemas Location Outspan	33.24°S	27.0°E	28 741	11

#### d. Water storage changes

Water storage changes are computed from GRACE observations of the temporal gravity field changes. Eight years of GRACE (release 5) data have been used in this study, and they come from two data processing centers: GeoForschungsZentrum (GFZ; Dahle et al. 2013) and Center for Space Research, The University of Texas at Austin (CSR; Bettadpur 2012). The data are provided as spherical harmonic coefficients of the monthly gravity field snapshots. Wahr et al. (1998) provide the mathematics (section 2a) to invert the monthly snapshots to surface mass anomalies  $\Delta M$ . According to them, those surface mass anomalies represent anomalies in water storage over the continental landmasses. For this reason, we will denote the water storage changes estimated from

GRACE as  $dM/dt$ . The GRACE data as such are plagued by a lot of errors, which necessitates further processing prior to their usage (see appendix for details). The postprocessed  $\Delta M$  are taken on a  $0.5^\circ \times 0.5^\circ$  grid and then aggregated over catchments to derive catchment-specific time series. The time series are then numerically differentiated using three-point central differences to arrive at the required flux quantity, surface mass changes  $dM/dt$ .

In this study, we use three different GRACE-based datasets in order to gain information on differences between products from different data providers or filtering methods. Therefore, the GFZ data have been filtered with both a regularization filter (GFZ<sub>reg</sub>; see appendix for details) as well as the widely used decorrelation (or destriping) filter in combination with a

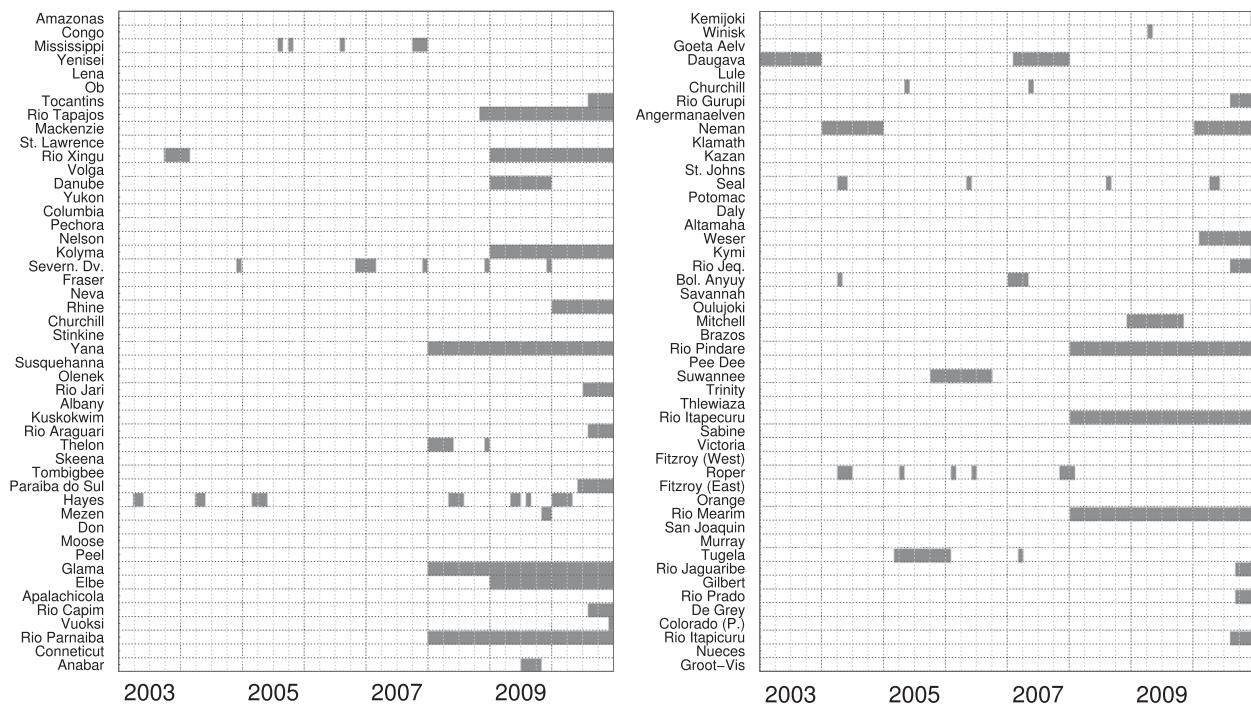


FIG. 3. Overview of the amount of runoff observations for every catchment. Each box corresponds to a single month of a specific catchment. White boxes indicate that observations are available for that month; while gray boxes depict missing data.

Gaussian filter ( $GFZ_{dstr}$ ). The data from the CSR have been filtered with the regularization filter only ( $CSR_{reg}$ ).

#### e. Moisture flux divergence

Vertically integrated moisture flux divergences can be obtained from global atmospheric reanalysis models. These models are combinations of a numerical model and an analysis scheme where the numerical model predicts the state of the earth's atmosphere, oceans, land surface, and upper soil layers. The analysis scheme, on the other hand, assimilates a large number of observations into the numerical model. Because of the very nature of assimilation, the combination of an analysis scheme forces the estimates to stay both in reasonable physical and numerical bounds, which, on the flip side, disturbs the model's interior mass budgets (Trenberth et al. 2011; Trenberth and Fasullo 2013).

The present-day reanalyses models assimilate nearly  $7 \times 10^6$ – $9 \times 10^6$  observations every 6–12 h, which goes to show their data-intensive nature. These reanalyses usually cover a period of multiple decades and provide estimates of the most important hydrological, hydrometeorological, and atmospheric quantities globally. Kalnay et al. [1996; see also Kistler et al. (2001)] categorize the gridded variables from reanalyses into three distinct classes according to the influence of the assimilated observations on them. The type A variables are

mainly influenced by observations (e.g., upper-air temperatures or horizontal winds). Type B variables are influenced by both the observations as well as the model (e.g., surface and 2-m temperatures), while type C variables are solely derived from the model (e.g., precipitation and runoff).

The vertically integrated moisture flux divergences are derived from horizontal wind fields—a type A variable—and specific humidity—a type B variable [see Eq. (4)]. Therefore, the moisture flux divergence can be assumed to be more reliable and stable than, for example, modeled evapotranspiration. Here, data from three widely used state-of-the-art reanalysis models have been used: ERA-Interim; MERRA from the Goddard Space Flight Center (GSFC), National Aeronautics and Space Administration (NASA; Rienecker et al. 2011); and Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) from the National Centers for Environmental Prediction (NCEP), National Oceanic and Atmospheric Administration (NOAA).

#### f. Land surface models

In this study, we also used monthly runoff and evapotranspiration from five different land surface models driven by the GLDAS (Rodell et al. 2004b) and MERRA-Land (Reichle et al. 2011). In contrast to traditional hydrological or land surface models, these estimates are strongly constrained by observations (similar to

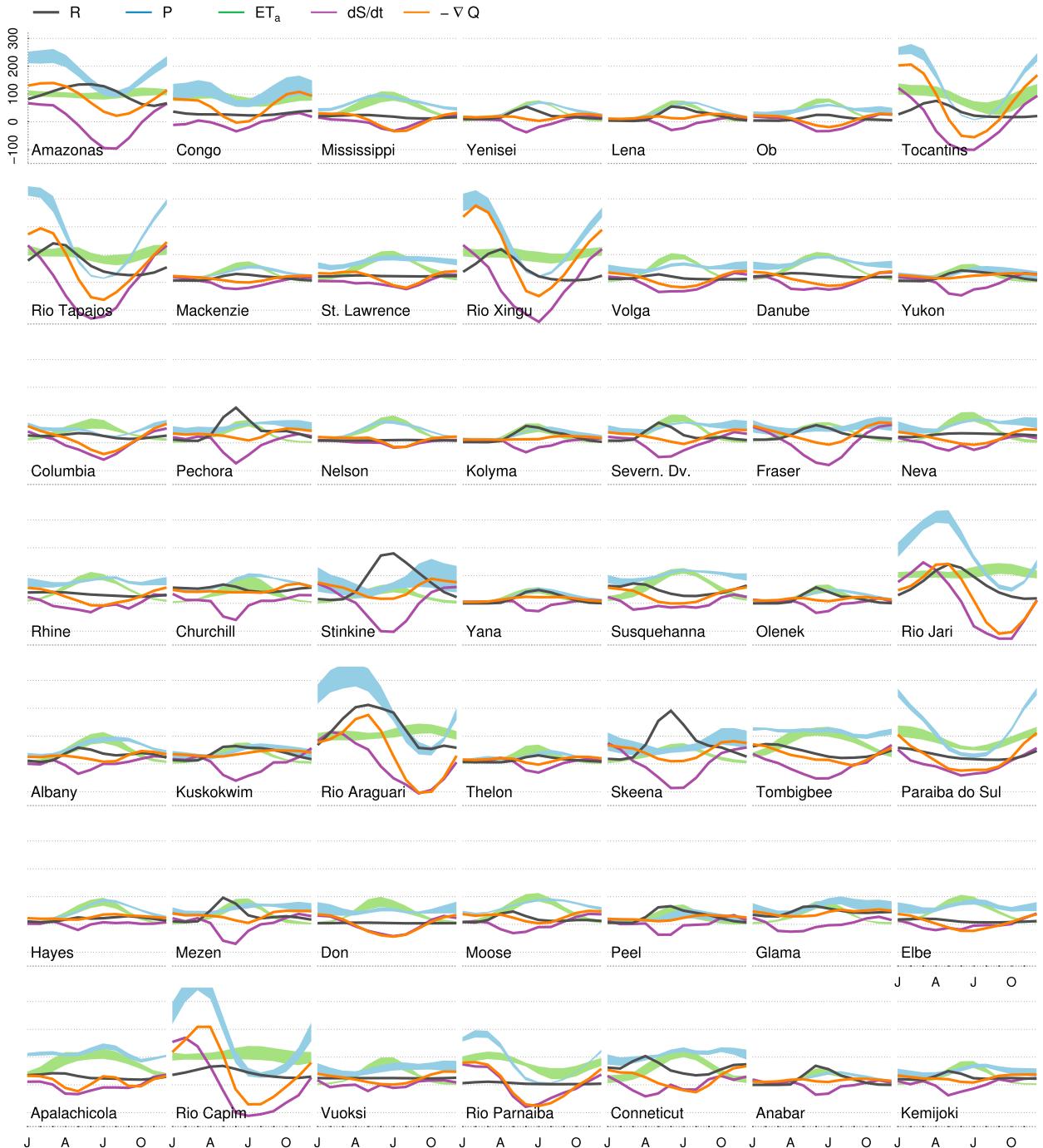


FIG. 4. Mean annual cycle of the different budget quantities for catchments 1–49. The precipitation bands are derived from the maximum and minimum estimates of GPCC, GPCP, CRU, DEL, and CPC while the evapotranspiration bands are given by GLDAS, ERA-Interim, GLEAM, MERRA-Land, MOD16, and FLUXNET MTE. Water storage changes, moisture flux convergences (i.e., negative divergences), and runoff are taken from the CSR<sub>reg</sub>, ERA-Interim, and GRDC dataset, respectively.

atmospheric reanalyses) in order to provide more reliable estimates of hydrological variables.

MERRA-Land is an offline (without atmosphere) rerun of a revised version of MERRA's land component,

which is forced by the atmospheric reanalysis near-surface fields (air temperature, wind, and humidity), surface pressure, radiation, and observed precipitation based on the CPC precipitation product (see section 3b). MERRA-Land

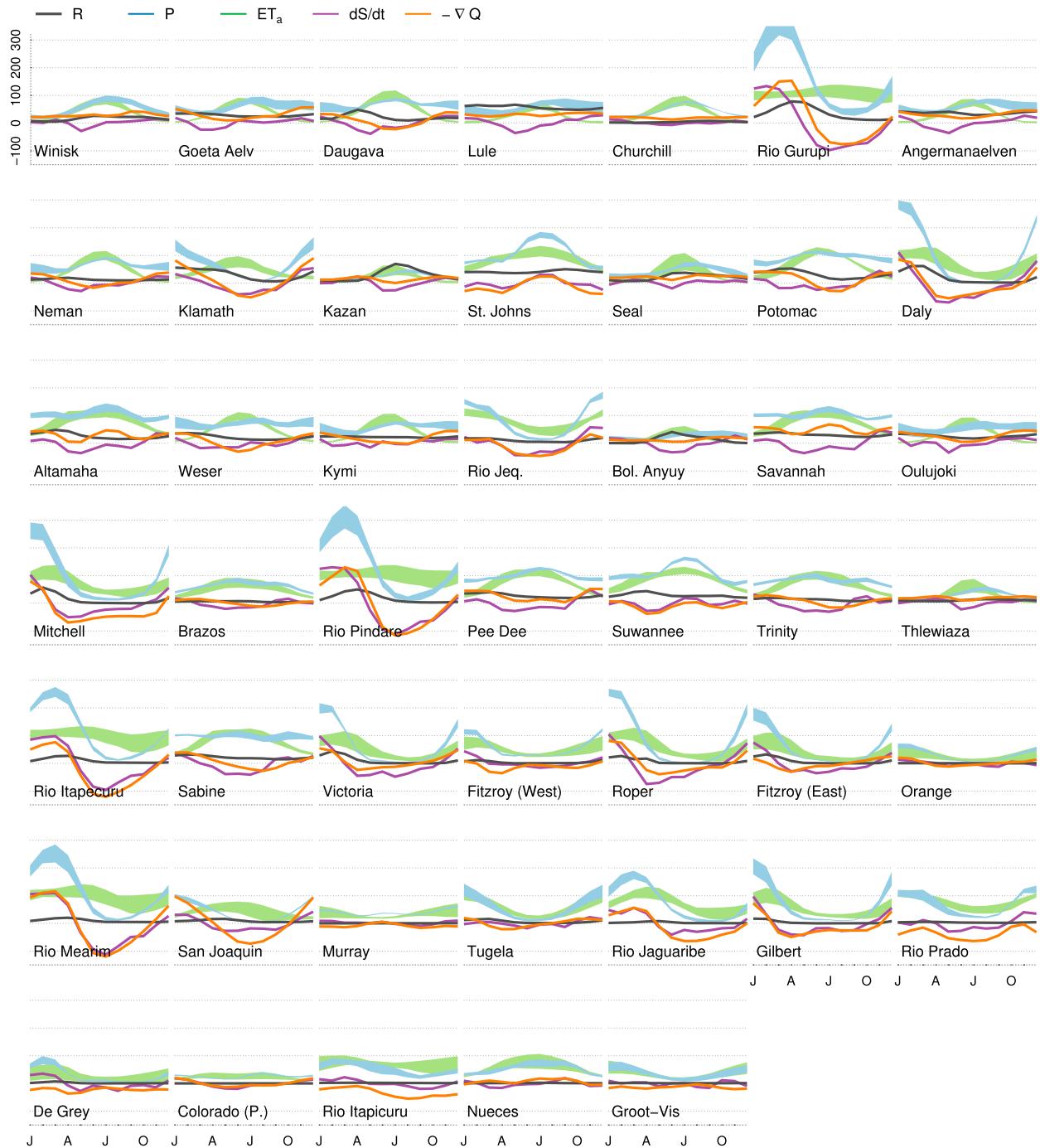


FIG. 5. As in Fig. 4, but for catchments 50–96.

thus provides a supplemental and improved set of land surface hydrological fields (Reichle et al. 2011).

GLDAS can be seen as a data assimilation framework, where different land surface models are combined with satellite- and ground-based observations in order to improve and constrain the simulated hydrological processes. In particular, the models currently driven by

the GLDAS are the Noah land surface model (LSM; Chen et al. 1996; Koren et al. 1999; Ek et al. 2003), the Common Land Model (CLM; Bonan 1998; Dai et al. 2003), Mosaic (Koster and Suarez 1996), and the Variable Infiltration Capacity (VIC) model (Liang et al. 1994). These four realizations are usually referred to as GLDAS1 and share the same forcing datasets:

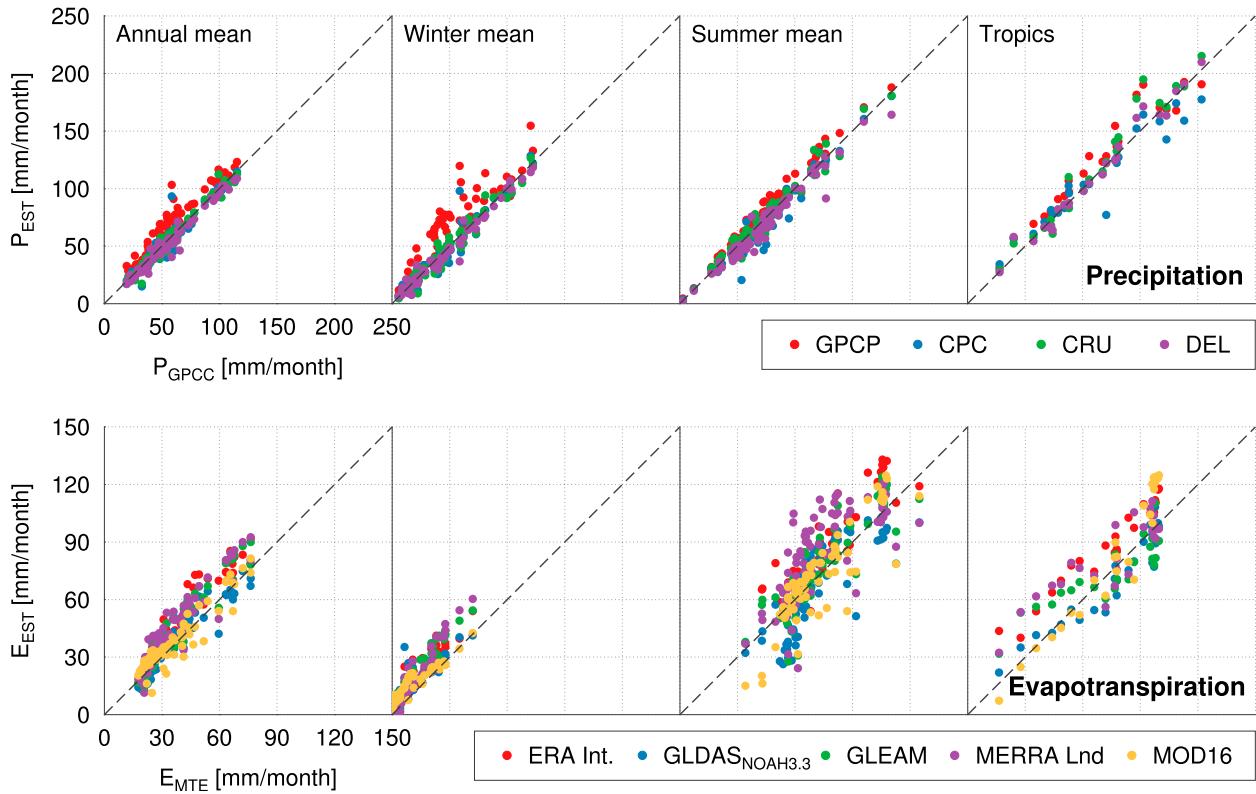


FIG. 6. Scatterplots between the different datasets of (top) precipitation and (bottom) evapotranspiration over the study regions with respect to GPCC (precipitation) and FLUXNET MTE (evapotranspiration). The first three columns show three different time scales: long-term annual, winter, and summer means. The tropical catchments have been removed from these scatterplots. The winter mean was derived from the average of December–February (DJF) for the Northern Hemisphere catchments and June–August (JJA) for Southern Hemisphere catchments. Similarly, depending on the hemispherical location of the catchments, their respective summer months were used for computing the summer mean. The fourth column shows the long-term mean for the 25 tropical catchments.

- meteorological data from the Global Data Assimilation System (GDAS; Derber et al. 1991),
- downscaled precipitation from the CPC's Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997), and
- satellite-derived downward radiation from the Air Force Weather Agency Agricultural Meteorology Modeling System (AFWA AGRMET; Kopp and Kiese 1996).

An overview of the different models and model forcings for GLDAS1 is given in Rodell et al. (2004b). The first GLDAS2 dataset was recently released and provides hydrological variables for the period 1948–2010. This version is currently based on the Noah3.3 LSM and uses forcing data from the Princeton meteorological forcing data (Sheffield et al. 2006) only. In this study, we use the total runoff estimates from the four GLDAS1 and the single GLDAS2 realizations.

It should be noted that these models do not consider a routing scheme for estimating the runoff (R. H.

Reichle 2013, personal communication; M. Rodell 2012, personal communication). Therefore, integrating fields of surface and subsurface runoff over a large area on monthly or shorter time scales might lead to inaccuracies in, for example, peak runoff as varying flow directions and velocities (due to the topography of the basin) are not considered. As a consequence, the impact of such routing schemes reduces when considering longer time scales (seasonal and annual). Therefore, for this study, we have used the modeled monthly runoff as provided by the data centers without applying any routing scheme in order to analyze the quality of these estimates on monthly and longer time scales.

## 4. Results

### a. Performance metrics

We compare the time series of runoff estimates from GRACE-based approaches with the observed runoff. To get a holistic understanding, we also compare them

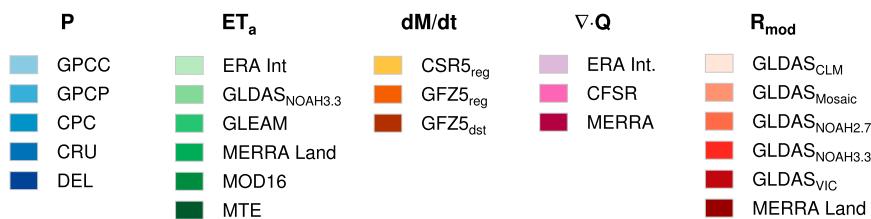


FIG. 7. Legend for the color codes used in Figs. 8, 11, and 12.

with the estimates from hydrological models and the runoff–precipitation ratio method. Such an analysis should provide us a balanced view on the utility of the GRACE-based estimates in terms of already available global-scale alternatives. In the current setup, we use the following three metrics to carry out our evaluation: correlation coefficient, percentage bias (PBIAS; Gupta et al. 1999) and Nash–Sutcliffe efficiency (NSE) coefficient (Nash and Sutcliffe 1970).

The correlation coefficient quantifies the level of common information content between observed and estimated runoff, although, as is well known, it is insensitive to the amplitude and the mean value. For the correlation analysis, we only show correlations above 0.25, which is approximately the 5% significance level with respect to 60 samples (i.e., 5 yr of data). The PBIAS,

$$\text{PBIAS} = \frac{\sum_{t=1}^T (R_{t,\text{obs}} - R_{t,\text{est}})}{\sum_{t=1}^T R_{t,\text{obs}}} = 1 - \frac{\bar{R}_{\text{est}}}{\bar{R}_{\text{obs}}}, \quad (12)$$

on the other hand, gives precisely this information about the relative difference between the long-term mean of the observations  $\bar{R}_{\text{obs}}$  and estimates  $\bar{R}_{\text{est}}$ . Finally, correlation and PBIAS are summarized by the NSE coefficient:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^T (R_{t,\text{obs}} - R_{t,\text{est}})^2}{\sum_{t=1}^T (R_{t,\text{obs}} - \bar{R}_{\text{obs}})^2}. \quad (13)$$

The NSE is a conservative quantity, and it is highly sensitive to the overall agreement between the observations and estimates: phase, amplitude, and mean. It ranges between  $-\infty$  and 1, where a value of 1 indicates that the estimated and observed runoff agree in terms of all the three variables mentioned above. Further, the NSE can be treated as a relative imbalance at monthly scales for the  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  methods as it involves the scaled mean of the monthly squared error.

### b. Notes on the interpretation of the results

The sheer number of combinations, due to the choice of multiple datasets for each of the budget components, forms an ensemble that helps to provide a broad overview of the approach in hand, but also makes the assessment a little cumbersome. The contributing datasets are shown in Fig. 7 together with a color code. The performance of the different approaches and data combinations at the monthly time scale is summarized in Figs. 8, 11, and 12. The upper part of these figures collects the values of the performance metrics into a matrix, where each cell denotes the individual performance values for each catchment and each estimate separately. In other words, each row of the matrix provides an overview of the performance of different combinations for a particular catchment, and each column provides an overview for a particular combination. The lower part summarizes each column of the matrix by sorting the performance metric values within prescribed intervals. The color scales in these matrices have been chosen in such a manner that darker colors indicate better performance and lighter colors indicate poorer performance. The datasets of a single combination can be identified with the aid of the color code and the legend (Fig. 7).

### c. Runoff from terrestrial water budget: $R_{\text{hyd}}$

The runoff estimates from the terrestrial water budget equation is the largest group of the four different runoff estimates compared here. This is obvious because the runoff estimates from the terrestrial water balance equation require three inputs: precipitation, evapotranspiration, and water storage change. The performance metric PBIAS (Fig. 8) shows a random behavior with no data combination providing an outright good performance consistently over all the catchments. However, for some catchments (Amazon, Rio Tapajos, and Rio Jari)  $R_{\text{hyd}}$  provides consistently good PBIAS values ( $\leq 25\%$ ). Also, for most of the catchments ( $\approx 60$ ) there is at least one data combination that provides a PBIAS value of nearly zero (see Fig. 9). In other words, there is at least one data combination that provides a reliable mean runoff value for most of the catchments. In all this, the MERRA-Land dataset performs

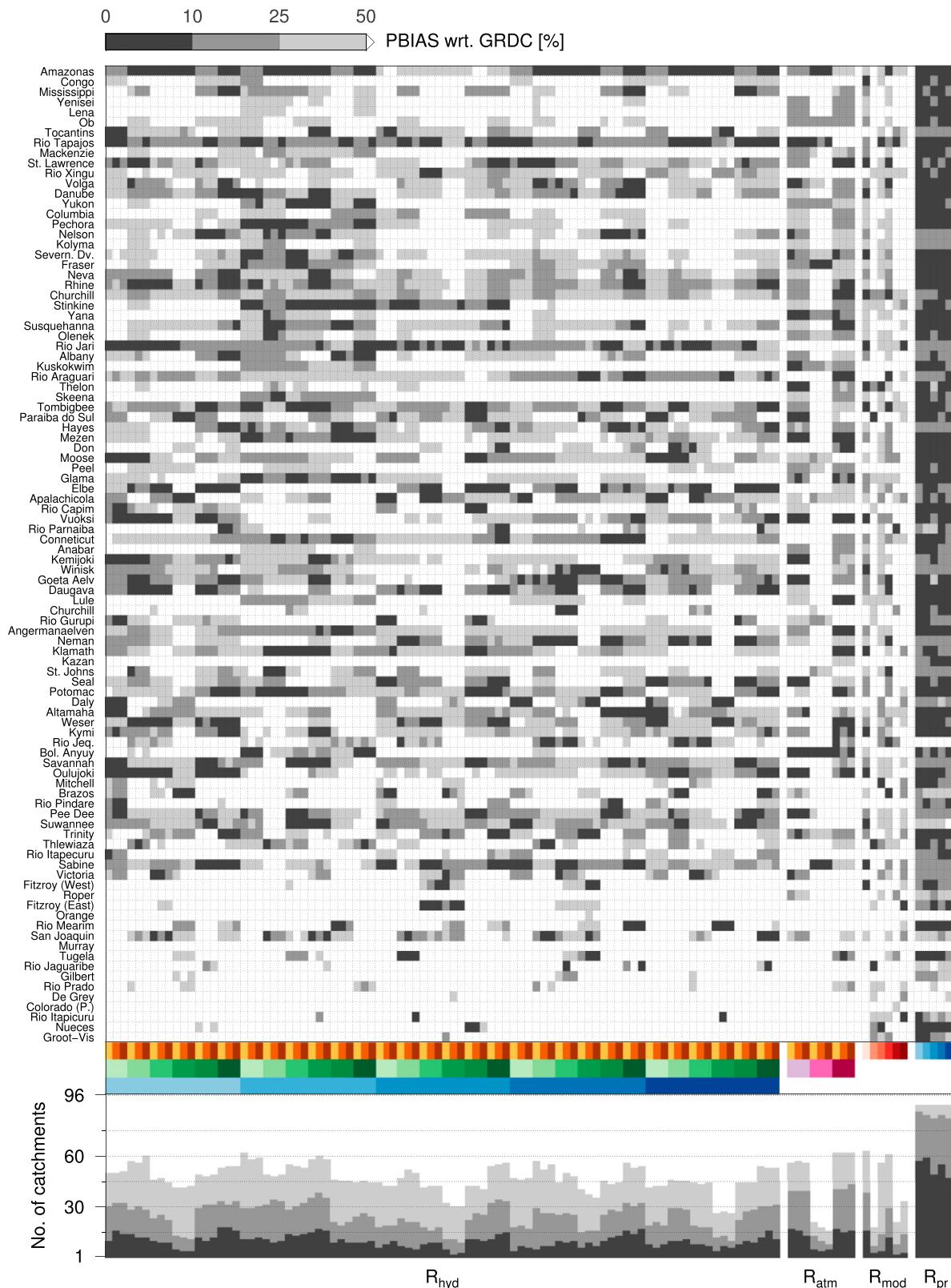


FIG. 8. PBIAS for the 96 study regions between observed and estimated runoff. A guideline on how to read the figure is given in section 4b while the color codes are given in Fig. 7.

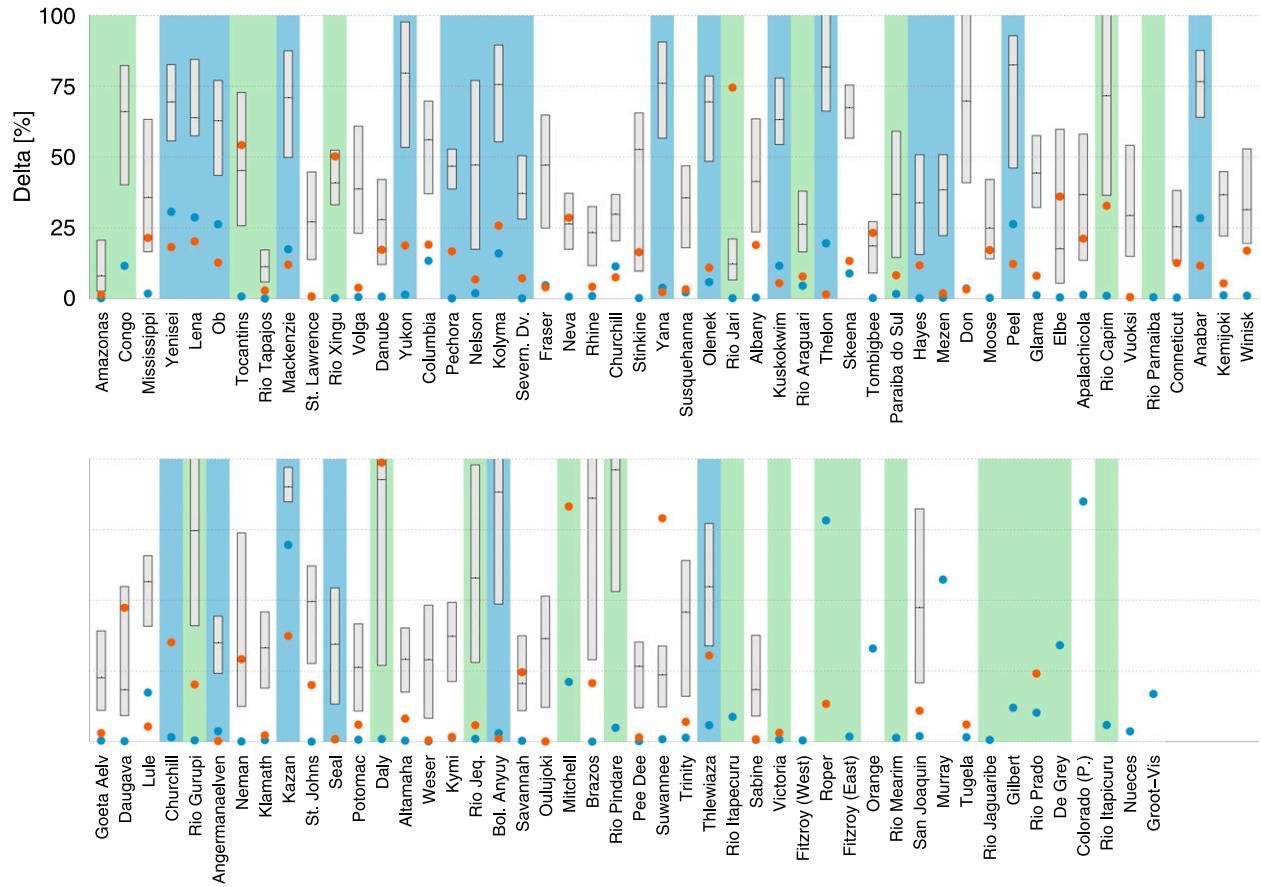


FIG. 9. Box plot showing the relative water cycle imbalance with respect to the observed mean annual runoff. The lower and upper bounds of the gray boxes depict the 25th and 75th percentile of the 90 ensemble members of  $R_{\text{hyd}}$ ; while the line within the boxes shows the median imbalance. The blue (red) dots are the minimum imbalances from  $R_{\text{hyd}}$  ( $R_{\text{atm}}$ ). The background colors indicate Arctic (blue) and tropical (green) catchments. The Arctic catchments have been chosen according to the entries in the ArcticRIMS database. For the tropical regimes, we have selected the basins within 23.5°S–23.5°N latitude.

consistently poorly except when it is combined with GPCP (see stacked bars in Fig. 8). The possible reason could be that GPCP is the wettest of all the precipitation datasets and MERRA-Land is the wettest of all the evapotranspiration datasets (see Fig. 6), and therefore their biases get canceled out. But when combined with the other datasets, the bias in MERRA-Land becomes visible. Similarly, GPCP does not perform very well when combined with MOD16 and FLUXNET MTE, while other precipitation datasets perform very well with them. The other interesting feature of the PBIAS matrix is the consistently poor performance in the Arctic catchments (Yenisei, Lena, Ob, Yukon, Thelon, Anabar, Kazan, Mackenzie, Kolyma, Churchill, Yana, Luleaelven, Olenek, Kuskokwim, and Peel), where there is a substantial underestimation of the runoff values (also see Fig. 10). However, we also observe that in these catchments the data combinations that involve GPCP are relatively well behaved. This

better performance might be due to the gauge undercatch correction performed in the GPCP dataset, which provides higher values for the precipitation values in the Arctic catchments.

In contrast to PBIAS, the correlation matrix (Fig. 11) though random to an extent, shows a homogeneous picture. In nearly 50% of the catchments, the correlation values are consistently more than 0.5 irrespective of the data combination, which demonstrates that most of the data combinations capture the dynamics of the catchment runoff to a large extent. Further, the Arctic and tropical catchments provide strong correlations with many data combinations, having a correlation upward of 0.75. Some catchments, however, show strong dependency on the data combination (e.g., Mackenzie and Bolshoy Anyuy provide very good correlation when using data combinations involving GLDAS<sub>Noah3.3</sub>). Similarly, the tropical catchments are strongly influenced by the choice of precipitation

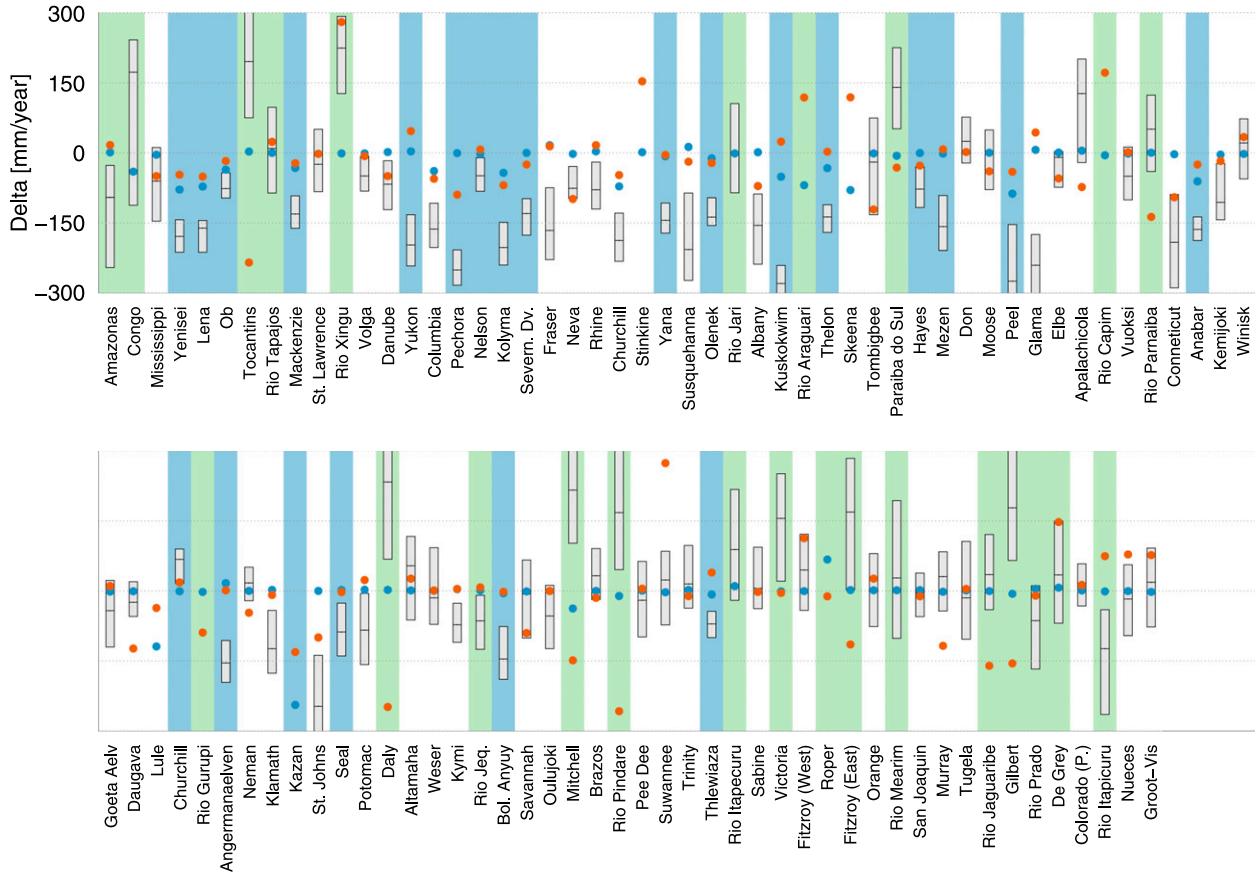


FIG. 10. As in Fig. 9, but for the absolute water cycle imbalance with respect to the observed mean annual runoff.

dataset, where the best performance is given by the use of GPCC and GPCP. As in the case of PBIAS, the MERRA-Land dataset produces the lowest correlation values. Lowest correlations are also found in catchments (e.g., St. Lawrence, Neva, Don, Vuoksi, San Joaquin, Orange, Murray, and Colorado) where the runoff does not have a clear seasonal cycle and is constant throughout the year. These catchments are affected by the biases and errors in the datasets, which produce short-term artifacts, thus reducing the correlation with observed runoff data (see supplemental material). An interesting aspect of Fig. 11 is the influence of the choice of GRACE dataset in the Arctic catchments (e.g., Yenisei, Ob, Pechora, Yana, Anabar, Stinkine, and Peel). The reason for the influence of the water storage dataset is that there is a clear time shift between precipitation and evapotranspiration in the Arctic catchments (see supplemental material), and therefore, the water storage changes have an important role to play in the runoff estimation. Nevertheless, it should also be noted that there is no single GRACE dataset that consistently performs well in all of the Arctic catchments. This is also corroborated by the histograms in Fig. 13.

NSE, as mentioned before, summarizes the PBIAS and correlation metrics, where a data combination should have less bias, and also must capture the dynamics of the catchment runoff. It is clearly evident that no data combination fulfills the twin requirements over all the catchments (see Fig. 12). The Amazon is the only catchment that provides a value of more than 0.75 and also only for select combinations. For example, the combinations involving CPC do not even produce a positive NSE value. Some reasonable (positive) values are provided by Rio Tapajos, Rio Xingu, Pechora, Severnaya Dvina, Stinkine, Tombigbee, Mezen, and Anabar over a variety of data combinations. Interestingly, many of these catchments are quite small (see Table 1) and are considered indiscernible by the GRACE mission. In spite of this optimism, Fig. 12 clearly demonstrates that the current status of global hydrological datasets does not allow a reasonable estimation of runoff globally. The main weakness of these datasets is the biases in the individual datasets, which limit the ability of the terrestrial water budget equation to estimate runoff.

To study any missing patterns in the matrix plots, we also study the impact of a particular dataset in all the

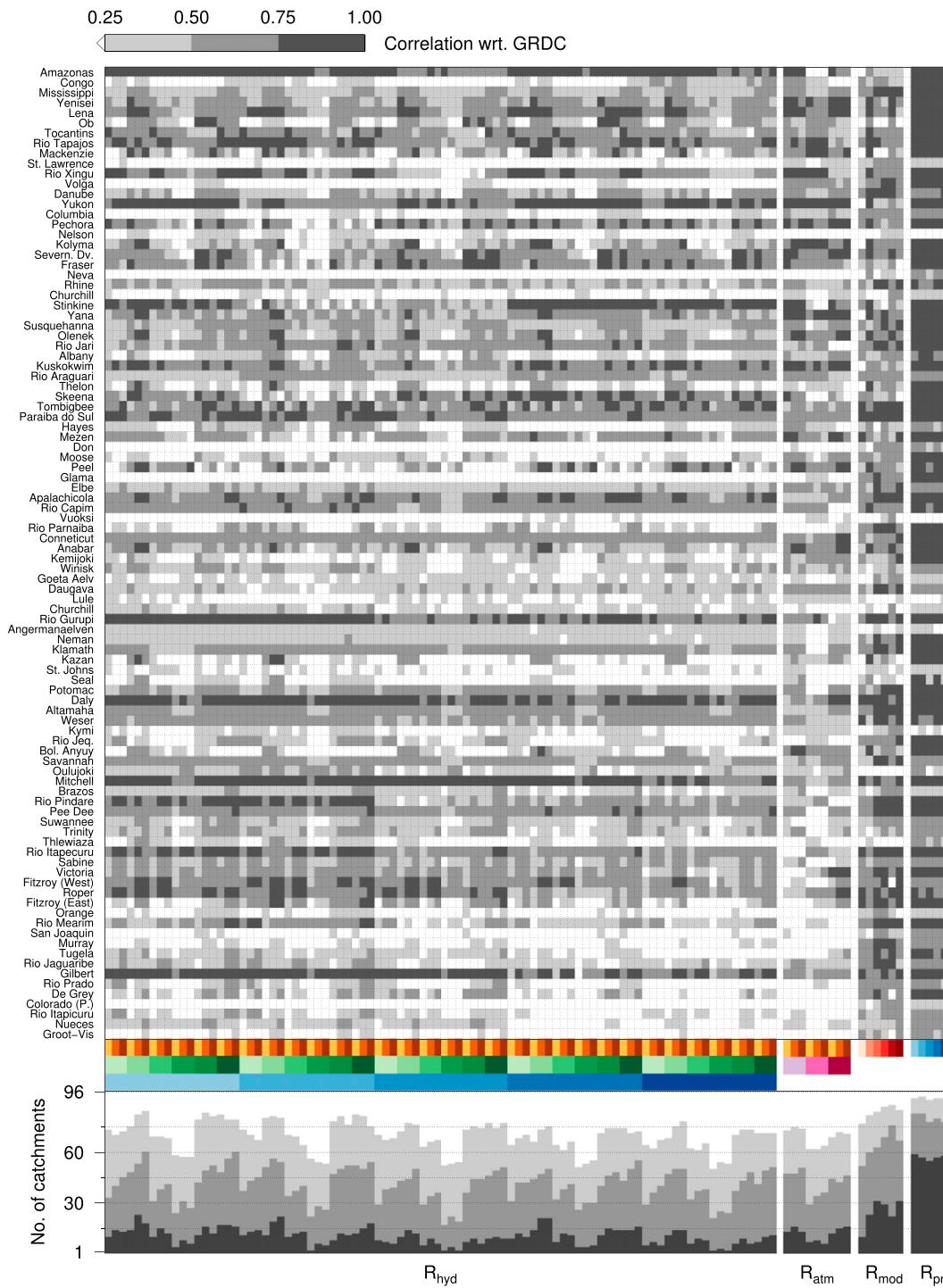


FIG. 11. As in Fig. 8, but for the correlation.

data combinations in which it was involved. This was carried out via histogram analysis as shown in Fig. 13. The histogram clearly shows that for all the datasets involved in the  $R_{\text{hyd}}$  approach, nearly 40% of the dataset provides a correlation of more than 0.5, but only 30% of the datasets have a relative bias of less than 25%. This

again confirms that the major problem of the  $R_{\text{hyd}}$  method lies in the biases in the individual datasets. Further, these histograms also indicate that CPC precipitation and MERRA-Land evapotranspiration do not provide good estimates as they consistently perform poorly in all three metrics. The GRACE datasets differ

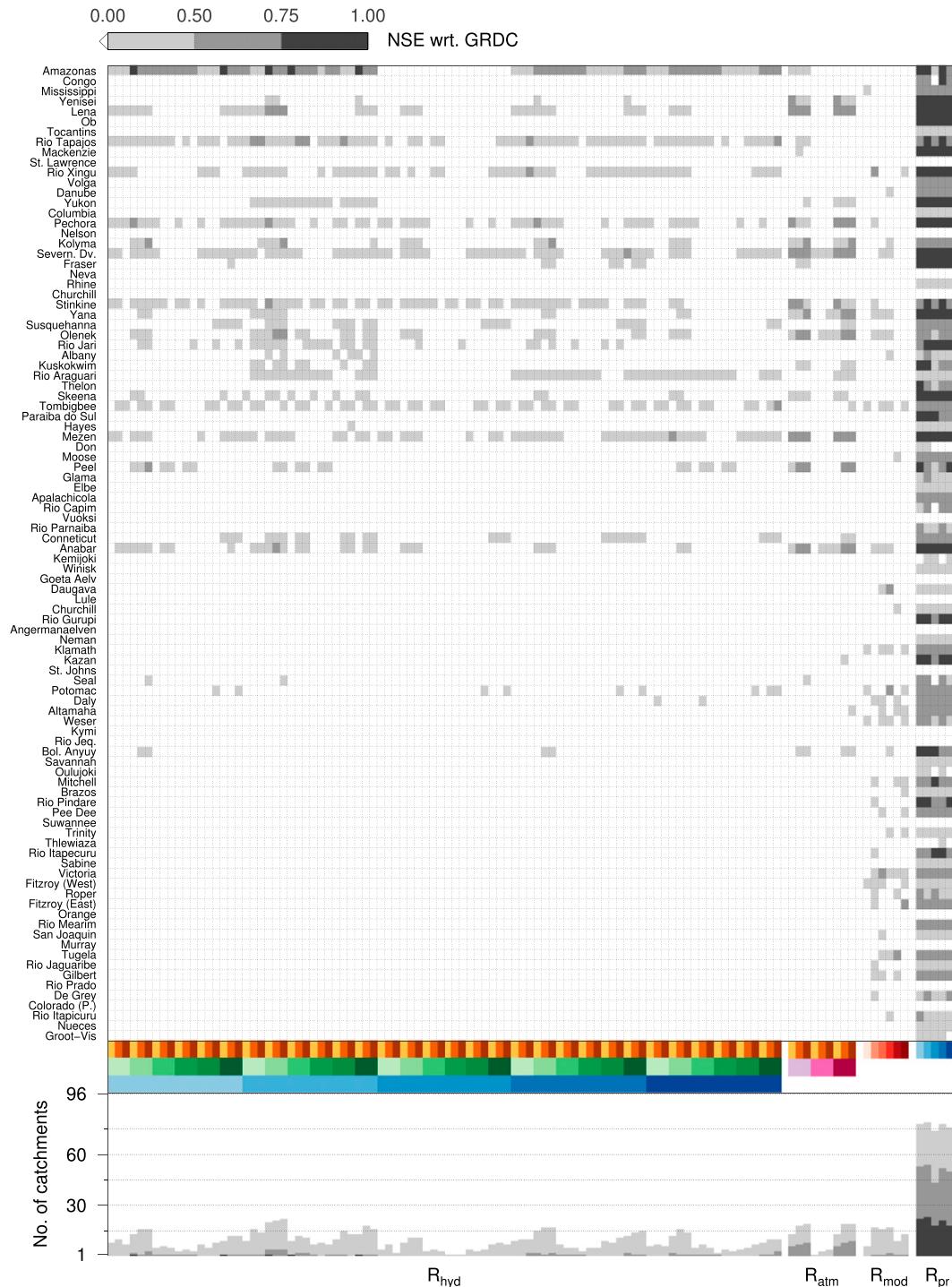


FIG. 12. As in Fig. 8, but for the NSE coefficient.

predominantly in the way they help capture the dynamics, but not otherwise. The biggest difference between the datasets of one variable is seen in evapotranspiration, where GLDAS<sub>Noah3.3</sub>, MOD16, and FLUXNET MTE perform consistently well in all three metrics.

#### d. Runoff from atmospheric water budget: $R_{\text{atm}}$

The  $R_{\text{atm}}$  approach uses only two data sources, and therefore, we can expect a better performance when compared with  $R_{\text{hyd}}$  as there is one less source of

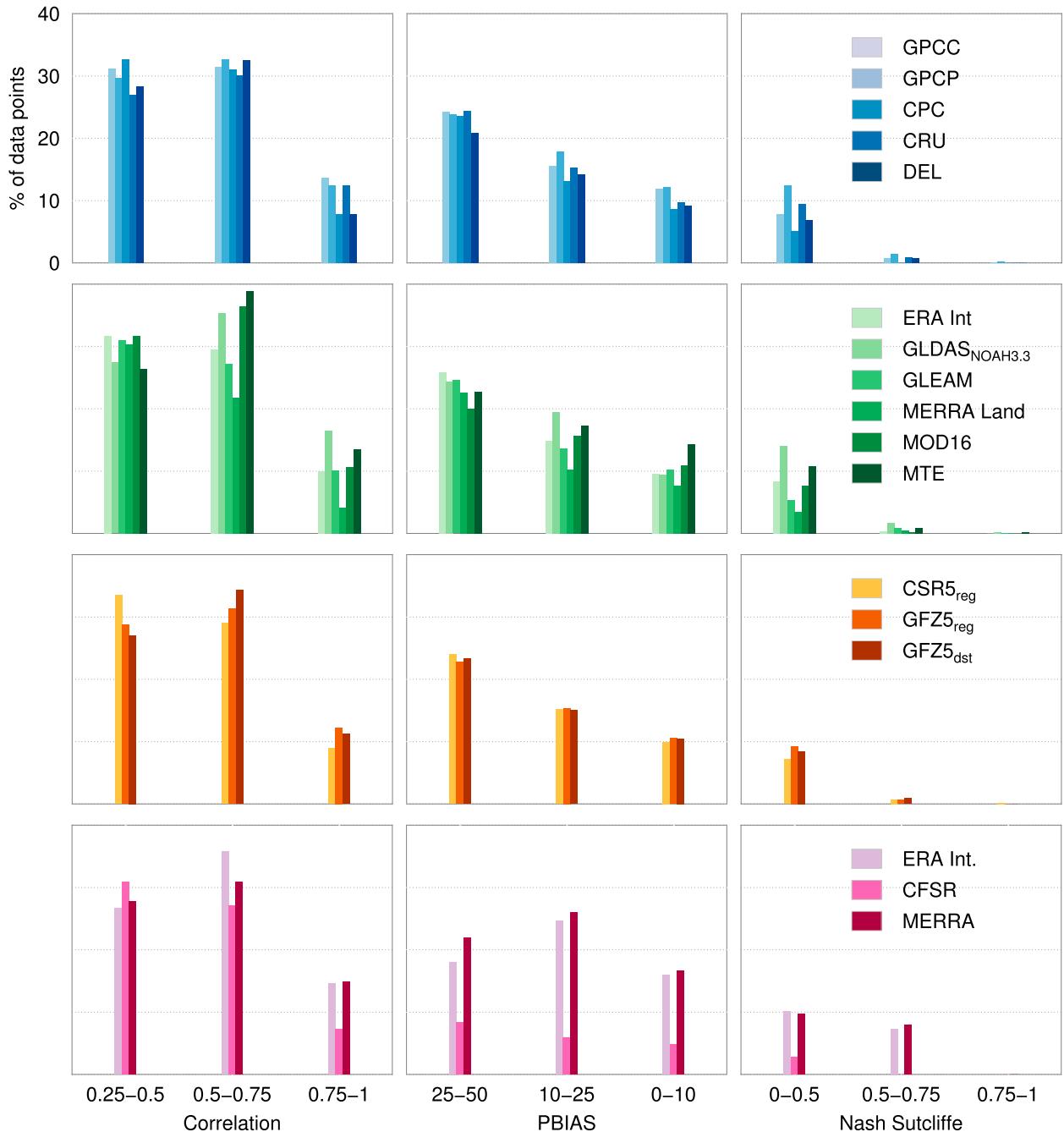


FIG. 13. Histograms showing the percentage of data points (i.e., all combinations and all catchments using a specific dataset) within the intervals from Figs. 8, 11, and 12. They thus serve as a summary of the matrix plots and show the individual performance of every dataset. The plot shows separate histograms for (from top to bottom) precipitation (1728 data points), evapotranspiration (1440 data points), water storage changes (2880 data points), and moisture flux divergences (288 data points). The number of data points is the product of the number of catchments and the number of combinations using a single dataset (e.g., 96 catchments  $\times$  18 combinations using GPCC precipitation = 1728 data points).

potential errors. However, the PBIAS values from  $R_{\text{atm}}$  (Fig. 8) show a similar random behavior like the estimates from  $R_{\text{hyd}}$ . Also, the number of catchments with PBIAS values  $\leq 10\%$  is similar compared to  $R_{\text{hyd}}$  (see

stacked bars in Fig. 8). On the other hand, there are slightly more catchments with bias values between 10% and 25%, which is mainly due to a better performance over Arctic catchments (e.g., Yenisei, Lena, Ob,

Kuskokwim, and Thelon). Here, while  $R_{\text{hyd}}$  tends to underestimate runoff (cf. blue and orange dots in Fig. 10),  $R_{\text{atm}}$  shows a better water budget closure with remaining relative imbalances of less than 25%. As many precipitation datasets suffer from an underestimation of precipitation in high-latitude regions (e.g., due to gauge undercatch errors; Lorenz and Kunstmann 2012), moisture flux divergences might be more reliable than those derived from precipitation and evapotranspiration over these regions. Of the three moisture flux divergence datasets, ERA-Interim and MERRA show a similar overall performance while CFSR reaches PBIAS values ( $\leq 25\%$ ) over less than 30 catchments, indicating large biases in the reanalysis model.

In terms of correlation (Fig. 11),  $R_{\text{atm}}$  shows slightly worse results than the best combinations from  $R_{\text{hyd}}$ . However, there are still about 45 catchments where moisture fluxes from ERA-Interim and MERRA reach correlations  $\geq 0.5$ , indicating reasonable dynamics in the runoff estimates from these combinations (see stacked bars in Fig. 11). Even if the better results over Arctic basins are not visible in the correlation matrix, there are still some catchments where  $R_{\text{atm}}$  higher correlations than  $R_{\text{hyd}}$  (e.g., Volga, Columbia, and Connecticut).

Finally, the matrix showing the NSE values (Fig. 12) reveals that the combinations using moisture fluxes from ERA-Interim and MERRA perform slightly better than the  $R_{\text{hyd}}$  estimates. Positive NSE values can be observed over at least 15 catchments, while there are only 8 catchments with values  $\geq 0.5$  (see stacked bars in Fig. 12). The matrix shows that  $R_{\text{atm}}$  provides better results mainly over Arctic catchments (e.g., Yenisei and Yana). Here, all combinations (except for those using CFSR moisture fluxes) reach values greater than 0.5, which indicates good agreement with the runoff observations.

The histograms (bottom row in Fig. 13) confirm the similar performance of moisture fluxes from ERA-Interim and MERRA and the significant biases in the CFSR dataset. Over roughly 50% of all the  $R_{\text{atm}}$  combinations and catchments involving moisture fluxes from ERA-Interim, correlations  $\geq 0.5$  with respect to observed runoff can be observed. In terms of PBIAS, MERRA has slightly more data points with biases  $\leq 50\%$ . Finally, the NSE shows that there is no superior data source of moisture flux divergences as both ERA-Interim and MERRA reach values between 0 and 0.5 (0.5 and 0.75) over about 10% (8%) of the data points.

#### e. Runoff from land surface models: $R_{\text{mod}}$

In terms of PBIAS (Fig. 8), the  $R_{\text{mod}}$  estimates from GLDAS<sub>CLM</sub> and GLDAS<sub>Noah3.3</sub> show lower biases than the other land surface models. However, compared with  $R_{\text{hyd}}$  and  $R_{\text{atm}}$ , the models seem to suffer from some

large biases over many catchments. For large basins like the Amazon, this can be partly explained with the lack of a routing scheme, which denies the correct estimation of peak runoff and thus creates large differences with respect to the observations. However, even over these large basins, there are single models that show PBIAS values  $\leq 10\%$  (e.g., GLDAS<sub>Noah3.3</sub> over Amazon and Volga or GLDAS<sub>CLM</sub> over Congo). This indicates that the lack of a routing scheme cannot fully explain the large deviations from the observations. Interestingly, even if GLDAS<sub>CLM</sub>, GLDAS<sub>Mosaic</sub>, GLDAS<sub>Noah2.7</sub>, and GLDAS<sub>VIC</sub> are driven by the same forcing data, the PBIAS shows some significant scattering, as there is not a single catchment where all these models perform consistently well. Overall, GLDAS<sub>Mosaic</sub> and GLDAS<sub>VIC</sub> show the largest biases with values smaller than 50% over less than 40 catchments (see stacked bars in Fig. 8).

In contrast to the PBIAS analysis, the correlations between  $R_{\text{mod}}$  and the observations draw a much more homogeneous picture (Fig. 11). There are many catchments with correlation values greater than 0.5 from all models. This holds true for large catchments (e.g., Mississippi, Volga, and Danube) but also for very small basins with low discharge (e.g., Rio Mearim, Gilbert, or Rio Prado). Thus, there is no dependency on a catchment's area or mean discharge, if land surface models are able to provide reasonable runoff dynamics. This performance is even more impressive when compared with the results from  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  as most land surface models show correlations of at least 0.5 (0.75) over 60 (30) and more catchments. Moreover,  $R_{\text{mod}}$  performs quite well over regions where  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  do not capture the runoff dynamics. These are either catchments where runoff has a much smaller amplitude compared to the other water cycle variables (e.g., Mississippi, Volga, Rhine, Don, and Neman) and/or no clear distinct annual cycle in precipitation (e.g., Rhine, Don, Daugava, Neman, Potomac, Altamaha, Weser, or Savannah). In these cases, the biases of the data sources of precipitation and evapotranspiration create short-term variations that result in low correlation values.

The NSE matrix (Fig. 12) shows a similar performance of  $R_{\text{mod}}$  and  $R_{\text{hyd}}$ . However, the two methods often succeed in totally different catchments. Especially over smaller catchments with low discharge,  $R_{\text{mod}}$  seems to provide reasonable runoff estimates (e.g., Potomac, the western Fitzroy, Victoria, or Tugela). Here, we see a very good agreement in the time series between the observed and estimated runoff. Furthermore, there are several catchments where most of the  $R_{\text{mod}}$  estimates provide NSE values greater than 0 (e.g., Tombigbee, Klamath, or Victoria), which should give evidence in the reliability of the modeled runoff in these regions.

### *f. Runoff from precipitation–runoff ratio: $R_{\text{pr}}$*

The rudimentary approach based on precipitation–runoff ratio is by far the most bias-free approach, with more than 45 (75) catchments that provide a PBIAS value of less than 10% (25%; see stacked bars in Fig. 8). Surprisingly, these values hold well irrespective of the precipitation dataset being used. Nevertheless, this method also fails in some catchments (St. Lawrence, Nelson, Churchill, Vuoksi, Rio Jequitinhonha, Suwanee, Orange, Murray, De Gray, and Rio Prado), wherein there is not a clear relationship between precipitation and runoff. The correlation values (Fig. 11) also demonstrate that the  $R_{\text{pr}}$  method captures the dynamics accurately, leading to a very good performance. Here again the choice of precipitation dataset does not matter at all for performance of the method, as nearly 60 catchments provide a correlation of more than 0.75. The catchments that returned poor PBIAS values also perform poorly in terms of correlation. The NSE values (Fig. 12) paint the same picture as PBIAS and correlation values. However, the CPC dataset performs relatively worse than the other precipitation datasets. In spite of this, the  $R_{\text{pr}}$  method performs the best of all the methods analyzed here.

This method is successful in catchments where there is a strong annual cycle (cyclostationarity) in both the precipitation and runoff datasets, which directly translates into a strong relationship in the  $P$ – $R$  ratio. This method fails when the runoff has a nonstationary behavior (e.g., see the time series of Nelson, St. Lawrence, or Vuoksi in the supplemental material), thus resulting in a lack of relationship in said ratio. Another interesting attribute of this method is that the ratio absorbs the biases in the precipitation datasets: the higher the biases, the smaller the ratio and vice versa. Although the  $P$ – $R$  ratio deals only with the stable relationship between runoff and precipitation, it implicitly demands the stability of the other two components of the water cycle: evapotranspiration and water storage change. For example, it is precisely because of the instability in water storage change that the method fails in catchments like St. Lawrence and Nelson, where the precipitation and evapotranspiration have a very strong annual cycle. Therefore, it can be said that the influence of evapotranspiration and water storage change is implicit in the  $P$ – $R$  ratio.

A theoretical advantage of  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  over models (without routing) and  $R_{\text{pr}}$  is that anthropogenic influences should not play a role when considering the level of agreement with respect to GRDC. As the predictions of  $R_{\text{pr}}$  are based on the mean monthly runoff–precipitation ratio, anthropogenic effects have to be

treated separately, or, more precisely, those effects cannot be taken into account when no runoff observations are available. However, we could not detect a better performance of the budget approaches compared to  $R_{\text{pr}}$  and  $R_{\text{mod}}$  over such catchments.

### *g. Water budget closure*

As  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  are based on the terrestrial and atmospheric–terrestrial water balance, they enable us to study the water budget (mis)closure at different time scales. The PBIAS value, which contains the mean of the differences between the observations and the estimates, provides the relative long-term imbalance of the water budget from these two methods. This can be expressed as follows:

$$\text{PBIAS}_{\text{hyd}} = \frac{\sum_{t=1}^T (R_{t,\text{obs}} - R_{t,\text{hyd}})}{\sum_{t=1}^T R_{t,\text{obs}}} = \frac{\sum_{t=1}^T \delta_{t,\text{hyd}}}{\sum_{t=1}^T R_{t,\text{obs}}} = \frac{\bar{\delta}_{\text{hyd}}}{\bar{R}_{\text{obs}}}, \quad (14a)$$

and similarly,

$$\text{PBIAS}_{\text{atm}} = \frac{\bar{\delta}_{\text{atm}}}{\bar{R}_{\text{obs}}}. \quad (14b)$$

At the annual scale, for almost all of the catchments, there is at least one data combination from either method that enables us to close the water budget (cf. Fig. 10 and annual mean and tropics of  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  in Fig. 14), but as mentioned in section 4c, the data combination differs from catchment to catchment. It is also clear that the  $R_{\text{hyd}}$  method, despite dealing with biases from three water balance equation variables, is able to close water budget better than the  $R_{\text{atm}}$  method (except for the Arctic catchments). This is conspicuous, especially in the tropics. The PBIAS values also show huge imbalances for the low discharge catchments (bottom part of Fig. 8), but care must be taken in interpreting these values. In these catchments, the imbalances are effectively divided by values close to zero, and therefore, those PBIAS values end up being orders of magnitude higher than for the catchments significant runoff (catchments on the right-hand side in the bottom rows of Figs. 9 and 10). To this end we also look at the absolute water budget imbalance.

Each of the data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences might be biased over certain regions and must thus be considered as a potential source of errors.

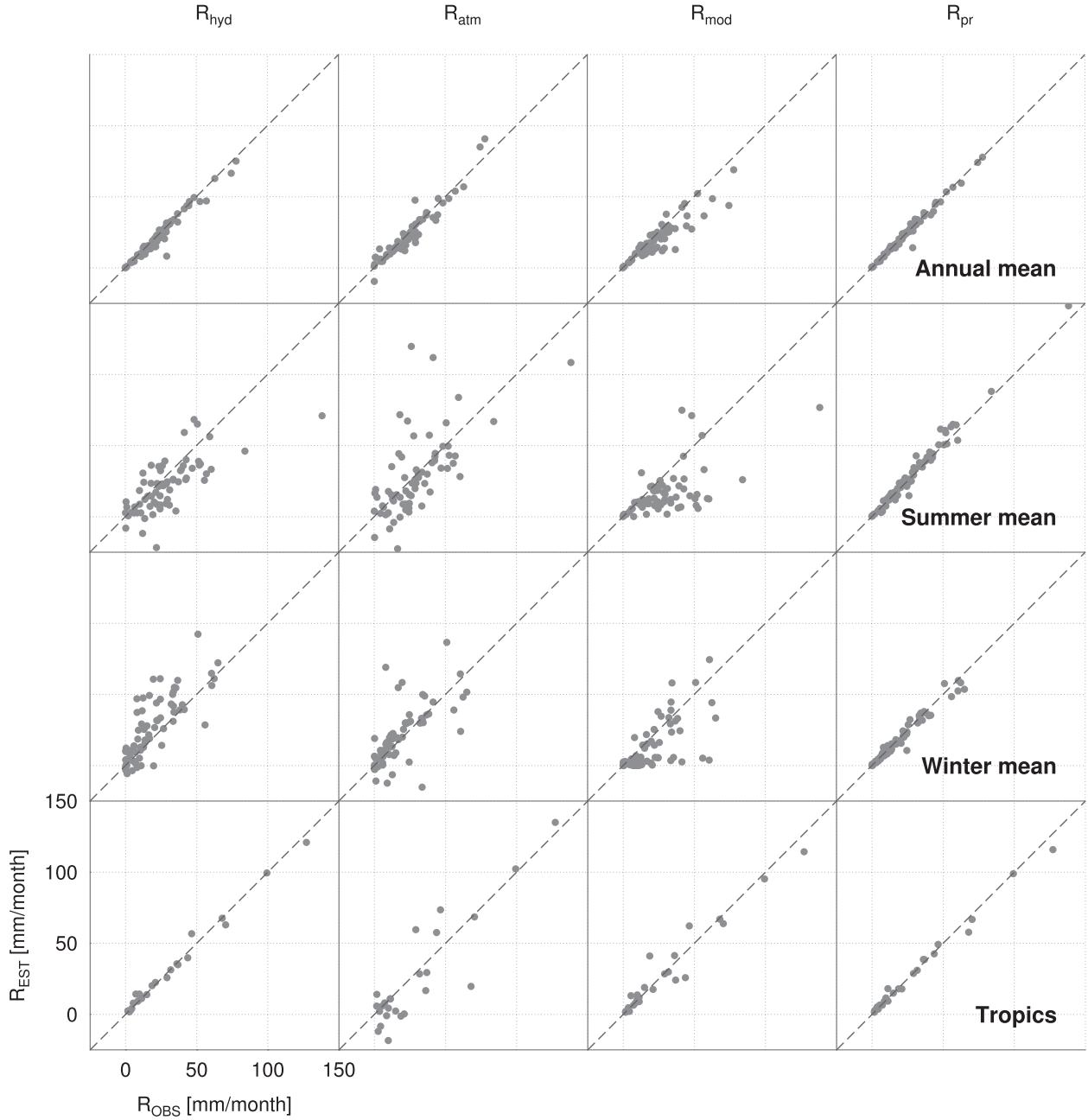


FIG. 14. Scatterplots of the (from top to bottom) mean annual runoff, summer and winter averages, and mean annual runoff over the tropical catchments. Each dot represents a single catchment. The  $x$  values are given by the GRDC observations while the  $y$  values are derived from the (from left to right) four approaches' best possible combinations for each catchment individually in terms of the PBIAS. The summer means are computed from the JJA averages over the Northern and the DJF averages over the Southern Hemisphere catchments.

In some cases, such biases can get canceled out when computing  $R_{\text{hyd}}$  or  $R_{\text{atm}}$ , leading to random errors that scatter around zero. This is the case for catchments like Rio Tapajos, St. Lawrence, Rio Jari, Tombigbee, or even small basins like San Joaquin (Fig. 10). In these cases, the scattering gives information on the

uncertainty level of the water cycle variables. Over most catchments, however, either a positive or negative bias for most combinations can be observed. This indicates general shortcomings in the current quality of the different water cycle variables. Most obviously, this can be observed for the Arctic (tropical) catchments, where the

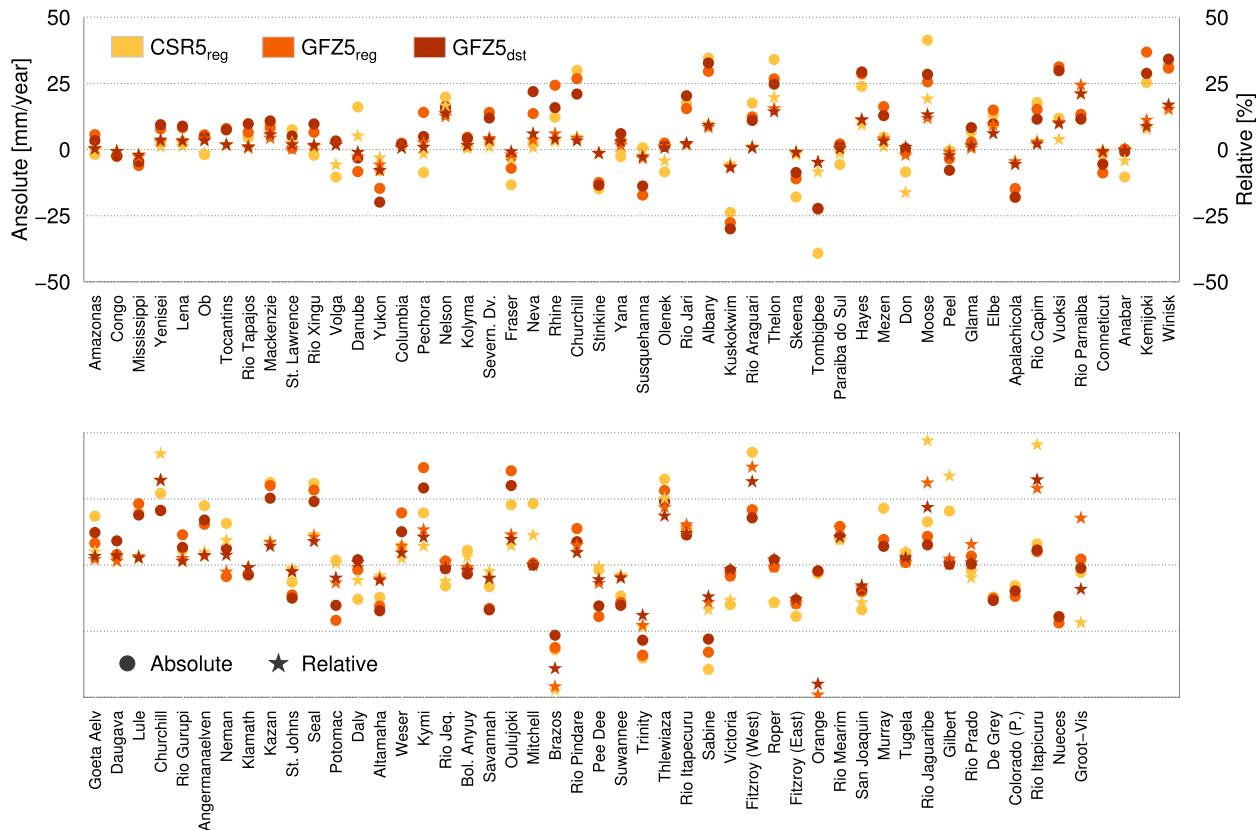


FIG. 15. Contribution of water storage changes from the three GRACE datasets to the long-term water budget over the 96 study regions. The circles show the absolute contribution ( $\text{mm yr}^{-1}$ ) while the stars depict the relative contribution with respect to the mean annual runoff.

majority of  $R_{\text{hyd}}$  combinations consistently tend to underestimate (overestimate) runoff (Fig. 10).

The main reason for these heavily biased runoff estimates can be assumed to originate from uncertainties in the precipitation and evapotranspiration datasets. According to Fig. 15, there is no clear contribution of water storage changes from GRACE to the catchments showing large remaining imbalances. For example, for Tocantins, where  $R_{\text{hyd}}$  is significantly overestimated, the GRACE datasets show an increase of water storage of less than  $10 \text{ mm yr}^{-1}$ . The same holds true for several Arctic catchments (e.g., Yenisei or Lena). In general, the total annual water storage as depicted by GRACE is less than  $\pm 50 \text{ mm yr}^{-1}$  for all catchments, but for most of the catchments, the total annual water storage values are within  $\pm 25 \text{ mm yr}^{-1}$ . Therefore, for the majority of the analyzed catchments, as GRACE does not contribute a lot to the imbalances, the biases in the precipitation and evapotranspiration datasets contribute heavily to the over- and underestimation.

At the monthly time scales, use can be made of the NSE to study the water budget closure because it contains the mean-squared error (MSE) of estimates with

respect to the observations in the numerator and the variance of the observations in the denominator. Thus, it is the normalized mean squared error of the estimates:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^T (R_{t,\text{obs}} - R_{t,\text{est}})^2}{\sum_{t=1}^T (R_{t,\text{obs}} - \bar{R}_{\text{obs}})^2}, \quad (15)$$

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (R_{t,\text{obs}} - R_{t,\text{est}})^2, \quad (16a)$$

$$\sigma_{\text{obs}}^2 = \frac{1}{T} \sum_{t=1}^T (R_{t,\text{obs}} - \bar{R}_{\text{obs}})^2, \quad \text{and} \quad (16b)$$

$$\text{therefore } \text{NSE} = 1 - \frac{\text{MSE}}{\sigma_{\text{obs}}^2}, \quad (16c)$$

where  $\sigma_{\text{obs}}^2$  is the variance of the observed runoff. Like PBIAS, when the catchment runoff is close to zero, the NSE values become bloated. Nevertheless, the closure at monthly time scales is far from being acceptable. Only

around 15 catchments provide a positive value (Fig. 12), which clearly indicates that the datasets are not mature enough to close the water balance at monthly time scales.

## 5. Summary and conclusions

In this contribution, we have estimated runoff from water balance equations mainly using GRACE water storage changes. A wide variety of hydrological and hydrometeorological datasets have been used to arrive at the runoff estimates. The estimates were then compared to observed runoff. This was only possible in about 96 catchments because of the limited time span of the GRACE dataset and the limited data availability and large data gaps in the GRDC database. To quantify the estimates holistically, we included runoff estimates from sophisticated hydrological land surface models and an empirical model involving the ratio between precipitation and runoff in the analysis. The evaluation of the estimates was carried out using three performance metrics, namely, PBIAS, correlation coefficient, and NSE.

Based on the assessment, the following conclusions are drawn.

- (i) Overall, the estimates from the precipitation–runoff ratio, though being a simple method, provide by far the best results when compared to the budget-based methods and the land surface models. This also implies that for most catchments, the impact of evapotranspiration and water storage changes on runoff underlies a stable annual cycle, which remains constant from year to year. However, the method can only succeed when the monthly relationship between runoff and precipitation remains constant over the years. For catchments with more or less random time series of precipitation or runoff, no stable precipitation–runoff ratio can be derived. For such catchments, the member of  $R_{\text{pr}}$  showed consistently bad results.
- (ii) In the water-balance-based estimates, there is no combination of datasets that is consistently standing out as the best, and there is also no combination that behaved consistently over different catchments. In particular, the PBIAS analysis shows highly variable results from the  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  estimates. The results further indicate that  $R_{\text{hyd}}$  tends to underestimate (overestimate) runoff over Arctic (tropical) catchments. For Arctic catchments, however,  $R_{\text{atm}}$  shows better PBIAS values. As a consequence, the quality of different data sources of precipitation, evapotranspiration, water storage changes, and moisture fluxes differ significantly from catchment to catchment. The
- (iii) This indicates that one cannot extrapolate the performance of a combination or method to other basins without thorough validation. To a certain extent, this holds also true for the estimates from the precipitation–runoff ratio. Performance of combinations thus requires scrutiny over individual catchments. This also has strong implications for the regionalization of these estimates to adjacent catchments.
- (iv) The estimates from  $R_{\text{hyd}}$  and  $R_{\text{atm}}$  perform well in catchments where the runoff exhibits a strong annual cycle combined with a large volume of runoff. Over such catchments, most combinations reach correlations of at least 0.75. Basins where the magnitudes of the variations in the time series of runoff are far below the other water cycle variables suffer from biases in the different data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences. These biases create short-time variations, which deny a reasonable estimation of runoff.
- (v) Further, the water-balance-based methods are reliant on GRACE datasets, which are reported to be of inferior spatial resolution compared to the other datasets used here. In spite of this shortcoming, the  $R_{\text{hyd}}$  estimates still provide good performance in small catchments, but only those with a strong annual cycle. Moreover, the different GRACE datasets show significantly different results in catchments with a time shift between peak runoff, precipitation, and evapotranspiration. In these cases, GRACE has to take care of the water storage and thus has a major influence on the performance of the  $R_{\text{hyd}}$  estimates.
- (vi) For most catchments, there is still a significant relative water budget imbalance with respect to the mean runoff. The majority of combinations from  $R_{\text{hyd}}$  suffer from a remaining imbalance of more than 25% over many basins. However, over most catchments, there is at least a single combination that leads to a relative imbalance close to 0%, indicating a reasonable water budget closure. Moreover, over several catchments, the absolute imbalance scatters around zero. Here, the water budget residual can be assumed to be randomly distributed between the different combinations of  $R_{\text{hyd}}$ . This emphasizes the use of ensemble-based methods for analyzing the water cycle over such catchments.
- (vii) The models performed comparatively poorly when compared with the other approaches. A reason for this might be due to the lack of a routing scheme. However, even on annual time scales, significant

biases could be observed when compared with the observations. It could be also detected that over several catchments, modeled runoff might show a significant time lag. However, this is again not consistent throughout all the study regions, which makes it inevitable to evaluate such data sources both in the individual catchments and also globally.

In general, none of the analyzed methods is able to provide reasonable runoff estimates globally. There are, however, several regions where the independent estimates from  $R_{\text{hyd}}$ ,  $R_{\text{atm}}$ , and  $R_{\text{mod}}$  are able to successfully reproduce observed runoff. But as the data combinations and models that perform well differ from catchment to catchment, we cannot rely on a single dataset for analyzing the basin-scale water cycle. Therefore, there is still an urgent need for in situ observations of the major water cycle variables, which are mandatory for further necessary improvements of global hydrological datasets.

**Acknowledgments.** We would like to thank the Helmholtz Alliance for Remote Sensing and Earth System Dynamics for funding this study. Further, we gratefully acknowledge the support of projects SN13/1, BA1150/11, KU2090/1 by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program SPP1257 *Mass Transport and Mass Distribution in the System Earth*. We thank the following data providers: Global Runoff Data Centre (GRDC); Global Precipitation Climatology Project (GPCP); Global Precipitation Climatology Centre (GPCC); University of Delaware (UDEL); Numerical Terradynamic Simulation Group at the University of Montana (NTSG UMT); Max Planck Institute for Biogeochemistry (MPI BGC); European Centre for Medium-Range Weather Forecasts (ECMWF); The University of Texas at Austin, Center for Space Research (CSR); NASA's Goddard Earth Science Data and Information Services Center (GES DISC); NOAA's Earth System Research Laboratory, Physical Sciences Division (ESRL PSD); and NCAR's Computational and Information Systems Laboratory Research Data Archive (CISL RDA). Further thanks are due to Dr. Diego Miralles (University of Bristol) for kindly providing the GLEAM evapotranspiration data. All graphics in the document were produced with the Generic Mapping Tools (GMT; [Wessel and Smith 1991](#)).

## APPENDIX

### Treating the Errors in GRACE Mass Anomalies

The major sources of errors in the GRACE data are aliasing of residual tidal signal ([Seo et al. 2008](#)), poor

estimates of  $C_{20}$  coefficients ([Chen et al. 2005](#)), and high-frequency noise in the spherical harmonic coefficients due to orbit geometry ([Kusche et al. 2009](#)). These errors are treated at the spherical harmonic coefficient level.

#### a. Tidal residuals

The tidal residuals alias frequencies that depend on the orbit geometry of GRACE, sampling rate of GRACE, and the natural frequency of the corresponding tides, which enables them to be modeled and removed from the signal to a large extent. This is accomplished by performing a least squares spectrum analysis on the residuals of GRACE temporal geoid anomalies, which are obtained after removing the mean annual behavior. The mean annual behavior is computed by taking the average of the 12 calendar months as is done for the precipitation–runoff ratio (8) [see [Tourian et al. \(2011\)](#) for the efficacy of this method]. The spectral analysis is performed only for select tidal frequencies:  $M_1$ ,  $S_1$ ,  $S_2$ ,  $P_1$ ,  $Q_1$ ,  $O_1$  and  $N_2$ . The estimates of the aliased frequencies are subsequently subtracted from the residuals of the spherical harmonic coefficients, which are then added back to the mean annual behavior to obtain the corrected monthly anomalies.

#### b. Erroneous $C_{20}$ coefficients

The erroneous  $C_{20}$  coefficients are replaced with those estimated from satellite laser ranging (SLR) in the case of the CSR monthly solutions as suggested by [Cheng and Ries \(2014\)](#). However, the  $C_{20}$  coefficients in the GFZ monthly solutions are used as they are well behaved. This is due to the use of EIGEN-6C gravity field model as a background model in the processing scheme ([Dahle et al. 2013](#)), which includes data from GRACE, Gravity Field and Steady-State Ocean Circulation Explorer (GOCE), and Laser Geodynamics Satellite ( Lageos; SLR) and also data from airborne and terrestrial gravimetry.

#### c. High-frequency noise

The high-frequency noise in the spherical harmonic coefficients manifests itself as stripe-like patterns in the spatial domain. This is dealt with by filtering, and here we apply two methods of filtering: a decorrelation filter in combination with a Gaussian filter with a filter radius of 350 km ([Swenson and Wahr 2006](#)) and a regularization filter as proposed by [Kusche \(2007\)](#), but with certain modifications. The modifications are that while [Kusche \(2007\)](#) computes the signal covariance from a combination of different models, we compute the signal covariance from the GRACE data itself, as described by [Sasgen et al. \(2006\)](#); [Kusche \(2007\)](#) also uses stationary signal and noise covariances, whereas we use cyclostationary signal

covariances and nonstationary noise covariances. For the GFZ dataset, standard deviations of the spherical harmonic coefficients are provided, which were used for the simulation of noise covariances. The standard deviations for the CSR dataset are not provided, and therefore, we have used standard deviations from the release 4 dataset for the noise covariance simulation.

## REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, doi:[10.1175/1525-7541\(2003\)004<1147:TVGCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGCP>2.0.CO;2).
- Alsdorf, D. E., E. Rodríguez, and D. P. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, **45**, RG2002, doi:[10.1029/2006RG000197](https://doi.org/10.1029/2006RG000197).
- Barrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala, 2009: The ERA-Interim archive version 1.0. ERA Rep. Series, Rep. 1, ECMWF, 16 pp. [Available online at [http://old.ecmwf.int/publications/library/ecpublications/\\_pdf/era/era\\_report\\_series/RS\\_1.pdf](http://old.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_1.pdf).]
- Bettadpur, S., 2012: GRACE UTCSR level-2 processing standards document (for level-2 product release 0005). Rev. 4.0, Doc. GRACE 327-742 (CSR-GR-12-xx), Center for Space Research, The University of Texas at Austin, 17 pp. [Available online at [ftp://podaac.jpl.nasa.gov/allData/grace/docs/L2-CSR0005\\_ProcStd\\_v4.0.pdf](ftp://podaac.jpl.nasa.gov/allData/grace/docs/L2-CSR0005_ProcStd_v4.0.pdf).]
- Bonan, G. B., 1998: The land surface climatology of the NCAR Land Surface Model coupled to the NCEP Community Climate Model. *J. Climate*, **11**, 1307–1326, doi:[10.1175/1520-0442\(1998\)011<1307:TLSCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<1307:TLSCOT>2.0.CO;2).
- Chen, F., and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268, doi:[10.1029/95JD02165](https://doi.org/10.1029/95JD02165).
- Chen, J. L., M. Rodell, C. R. Wilson, and J. S. Famiglietti, 2005: Low degree spherical harmonic influences on Gravity Recovery and Climate Experiment (GRACE) water storage estimates. *Geophys. Res. Lett.*, **32**, L14405, doi:[10.1029/2005GL022964](https://doi.org/10.1029/2005GL022964).
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, doi:[10.1029/2007JD009132](https://doi.org/10.1029/2007JD009132).
- Cheng, M., and J. Ries, 2014: Monthly estimates of C20 from 5 SLR satellites based on GRACE RL05 models. GRACE Tech. Note 07, Center for Space Research, The University of Texas at Austin, 1 pp. [Available online at [ftp://podaac.jpl.nasa.gov/allData/grace/docs/TN-07\\_C20\\_SLR.txt](ftp://podaac.jpl.nasa.gov/allData/grace/docs/TN-07_C20_SLR.txt).]
- Crowley, J. W., J. X. Mitrovica, R. C. Bailey, M. E. Tamisiea, and J. L. Davis, 2006: Land water storage within the Congo basin inferred from GRACE satellite gravity data. *Geophys. Res. Lett.*, **33**, L19402, doi:[10.1029/2006GL027070](https://doi.org/10.1029/2006GL027070).
- Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer, 2013: GFZ GRACE level-2 processing standards document for level-2 product release 0005. Scientific Tech. Rep. STR12/02, GFZ German Research Centre for Geosciences, 26 pp., doi:[10.2312/GFZ.b103-1202-25](https://doi.org/10.2312/GFZ.b103-1202-25).
- Dai, Y., and Coauthors, 2003: The Common Land Model. *Bull. Amer. Meteor. Soc.*, **84**, 1013–1023, doi:[10.1175/BAMS-84-8-1013](https://doi.org/10.1175/BAMS-84-8-1013).
- 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, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the national meteorological center. *Wea. Forecasting*, **6**, 538–547, doi:[10.1175/1520-0434\(1991\)006<0538:TNGOAS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2).
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of NOAH land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:[10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296).
- Entekhabi, D., and Coauthors, 2010: The soil moisture active passive (SMAP) mission. *Proc. IEEE*, **98**, 704–716, doi:[10.1109/JPROC.2010.2043918](https://doi.org/10.1109/JPROC.2010.2043918).
- 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, doi:[10.1175/2011JHM1388.1](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. [Available online at <http://iahs.info/uploads/dms/309015.pdf>.]
- , —, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles*, **16**, 1042, doi:[10.1029/1999GB001254](https://doi.org/10.1029/1999GB001254).
- , U. Loosar, A. Pietroniro, and R. D. Robarts, 2012: Rationale for monitoring discharge on the ground. *J. Hydrometeor.*, **13**, 1977–1986, doi:[10.1175/JHM-D-11-0126.1](https://doi.org/10.1175/JHM-D-11-0126.1).
- Fersch, B., H. Kunstmann, A. Bárdossy, B. Devaraju, and N. Sneeuw, 2012: Continental-scale basin water storage variation from global and dynamically downscaled atmospheric water budgets in comparison with GRACE-derived observations. *J. Hydrometeor.*, **13**, 1589–1603, doi:[10.1175/JHM-D-11-0143.1](https://doi.org/10.1175/JHM-D-11-0143.1).
- GRDC, 2013: Tenth meeting of the GRDC Steering Committee, 15–17 June 2011, Koblenz, Germany. GRDC Rep. Series, Rep. 42, Global Runoff Data Centre, 31 pp., doi:[10.5675/GRDC\\_Report\\_42](https://doi.org/10.5675/GRDC_Report_42).
- Gupta, H., S. Shrooshian, and P. O. Yapo, 1999: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.*, **4**, 135–143, doi:[10.1061/\(ASCE\)1084-0699\(1999\)4:2\(135\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)).
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2013: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, doi:[10.1002/joc.3711](https://doi.org/10.1002/joc.3711).
- Hrachowitz, M., and Coauthors, 2013: A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrol. Sci. J.*, **58**, 1198–1255, doi:[10.1080/02626667.2013.803183](https://doi.org/10.1080/02626667.2013.803183).
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:[10.1029/2009GL040000](https://doi.org/10.1029/2009GL040000).
- Jones, P., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, doi:[10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2).
- Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using

- a biosphere model. *Biogeosciences*, **6**, 2001–2013, doi:[10.5194/bg-6-2001-2009](https://doi.org/10.5194/bg-6-2001-2009).
- , and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, doi:[10.1038/nature09396](https://doi.org/10.1038/nature09396).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kerr, Y. H., and Coauthors, 2010: The SMOS Mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE*, **98**, 666–687, doi:[10.1109/JPROC.2010.2043032](https://doi.org/10.1109/JPROC.2010.2043032).
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–268, doi:[10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2).
- Kopp, T. J., and R. B. Kiess, 1996: The air force global weather central cloud analysis model. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 220–222.
- Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.*, **104**, 19569–19585, doi:[10.1029/1999JD900232](https://doi.org/10.1029/1999JD900232).
- Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Tech. Rep. Series on Global Modeling and Data Assimilation, Vol. 9, 60 pp. [Available online at <http://gmao.gsfc.nasa.gov/pubs/docs/Koster130.pdf>.]
- Kummerow, C., and Coauthors, 2000: The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *J. Appl. Meteor.*, **39**, 1965–1982, doi:[10.1175/1520-0450\(2001\)040<1965:TSOTTR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2).
- Kusche, J., 2007: Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *J. Geod.*, **81**, 733–749, doi:[10.1007/s00190-007-0143-3](https://doi.org/10.1007/s00190-007-0143-3).
- , R. Schmidt, S. Petrovic, and R. Rietbroek, 2009: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *J. Geod.*, **83**, 903–913, doi:[10.1007/s00190-009-0308-3](https://doi.org/10.1007/s00190-009-0308-3).
- Landerer, F. W., J. O. Dickey, and A. Günther, 2010: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *J. Geophys. Res.*, **115**, D23115, doi:[10.1029/2010JD014584](https://doi.org/10.1029/2010JD014584).
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, **99**, 14 415–14 428, doi:[10.1029/94JD00483](https://doi.org/10.1029/94JD00483).
- Longuevergne, L., B. R. Scanlon, and C. R. Wilson, 2010: GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resour. Res.*, **46**, W11517, doi:[10.1029/2009WR008564](https://doi.org/10.1029/2009WR008564).
- Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.*, **13**, 1397–1420, doi:[10.1175/JHM-D-11-088.1](https://doi.org/10.1175/JHM-D-11-088.1).
- Matsuura, K., and C. J. Willmott, 2012: Terrestrial precipitation: 1900–2010 gridded monthly time series (version 3.02). Center for Climatic Research, University of Delaware, Newark, DE. [Available online at [http://climate.geog.udel.edu/~climate/html\\_pages/download.html#P2011rev](http://climate.geog.udel.edu/~climate/html_pages/download.html#P2011rev).]
- Miralles, D. G., R. A. M. de Jeu, J. H. Gash, T. R. H. Holmes, and A. J. Dolman, 2011: An application of GLEAM to estimating global evaporation. *Hydrol. Earth Syst. Sci. Discuss.*, **8**, 1–27, doi:[10.5194/hessd-8-1-2011](https://doi.org/10.5194/hessd-8-1-2011).
- Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running, 2007: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.*, **111**, 519–536, doi:[10.1016/j.rse.2007.04.015](https://doi.org/10.1016/j.rse.2007.04.015).
- , M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, **115**, 1781–1800, doi:[10.1016/j.rse.2011.02.019](https://doi.org/10.1016/j.rse.2011.02.019).
- Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.*, **38**, L06492, doi:[10.1029/2010GL046230](https://doi.org/10.1029/2010GL046230).
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:[10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Parkinson, C., 2003: Aqua: An earth-observing satellite mission to examine water and other climate variables. *IEEE Trans. Geosci. Remote Sens.*, **41**, 173–183, doi:[10.1109/TGRS.2002.808319](https://doi.org/10.1109/TGRS.2002.808319).
- Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics, 520 pp.
- Reichle, R. H., R. D. Koster, G. J. M. D. Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Tour, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, doi:[10.1175/JCLI-D-10-05033.1](https://doi.org/10.1175/JCLI-D-10-05033.1).
- Rieger, J., and M. J. Tourian, 2014: Characterization of runoff-storage relationships by satellite gravimetry and remote sensing. *Water Resour. Res.*, **50**, 3444–3466, doi:[10.1002/2013WR013847](https://doi.org/10.1002/2013WR013847).
- , —, B. Devaraju, and N. Sneeuw, 2012: Analysis of GRACE uncertainties by hydrological and hydro-meteorological observations. *J. Geodyn.*, **59–60**, 16–27, doi:[10.1016/j.jog.2012.02.001](https://doi.org/10.1016/j.jog.2012.02.001).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- Roads, J. O., S. C. Chen, A. K. Guetter, and K. P. Georgakos, 1994: Large-scale aspects of the United States hydrologic cycle. *Bull. Amer. Meteor. Soc.*, **75**, 1589–1610, doi:[10.1175/1520-0477\(1994\)075<1589:LSAOTU>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<1589:LSAOTU>2.0.CO;2).
- Rodell, M., J. S. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004a: Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophys. Res. Lett.*, **31**, L20504, doi:[10.1029/2004GL020873](https://doi.org/10.1029/2004GL020873).
- , and Coauthors, 2004b: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, doi:[10.1175/BAMS-85-3-381](https://doi.org/10.1175/BAMS-85-3-381).
- , I. Velicogna, and J. Famiglietti, 2009: Satellite-based estimates of ground-water depletion in India. *Nature*, **460**, 999–1002, doi:[10.1038/nature08238](https://doi.org/10.1038/nature08238).
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1).
- Salvucci, G. D., and P. Gentile, 2013: Emergent relation between surface vapor conductance and relative humidity profiles yields evaporation rates from weather data. *Proc. Natl. Acad. Sci. USA*, **110**, 6287–6291, doi:[10.1073/pnas.1215844110](https://doi.org/10.1073/pnas.1215844110).
- Sasgen, I., Z. Martinec, and K. Fleming, 2006: Wiener optimal filtering of GRACE data. *Stud. Geophys. Geod.*, **50**, 499–508, doi:[10.1007/s11200-006-0031-y](https://doi.org/10.1007/s11200-006-0031-y).
- Schlosser, C. A., and P. R. Houser, 2007: Assessing a satellite-era perspective of the global water cycle. *J. Climate*, **20**, 1316–1338, doi:[10.1175/JCLI4057.1](https://doi.org/10.1175/JCLI4057.1).

- Schmidt, R., F. Flechtner, U. Meyer, K.-H. Neumayer, C. Dahle, R. König, and J. Kusche, 2008: Hydrological signals observed by the GRACE satellites. *Surv. Geophys.*, **29**, 319–334, doi:[10.1007/s10712-008-9033-3](https://doi.org/10.1007/s10712-008-9033-3).
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, doi:[10.1007/s00704-013-0860-x](https://doi.org/10.1007/s00704-013-0860-x).
- Schrama, E. J. O., and B. Wouters, 2011: Revisiting Greenland Ice Sheet mass loss observed by GRACE. *J. Geophys. Res.*, **116**, B02407, doi:[10.1029/2009JB006847](https://doi.org/10.1029/2009JB006847).
- Seitz, F., M. Schmidt, and C. K. Shum, 2008: Signals of extreme weather conditions in central Europe in GRACE 4-D hydrological mass variations. *Earth Planet. Sci. Lett.*, **268**, 165–170, doi:[10.1016/j.epsl.2008.01.001](https://doi.org/10.1016/j.epsl.2008.01.001).
- Seo, K. W., C. R. Wilson, S. C. Han, and D. E. Waliser, 2008: Gravity Recovery and Climate Experiment (GRACE) alias error from ocean tides. *J. Geophys. Res.*, **113**, B03405, doi:[10.1029/2006JB004747](https://doi.org/10.1029/2006JB004747).
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, doi:[10.1175/JCLI3790.1](https://doi.org/10.1175/JCLI3790.1).
- , C. R. Ferguson, T. J. Troy, E. F. Wood, and M. F. McCabe, 2009: Closing the terrestrial water budget from satellite remote sensing. *Geophys. Res. Lett.*, **36**, L07403, doi:[10.1029/2009GL037338](https://doi.org/10.1029/2009GL037338).
- Shiklomanov, A. I., R. B. Lammers, and C. J. Vörösmarty, 2002: Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Trans. Amer. Geophys. Union*, **83**, 13–17, doi:[10.1029/2002EO000007](https://doi.org/10.1029/2002EO000007).
- Swenson, S., and J. Wahr, 2006: Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. *J. Hydrometeor.*, **7**, 252–270, doi:[10.1175/JHM478.1](https://doi.org/10.1175/JHM478.1).
- Syed, T. H., J. S. Famiglietti, J. Chen, M. Rodell, S. I. Seneviratne, P. Viterbo, and C. R. Wilson, 2005: Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land–atmosphere water balance. *Geophys. Res. Lett.*, **32**, L24404, doi:[10.1029/2005GL024851](https://doi.org/10.1029/2005GL024851).
- , —, V. Zlotnicki, and M. Rodell, 2007: Contemporary estimates of pan-Arctic freshwater discharge from GRACE and reanalysis. *Geophys. Res. Lett.*, **34**, L19404, doi:[10.1029/2007GL031254](https://doi.org/10.1029/2007GL031254).
- , —, and D. P. Chambers, 2009: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J. Hydrometeor.*, **10**, 22–40, doi:[10.1175/2008JHM993.1](https://doi.org/10.1175/2008JHM993.1).
- Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, 2004: GRACE measurements of mass variability in the Earth system. *Science*, **305**, 503–505, doi:[10.1126/science.1099192](https://doi.org/10.1126/science.1099192).
- Tourian, M. J., J. Rieger, N. Sneeuw, and B. Devaraju, 2011: Outlier identification and correction for GRACE aggregated data. *Stud. Geophys. Geod.*, **55**, 627–640, doi:[10.1007/s11200-009-9007-z](https://doi.org/10.1007/s11200-009-9007-z).
- , N. Sneeuw, and A. Bárdossy, 2013: A quantile function approach to discharge estimation from satellite altimetry (ENVISAT). *Water Resour. Res.*, **49**, 4174–4186, doi:[10.1002/wrcr.20348](https://doi.org/10.1002/wrcr.20348).
- Trenberth, K. E., and J. T. Fasullo, 2013: Regional energy and water cycles: Transports from ocean to land. *J. Climate*, **26**, 7837–7851, doi:[10.1175/JCLI-D-13-00008.1](https://doi.org/10.1175/JCLI-D-13-00008.1).
- , 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, doi:[10.1175/JHM600.1](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, doi:[10.1175/2011JCLI4171.1](https://doi.org/10.1175/2011JCLI4171.1).
- Vörösmarty, C. J., and Coauthors, 2001: Global water data: A newly endangered species. *Eos, Trans. Amer. Geophys. Union*, **82**, 54–58, doi:[10.1029/01EO00031](https://doi.org/10.1029/01EO00031).
- , and Coauthors, 2010: Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561, doi:[10.1038/nature09440](https://doi.org/10.1038/nature09440).
- Wahr, J., M. Molenaar, and F. Bryan, 1998: The time-variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.*, **103**, 30 205–30 230, doi:[10.1029/98JB02844](https://doi.org/10.1029/98JB02844).
- Werth, S., A. Günther, R. Schmidt, and J. Kusche, 2009: Evaluation of GRACE filter tools from a hydrological perspective. *Geophys. J. Int.*, **179**, 1499–1515, doi:[10.1111/j.1365-246X.2009.04355.x](https://doi.org/10.1111/j.1365-246X.2009.04355.x).
- Wessel, P., and W. H. F. Smith, 1991: Free software helps map and display data. *Eos, Trans. Amer. Geophys. Union*, **72**, 441–446, doi:[10.1029/90EO00319](https://doi.org/10.1029/90EO00319).
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:[10.1175/1520-0477\(1997\)078<2539:GPAYMA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2).
- Zaitchik, B. F., M. Rodell, and F. Olivera, 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resour. Res.*, **46**, W06507, doi:[10.1029/2009WR007811](https://doi.org/10.1029/2009WR007811).
- Zhou, X., Y. Zhang, Y. Wang, H. Zhang, J. Vaze, L. Zhang, Y. Yang, and Y. Zhou, 2012: Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *J. Hydrol.*, **470–471**, 269–279, doi:[10.1016/j.jhydrol.2012.09.002](https://doi.org/10.1016/j.jhydrol.2012.09.002).