



Are we approaching water budget closure at basin scale for the right reasons?

Julien Lefebve¹, Sylvain Biancamaria¹, Alejandro Blazquez¹, Simon Munier², Laetitia Gal³, Rômulo Augusto Jucá Oliveira³, Jordi Etchanchu⁴

5 ¹Universite de Toulouse, LEGOS (CNES, CNRS, IRD, UT), Toulouse, 31400, France

²CNRM, Toulouse, 31057, France

³Hydro Matters, Toulouse, 31400, France

⁴HSM, Montpellier, 34093, France

10 *Correspondence to:* Julien Lefebve (julien.lefebve@cnes.fr)

Abstract.

The water cycle represents the continuous circulation of water through the atmosphere, the continents and the oceans. At the basin scale, this cycle can be decomposed into incoming and outgoing fluxes as well as changes of water stored within the basin. The relationship between these fluxes and changes in storage is called the water budget. Closing the water-budget at
15 basin scale, using observations, model outputs or a combination of both, provides a deeper understanding of water fluxes, storage changes as well as the underlying hydrological processes.

Satellite measurements, especially those from the Gravity Recovery And Climate Experiment (GRACE, 2002-2017) and GRACE Follow-On (GRACE-FO, 2018-present) missions, now allows to estimate the water budget closure level for large basins worldwide, including those poorly or not instrumented at all. One of the outgoing fluxes for exorheic basins, river
20 discharge, is not available for many basins, particularly during the twenty-first century due to the spatial and temporal heterogeneity of openly accessible in-situ river discharge. To address this issue, several studies have attempted to close the basin-scale water-budget by estimating river discharge as the simple sum of runoff calculated from Land Surface Models. Despite well-documented significant errors in runoff estimation, uncertainties associated with these approaches are often neglected as river discharge is, in most cases, the flux with the smallest amplitude in the water budget. In this study, using an
25 ensemble approach to assess the state-of-the-art estimates of water cycle fluxes and stocks, we precisely evaluate the impact of river discharge errors on the conclusions drawn from the water budget analysis of 82 exorheic basins worldwide. When ensembles of products are used to estimate combination of precipitation, evapotranspiration and discharge that are the closest to the GRACE storage change variations, we demonstrate compensations between the components of the water budget can occur. Although river discharge is the flux with the smallest amplitude, its dispersion (36%), when derived from aggregated
30 runoff, is comparable to that of precipitation (44%) and evapotranspiration (34%) and a seemingly high level of water budget closure can be achieved with, for some basin, errors of more than 40% in the annual volume assessment of precipitation and evaporation. Errors caused by these compensations are larger for continental, sub-arctic and polar basins (around 15 %) than



for tropical, arid and temperate basins (around 6 %). Moreover, applying a simple routing scheme to the runoff does not significantly reduce these errors. In this study, we mitigate and assess these compensations by using a combination of in situ and satellite-based discharge estimates. Future studies should explore methods that further limit the compensation effect when attempting to approach a water budget closure at basin scale.

1 Introduction

The water cycle describes the continuous movement of water within the Earth and atmosphere. The study of the continental component of the water cycle at the basin scale can be conducted through the evaluation of our ability to approach the closure of its water budget (Dorigo et al., 2021). At basin scale, the water cycle can be indeed decomposed into incoming and outgoing fluxes as well as change of water stored in the basin. The incoming fluxes are liquid or solid precipitations (P) while the outgoing fluxes are evapotranspiration (ET) and discharge at the outlet (Q). The total amount of water stored in different forms on the surface and in the underground layers of the Earth (river, lakes, reservoirs, snow, ice, and groundwater) can be grouped under a common term: Total Water Storage (TWS). The launches of the GRACE gravimetry mission (2002) (Tapley et al., 2004) and its successor GRACE FO (2018) (Landerer et al., 2020) have enabled the observation of TWS monthly variations on a planetary scale, thus allowing monthly water budget closure assessment at large spatial scales (Rodell et al., 2018). Nevertheless, achieving a detailed understanding of the continental component of the water cycle at basin scale remains a challenge (Bloeschl et al., 2019). This challenge is amplified because the water cycle is increasingly perturbed by human activities and climate change (Abbott et al., 2019; Bonfils et al., 2020; Marvel et al., 2019; Padrón et al., 2020).

For this purpose, we will refer to Water Budget Evaluation (WBE) as the evaluation of the ability to approach the closure of the monthly water budget at basin scale. Achieving a high level of closure in a WBE increases confidence in the estimates of the individual budget components. Consequently, many studies have attempted to compute the WBE for different basins. These studies combine gravimetry-based TWS variation data with remote sensing products, models outputs, or direct measurements in order to evaluate monthly precipitation, evapotranspiration, and discharge at the outlet of exorheic basins. WBEs have been conducted mostly on large basin, for example, on the Amazon (Eom et al., 2017; Nie et al., 2016; Pellet et al., 2021; Syed et al., 2005), the Congo (Crowley et al., 2006), the Mississippi (Munier et al., 2014; Syed et al., 2005), the Indus (Hussain et al., 2020; Zhu et al., 2021), the Mackenzie (Morrow et al., 2011), but also on smaller ones like the Garonne (Biancamaria, 2019). Studies have also been performed on multiple basins simultaneously. For example Gao et al. (2010) analysed nine basins in the United States while Jiao et al. (2025) investigated ten basins in Asia. Others focus on major basins on each continent (Abolafia-Rosenzweig et al., 2021; Aires and Pellet, 2024a; Sahoo et al., 2011). To our knowledge, Lehmann et al. (2022) is the only study that has attempted a quasi-global approach, processing 189 basins. That study concluded that no single dataset allows for satisfactory results in calculating global-scale WBEs. Indeed there is considerable dispersion among datasets,



65 especially for P (Maggioni et al., 2016; Masunaga et al., 2019; Sun et al., 2018) and ET (Bai et al., 2022; Ma et al., 2024; McCabe et al., 2016; Talsma et al., 2018).

Concerning discharge, the spatial density of freely available discharge measurement stations is highly uneven worldwide, and their number has been declining since the 1980s (Alsdorf et al., 2007; Crochemore et al., 2020; Fekete et al., 2015; Hannah et al., 2011; Voroesmarty et al., 2001). Consequently, discharges at the outlets of some major basins are no longer available in open-access databases, making global-scale WBE attempts difficult. A common alternative is to use monthly modelled runoff, which is usually provided on a longitude–latitude grid. Runoff is the portion of precipitation that flows over the land surface into rivers, lakes or the ocean. It occurs when the soil is saturated (Dunne and Black, 1970) or when rainfall intensity exceeds the infiltration capacity of the soil (Horton, 1945). Runoff therefore does not equal discharge but can, however, be simulated across all emerged lands. For this reason many studies compute WBEs by simply summing runoff at the basin scale from Land Surface Models (LSM) such as the Global Land Data Assimilation System (GLDAS) (Nie et al., 2016; Pan et al., 2012; Zhu et al., 2021) or from reanalysis such as the fifth generation of the European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) or the second version of Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) (Lehmann et al., 2022; Li et al., 2026). This substitution of discharge at the outlet with runoff is a common source of errors and can distort WBE conclusions. First, models poorly account for direct anthropogenic impacts as dams and irrigation, as well as for climate change impacts (Galelli et al., 2025; Salwey et al., 2023; Shiklomanov and Lammers, 2009). These effects accumulate downstream, increasing the error on outlet discharge (Ghiggi et al., 2019; Jiao et al., 2025; Miao and Wang, 2020). The effect can be particularly large in heavily anthropogenically altered basins such as the Mekong (Räsänen et al., 2017), the Yellow River (Wang et al., 2025, 2016; Yu et al., 2013), or the Indus (Mahar et al., 2022). Secondly, to our knowledge, most runoff-based WBEs studies do not apply a routing model, or if they do, they do not indicate the method used. The omission of routing is often justified by two assumptions: (i) discharge is the weakest component of the water cycle (Lehmann et al., 2022) and (ii) the effect of propagation at a monthly scale should be negligible (Lorenz et al., 2014). Yet, median propagation speed ranges from 0.5 m.s⁻¹ to 2.4 m.s⁻¹ (Allen et al., 2018). At these speeds the delay of the peak flood at a river outlet can correspond to travel distances of 1 300 km–6 200 km within a month. This may be particularly significant for long rivers such as the Nile (7,088 km), the Amazon (6,575 km), the Yangtze (6,236 km), and the Mississippi (6,084 km), whose lengths were assessed by Liu et al. (2009). Moreover, dams and reservoirs alter natural discharge patterns in a basin-specific way (Hanasaki et al., 2006; Yoon et al., 2013). Using a simple runoff sum without routing in large, lake-rich basins (e.g., the Nile) therefore raises doubts about the validity of the resulting WBE. More generally, relying on Model-Based Discharge from aggregated runoff (MBD) rather than on Observations-Based Discharge (OBD) does not necessarily ensure a reliable water-budget closure.

95 Recent analyses illustrate the magnitude of the problem. Jiao et al. (2025) showed, for ten basins, that the Coefficient of Variation (CV) of discharge derived from runoff datasets is generally twice that of P and ET datasets. Lehmann et al. (2022) reported also a higher CV for these discharge and it states that compensation phenomena can occur within the water budget equation, which can yield a high level of closure for the wrong reasons. Luo et al. (2026) evaluated the error made by



substituting observed discharge with four unrouted LSM runoff on 62 basins. It obtained a discrepancy between 1% and 20% in precipitation estimation with an average of 6% and a discrepancy between 0.5% and 38% in evapotranspiration estimation with an average of 7%. It also showed a greater impact on humid basins than on arid basins.

Therefore, this study focuses on the following questions: (i) How can the results of Luo et al. (2026) be refined by using a larger, more comprehensive set of widely-used runoff, precipitation, and evapotranspiration datasets? (ii) Does a simple runoff-routing scheme reduce compensation errors in the water-budget closure? (iii) In which climate regimes are these errors most pronounced? To address these, we compare WBEs derived from routed and unrouted runoff data with discharge calculations based on direct measurements (including both in situ-based discharge and satellite-based discharge). The satellite-based discharge data enable us to incorporate basins of major societal relevance in this study and to assemble a sufficiently large sample for a robust statistical analysis across the principal climate types.

Satellite-based discharges are employed to compensate for the spatial and temporal heterogeneity of in situ-based discharges. Four categories of satellite-derived discharge methods are recognised (Lin et al., 2024). (1) Hydraulic parameter method (Birkinshaw et al., 2014; Bjerklie et al., 2003; Tarpanelli et al., 2015). (2) Hydrological model (Huang et al., 2022; Reichle et al., 2014; Sun et al., 2012). (3) At-Many-Stations Hydraulic Geometry (AMHG) (Brinkerhoff et al., 2019; Gleason et al., 2014; Ishikawa et al., 2025). (4) Empirical regression. (Elmi et al., 2021; Filippucci et al., 2022; Frappart et al., 2015; Tarpanelli et al., 2017; Tourian et al., 2013). Empirical regression methods, based on the correlation between hydrological parameters and in situ river discharge is generally considered the most accurate (Lin et al., 2024). Among empirical regression methods, those based on the calculation of Water Surface Elevation (WSE) are particularly suitable for river discharge calculations (Tarpanelli et al., 2015).

Radar altimetry is a widely used technique for estimating WSE. The first calculation of WSE on continental surfaces from a satellite altimetry mission was carried out by Koblinsky et al. (1993). This was achieved by retracking Geosat data at four stations along the Amazon River. Later, Birkett (1998) demonstrated that Topex/Poseidon could detect river water levels with a root-mean-square error of 11–25 cm. Subsequent studies built rating curves from in situ discharge and Topex/Poseidon WSE for the Ob (Kouraev et al., 2004), the Amazon (Zakharova et al., 2006), or the Chari (Coe and Birkett, 2004). Papa et al. (2010) combined Topex, Envisat and ERS-2 data to derive discharge for the Ganga-Brahmaputra. Paris et al. (2016) proposed a method for calculating the rating curve with a Bayesian approach, while Zakharova et al., 2020 introduced a polynomial method and Tourian et al. (2013) a quantile-based method for situations where altimetry and gauge records do not overlap. More recently, Saemian et al (2025) calculates altimetry discharge on 8,730 stations. However, most of these studies estimate discharge without combining altimetry missions to generate long-term series. The recent European Space Agency Climate Change Initiative (CCI) river-discharge project provides several remote-sensing-based products, among them altimetry-derived discharges obtained by concatenating measurements from numerous space missions across 18 basins to achieve the longest possible time series (Gal et al., 2024). Because the product delivers discharge estimates as close as possible to the basin outlet, it is especially relevant for basin-scale water-budget calculations.



In this paper, we evaluate whether calculating discharge using space altimetry and in situ data yields more reliable WBEs than those based on modelled runoff. Specifically, we (i) quantify the difference between OBD and MBD (both routed and unroun135 ted), (ii) assess the extent of error compensation within the water-budget, (iii) examine the resulting bias in WBE conclusions, and (iv) identify the types of basins where these errors are most pronounced. The analysis integrates in situ gauge records with radar-altimetry measurements for 82 basins selected for their diversity in size, climate, discharge magnitude and societal importance.

Section 2 describes the selected basins and the datasets used for TWS, precipitation, evapotranspiration, runoff and discharge.

140 Section 3 outlines the methods employed to compute the various discharge estimates, to assess datasets variability and to perform the WBE for each basin. Section 4 presents and discusses the results. Section 5 concludes the study and proposes directions for future research.

A table of acronyms is provided in the Appendix J.

145 2 Data

2.1 Study period and area

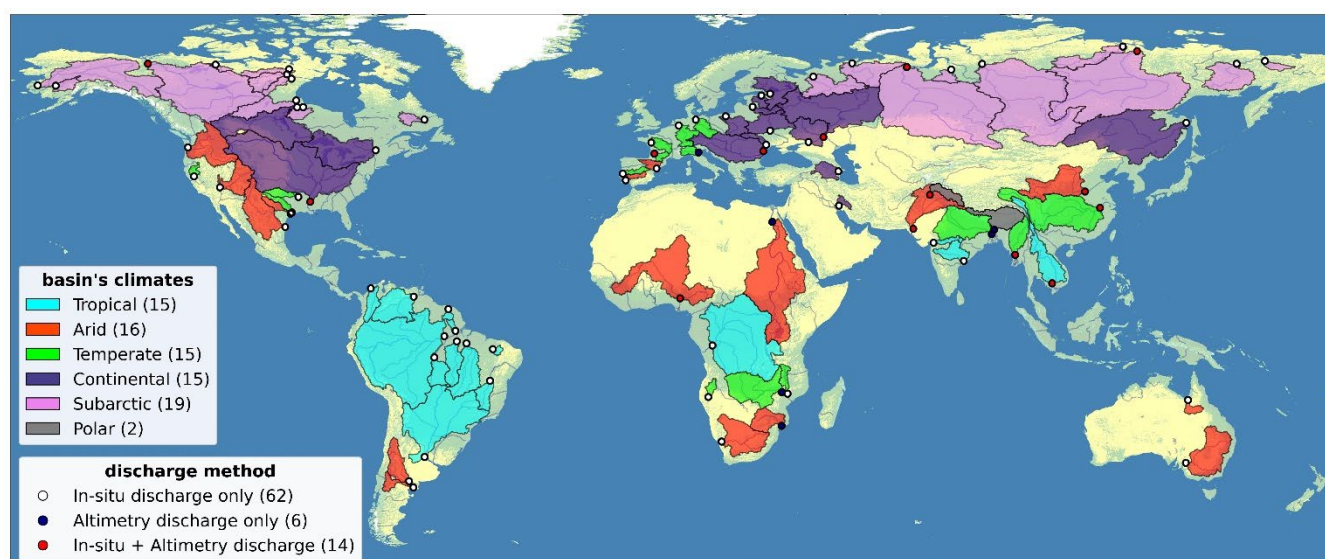
The study focuses on the period from 2003 to 2019, mainly due to the availability of TWS from GRACE and GRACE-FO missions (2002-2025) and the availability of a maximum number of P and ET products.

The basins were selected according to three criteria: (i) taken together they span a wide range of size, climate and latitude; (ii) 150 they provide several consecutive years of discharge observations for the study period (whether in situ or altimetry-derived); and (iii) they have a minimum area that matches the spatial resolution of GRACE and GRACE-FO. Gravimetry-based TWS variations can be resolved with an error of about 2 cm in terms of equivalent water height for basins larger than 63,000 km² (Vishwakarma et al., 2018). Biancamaria (2019) reported a relatively high level of water budget closure for the Garonne, a temperate basin of roughly 50 000 km². Consequently, we retained 82 basins with areas greater than 50 000 km², distributed 155 across all continents except the polar ice caps (Fig. 1). Together, they provide a representative coverage of tropical, arid, temperate, continental, and sub-arctic regions. The Brahmaputra and Upper Indus basins were added in order to include at least two polar-climate catchments.

A mask is calculated for each basin using the level 5 of the HydroBASIN database (Lehner and Grill, 2013) by selecting only the upstream area of the gauge station. This means removing the fraction of the total basin downstream the gauge (Fig. 2a). 160 The 82 sub-basin masks were then converted to a 1-degree resolution by selecting pixels that include at least 10% of the basin. For each pixel of every sub-basin mask, the ratio of the basin area contained in the pixel to the pixel's total area is used as a weighting factor to aggregate the gridded datasets of TWS, precipitation (P), evapotranspiration (ET) and runoff (Fig. 2a). Climates zones over the 82 exorheic basins are computed with a recent Köppen-Geiger classification for the period 1991-2020 (Beck et al., 2023). The classification includes thirty climate classes, each identified by a three-letter code. For this study the



165 classes were grouped into six major climate types: Tropical (classes Axx), Arid (Bxx), Temperate (Cxx), Continental (Dxx, except Dfc, Dwc, Dfd, and Dwd), Subarctic (Dfc, Dwc, Dfd, Dwd), and Polar (Exx). A basin's climate is assigned according to the dominant climate type, the one that occupies the largest share of the basin's surface (Fig. S1 to Fig. S3 in supplementary material). 43 of the 82 basins have a dominant climate covering more than 80 % of their area, and 70 basins have a dominant climate covering more than 50 % of their area. Only three basins have a dominant climate that accounts for less than 40 % of the basin surface: Columbia River (38%), Ob (38%), and Yellow River (37%). In this study, climate characteristics were assigned to each basin using only the upstream area of the gauge station. Consequently, the resulting climate estimates may not fully represent the climatic conditions across the entire basin.



175 **Figure 1: Attributed climates (polygons in colour) and the methods used for discharge observations for each basin selected in this study (white, black or red dots). The numbers in parentheses indicate the number of basins belonging to each climate type and the number of basins for which each observation method is applied**

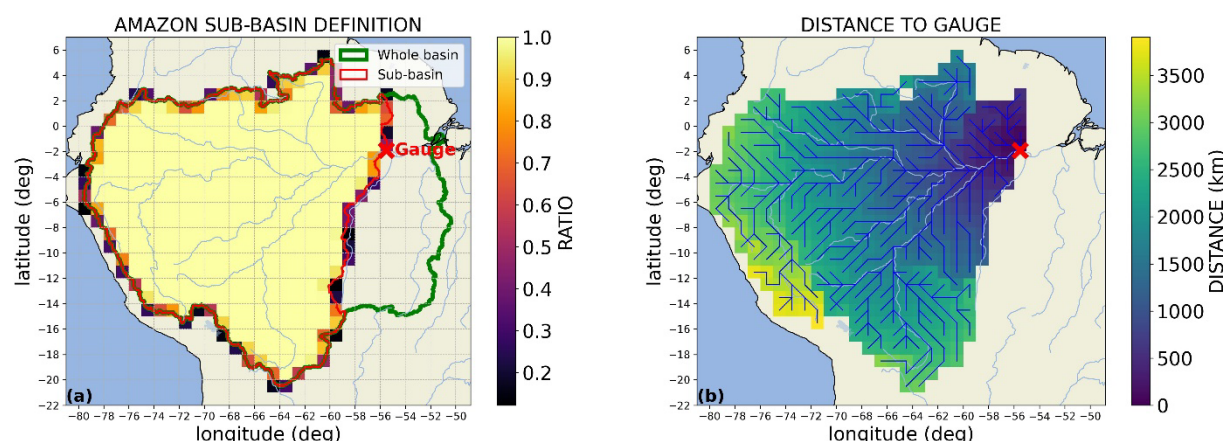


Figure 2: (a) Amazon sub-basin and sub-basin mask weighting factors within each 1 deg pixel, (b) routing and distance to the station computation for each pixel at 1 deg for the Amazon sub-basin

2.2 Datasets

This section presents the datasets used to calculate the WBE: total water storage variation (2.2.1), precipitation (2.2.2), evapotranspiration (2.2.3), runoff (2.2.4), and discharge (2.2.5). They are summarised in Table 1. It should be noted that we separate datasets based on Land Surface Model (LSM) without data assimilation (which we will simply refer to as LSM) and datasets based on reanalysis. Indeed, reanalysis use Land Surface Model (LSM) to simulate terrestrial surface processes by assimilating atmospheric observation data.

2.2.1 Total water storage

Each GRACE (2002-2017) and GRACE-FO (2018-present) space mission consists of two identical satellites flying in formation at a separation distance around 220 km and an altitude of approximately 500 km. The satellites continuously measure the distance between them with very high precision using a band K radar or a microwave measurement system. When the satellites pass over a region where the mass varies, this locally changes the gravitational force and causes slight variations in the inter-satellite distance. Measuring this distance enables the assessment of time-variations in the gravity field. The gravity field may be computed by decomposing it into a series of spherical harmonics or by dividing the Earth's surface into a grid of blocks or cells, each representing a mass concentration (MASCONs). These gravity field variations, once corrected for solid-Earth mass variations as earthquakes or glacial isostatic adjustment, are due to water mass transport at the Earth's surface. Note that there is a data gap between GRACE and GRACE-FO missions during the period 2017-2018.

This study uses the TWS product from the Service d'Activités en Gravimétrie Spatiale et Applications (SAGSA), version 2.1 (Blazquez et al., 2018) as well as 3 GRACE MASCON solutions: from the Jet Propulsion Laboratory (JPL) (RL06.1 CRI



Filtered Version 03) (Landerer and Swenson, 2012; NASA/JPL, 2021), from the Center for Space Research (CSR) (RL06 version 02) (Save et al., 2016), and from the Goddard Space Flight Center (GSFC) (RL06 V2.0) (Loomis et al., 2019).

SAGSA TWS is a monthly ensemble product calculated at a $1^\circ \times 1^\circ$ spatial resolution from spherical harmonics coefficients.

205 Here we choose a sub-ensemble that combines: (i) five GRACE data-processing methods (CSR, JPL, Technical University of Graz (TUGRAZ), French space agency (CNES), and German Research Centre for Geosciences (GFZ)), (ii) two Global Isostatic Adjustment (GIA) models (ICE6GD and Caron_2018), and (iii) two level of Decorrelation and Denoising Kernel filtering (DDK3 and DDK6). Only one realization in geocenter (TN13) and one C20 (TN14) are selected in this study (to reduce the number of combinations to 20 instead of 120 for the whole SAGSA V2.1 product). This assumption is based on the
210 fact that their impacts of a single basin are negligible with respect to the uncertainty of the other post-processing. We assume the uncertainty in gravimetry-based TWS variation to be the spread among the different combinations of this sub-ensemble.

2.2.2 Precipitation

Precipitation is the portion of the atmospheric water cycle that reaches the Earth's surface in the form of liquid water (rain) or
215 frozen particles (snow, hail). Complementing the global observational data network (e.g., rain gauges), which remains sparse in many regions (Kidd et al., 2017), various gridded precipitation products have been developed utilizing different techniques and data sources, incorporating both physical and statistical approaches.

In this study, we use precipitation data from the Frequent Rainfall Observations on GridS (FROGS) database (Roca et al., 2019). FROGS is an ensemble product that gathers a large set of precipitation products, including satellite, ground-based
220 observations, and reanalysis products, and provided in the same structure (temporal and spatial resolutions, i.e., at daily and 1° resolution, respectively). Among the products available in FROGS, we retained only those that are truly global, i.e., that provide coverage for latitudes higher than $\pm 60^\circ$, and that span the entire study period (2003 – 2019). These include the Global Precipitation Climatology Project (GPCP) (Huffman et al., 2023), NASA Integrated Multi-satellitE Retrievals for GPM (IMERG) (Huffman et al., 2020), Global Interpolated RAINfall Estimation (GIRAFE) (Konrad et al., 2025), The Global
225 Unified Gauge-Based Analysis Climate Prediction Center (CPC) (Schneider et al., 2013; Xie et al., 2007), Global Precipitation Climatology Center (GPCC) (Schneider et al., 2022), the Japanese 55-year reanalysis (JRA-55) (Kobayashi et al., 2015), the Japanese Reanalysis for Three Quarters of the Century (JRA-3Q) (Kosaka et al., 2024), version 2 of the Modern-Era Retrospective analysis for Research and Applications (MERRA2) (Gelaro et al., 2017), Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010), and the fifth generation of European Centre for Medium-Range Weather Forecasts (ERA5) (C3S,
230 2018) (Table 1). Regarding IMERG, two versions are used: V07 Final Calibrated (FC) with rain gauge calibration and V07 Final Uncalibrated (FU), in order to have at least two products based solely on satellite data. An additional version of GIRAFE with bias correction (GIRAFE-C) is also considered. In this case, the bias adjustments were performed on daily basis and pixel-by-pixel, considering the GPCC precipitation through the Quantile Delta Mapping (QDM) approach (Cannon et al., 2015), which accounts for spatial and seasonal variabilities.



235 The 12 selected precipitation products cover a wide range of sources including rain gauge measurements (GPCC, CPC), reanalysis (ERA, CFSR, MERRA2, JRA-55), satellite measurements with GPCC rain gauge calibration (GPCP, IMERG FC, GIRAFE-C), and satellite measurements without rain gauge calibration (IMERG FU, GIRAFE).

For each basin and each product, the 1° FROGS pixels are selected using the sub-basin mask defined in Sect. 2.1. The daily time series of each pixel are first scaled by the mask weighting factor and then aggregated to generate monthly time series.

240 Finally, the monthly series from all selected pixels are summed across the basin to obtain the total monthly precipitation at the basin scale.

2.2.3 Evapotranspiration

Evapotranspiration is the combined flux of water vapor from the land surface to the atmosphere, encompassing both the evaporation of water from soil and water bodies and the transpiration of water through plant leaves. There are many ways to calculate evapotranspiration; among them are the Penman-Monteith equation (Monteith and Lennox, 1965; Penman and Keen, 1948), and the surface-energy-balance method (Bastiaanssen et al., 1998). 12 global evapotranspiration products available for the 2003-2019 period have been selected for this study (see Table 1). These are ERA5-Land (Muñoz-Sabater et al., 2021), MERRA-2 (Global Modeling And Assimilation Office and Pawson, 2015), JRA-55 (Kobayashi et al., 2015), 3 variants of the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004): Variable Infiltration Capacity (VIC) (Beaudoin, 2020; Hamman et al., 2018; Liang et al., 1994), Noah (Beaudoin et al., 2020; Chen et al., 1996; Ek et al., 2003; Koren et al., 1999) and Catchment Land Surface Model (CLSM) (Koster et al., 2000; Li, 2020), versions 4.2a and 4.2b of the Global Land Evaporation Amsterdam Model (GLEAM) (Miralles et al., 2025), version 2 of the Breathing Earth System Simulator (BESS V2) (Li et al., 2023), version 5 of Operationnal Simplified Surface Energy Balance (SSEBop V5) (Senay et al., 2013), version 2 of Penman Monteith Leuning (PML V2) (Gan et al., 2018; Zhang et al., 2019) and Synthesized ET (Elnashar et al., 2021).

255 These evapotranspiration products are based on reanalysis (ERA5-Land, MERRA2, JRA55), LSM (GLDAS VIC, Noah, and CLSM), satellite data (GLEAM, BESS, SSEBop, PMLV2), and an ensemble product Synthesized ET. These products use different calculation methods. For example, GLEAM 4.2a and 4.2a are driven by the Penman-Monteith equation (whereas earlier versions relied on the Priestley-Taylor formulation). PML V2 uses the Penman-Monteith-Leuning equation (Leuning et al., 2008; Zhang et al., 2010). ERA5-Land computes ET through a surface-energy-balance equation coupled with a water-budget closure, while SSEBop solves the surface energy balance using surface temperature and vegetation conditions. Note that the version 2.1 of GLDAS used in this paper does not employ data assimilation and operates in an open-loop mode, which distinguishes its functioning from the reanalysis ERA5-Land, MERRA2, and JRA-55.

ET datasets were retrieved with a monthly resolution. The 1-degree mask was applied to the data using the same method as for precipitation, keeping the initial spatial resolution (see Fig. S4). Due to their spatial resolution, MERRA-2 (0.5°x 0.625°) and JRA-55 (0.5625° x 0.5625°) data were interpolated to 1 degree before applying the mask and the per-pixel weighting factor. Finally, the pixels were summed to obtain the monthly time series.



2.2.4 Runoff

270 Runoff was calculated from 4 LSM, 3 reanalysis and a machine learning model (see Table 1). Three GLDAS V2.1 monthly variants have been selected (Rodell et al., 2004): VIC (Beaudoin, 2020; Liang et al., 1994), Noah (Beaudoin et al., 2020; Chen et al., 1996; Ek et al., 2003; Koren et al., 1999) and CLSM (Koster et al., 2000; Li, 2020). The fourth selected LSM is the Interaction Sol-Biosphère-Atmosphère (ISBA) model (Munier and Decharme, 2022). The three selected reanalysis are
275 Reconstruction product (GRUN) relies on a machine-learning approach that estimates runoff from precipitation, temperature and observed runoff in small catchments (Ghiggi et al., 2021). We excluded GLDAS V2.2 to prevent circular reasoning, since it already incorporates GRACE data assimilation.

The 1-degree Total Runoff Integrating Pathways (TRIP) routing (Oki and Sud, 1998), as well as a 1-degree digital elevation model (Earth Resources Observation and Science (EROS) Center, 2010) (EROS) have been used to convert runoff into
280 discharge.

2.2.5 Discharge

In-situ discharge measurements were obtained primarily from the Global Runoff Data Centre (GRDC). Exceptions include: (1) the Amazon, Orinoco, Maroni, and Congo basins, for which data were sourced from the HYdro-geochemistry of the
285 AMazonian Basin project (HYBAM); (2) the Indus basin, using discharge from Indus Flood report documents of the Federal Flood Commission of Pakistan; and (3) the Mekong basin, where data were supplied by the Mekong River Commission. Altimetry-based discharge is taken either from the RD-ALTI discharge product of the ESA CCI discharge project (Gal et al., 2024), or calculated directly from water surface elevation (WSE) altimetry data acquired by the Topex, ERS2, Envisat, Jason-1, Jason-2, Saral, Jason-3, Sentinel-3A, Sentinel-3B missions, which are provided by the Centre de Topographie de l'Océan
290 et de l'Hydrosphère (CTOH) (Frappart et al., 2021).



Table 1. Datasets used in this study

Name	Method	Spatial Resolution	Reference
Terrestrial Water Storage (TWS)			
SAGSA	Spherical harmonics	1 ° x 1 °	Blazquez et al., 2018
GFSC	Mascon	0.5 ° x 0.5 °	Loomis et al., 2019
CSR	Mascon	0.5 ° x 0.5 °	Save et al., 2016
JPL	Mascon	0.25 ° x 0.25 °	Landerer and Swenson, 2012
Precipitation (P)			
GPCP	Satellite + rain gauges		Huffman et al., 2023
IMERG V07B FC	Satellite + rain gauges		Huffman et al., 2020
GIRAFE C	Satellite + rain gauges		Konrad et al., 2025
IMERG V07B FU	Satellite		Huffman et al., 2020
GIRAFE	Satellite	1 ° x 1 °	Konrad et al., 2025
GPCC	Rain gauges	as from FROGS	Schneider et al., 2022
CPC	Rain gauges	database	Xie et al., 2007
JRA-55	Reanalysis	Roca et al, 2019	Kobayashi et al., 2015
JRA-3Q	Reanalysis		Kosaka et al., 2024
MERRA2	Reanalysis		Gelaro et al., 2017
CFSR	Reanalysis		Saha et al., 2010
ERA5	Reanalysis		C3S, 2018
Evapotranspiration (ET)			
GLEAM V4.2a	Satellite+Model	0.1 ° x 0.1 °	Miralles et al., 2025
GLEAM V4.2b	Satellite+Model	0.1 ° x 0.1 °	
BESS V2	Satellite	0.05 ° x 0.05 °	Li et al., 2023
SSEBop V5	Satellite	1 km x 1 km	Senay et al., 2013
PML V2	Satellite	0.05 ° x 0.05 °	Zhang et al., 2019
Synthesized ET	Ensemble	1 km x 1 km	Elnashar et al., 2021
Runoff (R)			
ISBA	Land Surface Model	0.5 ° x 0.5 °	Munier and Decharme, 2022
GRUN	Machine Learning	0.5 ° x 0.5 °	Ghiggi et al., 2021
Evapotranspiration (ET) & Runoff (R)			
ERA5-Land	Reanalysis	0.1 ° x 0.1 °	Muñoz-Sabater et al., 2021
MERRA2	Reanalysis	0.5 ° x 0.625 °	Gelaro et al., 2017
JRA-55	Reanalysis	0.5625 ° x 0.5625 °	Kobayashi et al., 2015
GLDAS VIC V2.1	Land Surface Model	1 ° x 1 °	Rodell et al., 2004
GLDAS Noah V2.1	Land Surface Model	1 ° x 1 °	
GLDAS CLSM V2.1	Land Surface Model	1 ° x 1 °	
Discharge (Q)			
GRDC	Gauges	station	https://portal.grdc.bafg.de/



HYBAM	Gauges	station	https://hybam.obs-mip.fr/
Mekong River Commission	Gauges	station	https://www.mrcmekong.org/
Federal Flood Commission	Gauges	station	https://www.ffc.gov.pk/
ESA CCI	Altimetry + Gauges	virtual station	Gal et al., 2024
Water Surface Elevation (WSE)			
CTOH	Altimetry	virtual station	Frappart et al., 2021

3 Methods

For this study, we have developed and adapted several methods, which we will introduce in this section. First, we will define the metrics used to validate the methods and evaluate the results (Sect. 3.1). Second, we will present the methodology for calculating a reference discharge based on observations (in situ-based discharge and/or altimetry-based discharge) and also the methods for computing discharges from runoff data, with or without routing scheme (Sect. 3.2). Third, we will assess the reliability of the observations-based discharge as well as the methods for comparing this discharge with those calculated from runoff (Sect. 3.3). Fourth, we will detail WBE computation and the method for deriving the annual volumes of P and ET from it (Sect 3.4). Finally, we will describe the method for assessing the dispersion of each component of the water budget (Sect. 3.5).

3.1 Metrics

In this section we present the various metrics used in this study.

The Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) is a metric widely used in hydrology to evaluate the performance of a hydrological model (Eq. 1). It allows comparing an observed time series (Obs) with a time series simulated by a model (Mod). The NSE ranges from $-\infty$ to 1. A NSE of 1 indicates a perfect fit between the observed and simulated series. A NSE of 0 means that the model's simulation is no better than using the observed mean as a predictor. Negative NSE values indicate very poor agreement between model and the observations. Simulations are considered behavioural if $NSE > 0.5$ (Knoben et al., 2019).

$$NSE(\text{Obs}, \text{Mod}) = 1 - \frac{\frac{1}{T} \sum_{t=1}^T (\text{Mod}(t) - \text{Obs}(t))^2}{\frac{1}{T} \sum_{t=1}^T (\text{Mod}(t) - \overline{\text{Obs}(t)})^2} \quad (1)$$

where $\overline{\text{Obs}(t)} = \frac{1}{T} \sum_{t=1}^T \text{Obs}(t)$ is the long term mean of the observed time series.

The Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) is another goodness-of-fit indicator used in hydrology for comparing simulated and observed series (Eq.2). Althoff and Rodrigues (2021) consider the KGE to be a more reliable objective-function metric than the NSE, especially because the NSE is highly sensitive to peak flow. KGE also ranges from $-\infty$ to 1. A KGE of -



0.41 is a commonly used reference threshold to distinguish performances that are inferior to those of the mean-discharge model (corresponding to an NSE = 0). Simulations are considered behavioural if $KGE > 0.3$ (Knoben et al., 2019).

$$KGE(\text{Obs}, \text{Mod}) = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{\text{Mod}}}{\sigma_{\text{Obs}}} - 1\right)^2 + \left(\frac{\mu_{\text{Mod}}}{\mu_{\text{Obs}}} - 1\right)^2} \quad (2)$$

320 With r the Pearson correlation of the Mod and Obs time series, σ_x the standard deviation of x , and μ_x the mean of x .

For a clearer assessment of the differences between the Mod and Obs time series, this study also decomposes the KGE into its three component terms: the correlation term (KGE_{corr} , Eq. 3), the mean term (KGE_{mean} , Eq. 4), and the standard-deviation term (KGE_{std} , Eq. 5). The nearer each of these terms is to zero, the more it indicates that the two signals share, respectively, a high
325 level of correlation, a low bias, and a low standard-deviation discrepancy.

$$KGE_{\text{corr}}(\text{Obs}, \text{Mod}) = (r-1)^2 \quad (3)$$

$$KGE_{\text{mean}}(\text{Obs}, \text{Mod}) = \left(\frac{\mu_{\text{Mod}}}{\mu_{\text{Obs}}} - 1\right)^2 \quad (4)$$

$$KGE_{\text{std}}(\text{Obs}, \text{Mod}) = \left(\frac{\sigma_{\text{Mod}}}{\sigma_{\text{Obs}}} - 1\right)^2 \quad (5)$$

330 The coefficient of variation (CV) of a signal is defined as the ratio of the standard deviation to the mean. CV quantifies the spread of a data set relative to its mean, using the mean as a normalizing factor. Mean Coefficient of Variation (CV_m) can be computed for time-dependent vector (Eq. 6) and is commonly used to assess the dispersion of P, ET and Q datasets in water budget studies (Jiao et al., 2025; Lehmann et al., 2022).

$$CV_m(X_k) = \frac{\frac{1}{T} \sum_{t=1}^T \sigma_k(t)}{\frac{1}{T} \sum_{t=1}^T \mu_k(t)} = \frac{\sigma_k}{\mu_k} \quad (6)$$

335 With:

$$X_k(t) = \begin{bmatrix} X_{k,1}(t) \\ X_{k,2}(t) \\ \vdots \\ X_{k,N_k}(t) \end{bmatrix}, \mu_k(t) = \text{mean}_{i \in [1:N_k]} \begin{bmatrix} X_{k,1}(t) \\ X_{k,2}(t) \\ \vdots \\ X_{k,N_k}(t) \end{bmatrix} \text{ and } \sigma_k(t) = \text{std}_{i \in [1:N_k]} \begin{bmatrix} X_{k,1}(t) \\ X_{k,2}(t) \\ \vdots \\ X_{k,N_k}(t) \end{bmatrix}$$

However, because the CV_m is anchored to the mean of each variable, it cannot be used to compare the levels of variability among the different components of the water budget. Moreover, in most catchments the TWS variation crosses zero at least once a year, which would create a singularity when calculating a CV for the TWS variation. To overcome these limitations,
340 we introduce the Coefficient of Relative Variation (CRV). Considering a set of n time-dependent vectors, $CRV_j(X_k)$ of one of the n time-dependent vectors is defined in Eq. 8 as the ratio of the temporal mean of this vector standard deviation (i.e., σ_k in Eq. 6) to A_j , the mean amplitude of one of the n time-dependent vectors (Eq. 7). Unlike the CV, the CRV normalization factor is not the vector's mean, but a value that is the same for every time-dependent vector, which makes CRV comparable across vectors.



$$A_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \left(\max_{t \in [1:T]} X_{k,i}(t) - \min_{t \in [1:T]} X_{k,i}(t) \right) \quad (7)$$

$$CRV_j(X_k) = \frac{\sigma_k}{A_j} \quad (8)$$

The ratio of two CRV values equals the ratio of the corresponding mean standard deviations and, thus, the ratio of the two CV_{ms} multiplied by the ratio of the mean values (Eq. 9).

$$\frac{CRV_j(S_1)}{CRV_j(S_2)} = \frac{CV_m(S_1)}{CV_m(S_2)} \times \frac{\mu_1}{\mu_2} \quad (9)$$

3.2 Discharge computation

3.2.1 Altimetry-based discharge

Altimetry data were processed, mission by mission, using the ALTiS tool from the CTOH (<https://gitlab.com/ctoh/altis>) to produce time series of WSE. Continuous series of WSE and discharge were then generated using a methodology similar to that of the ESA CCI river discharge precursor project (ESA Climate Change Initiative River Discharge Precursor, 2025), which combines two different approaches (Gal et al., 2024):

- Overlapping dataset: When temporal overlap exists between in situ and WSE data, a Bayesian approach is used to derive the rating curve. Performance is assessed with the NSE and KGE, treating the altimetry-based discharges time series as the simulated series and the in situ discharge as the observed ones (Sect. 3.1). This approach was applied to the following basins: Indus (NSE = 0.71, KGE = 0.31), Mekong (NSE = 0.96, KGE = 0.98), Upper Indus (NSE = 0.57, KGE = 0.67), Volga (NSE = 0.86, KGE = 0.91), and Yellow River (NSE = 0.66, KGE = 0.77).
- Non-overlapping datasets: When no temporal overlap exists, a quantile-based matching method is used to relate discharge and WSE. These quantile-derived scatter plots are subsequently used as input for a Bayesian inversion to derive the rating curve. This methodology was applied to the Nile and Yangtze. Regarding the Nile, the in situ data from GRDC (1973-1984) are posterior to the construction of the Aswan Dam (1973), making the use of this method legitimate. For the Yangtze River, the in situ data (1946 – 2004) precede the commissioning of the Three-Gorges Dam (2003 – 2012), yet the method remains appropriate because (i) the altimetry period we employed (2008 – 2019) does not cover the reservoir-filling phase (2006), and (ii) the gauge at Datong is situated far downstream, so its discharge is strongly influenced by major tributaries such as the Yuan Jiang, Xiang, and Han rivers.

Once the daily altimetry-based discharge has been computed, the corresponding monthly discharge is obtained by averaging the instantaneous daily values over the month. For basins where the altimetry-derived discharges overlap with in-situ measurements, the performance of the monthly altimetry-based series was reassessed using the NSE and KGE metrics. The results are as follows: Indus (NSE = 0.73, KGE = 0.3), Mekong (NSE = 0.99, KGE = 0.79), Upper Indus (NSE = 0.55, KGE = 0.67), Volga (NSE = 0.88, KGE = 0.83), and Yellow River (NSE = 0.57, KGE = 0.73).

The daily discharges from the ESA CCI river discharge were processed in the same way in order to obtain monthly altimetry-based discharges.



The actual estimation of in situ discharge, for example by Acoustic Doppler Current Profiler (ADCP), requires making several back-and-forth trips on the river to minimize errors and takes time (Paris, 2015; Petrie et al., 2013). Thus, in most cases, in situ discharge data are calculated by measuring a water height and applying a rating curve previously evaluated from ADCP measurements. The altimetry-based discharge computation uses the same principle: deriving a discharge from a water height and a rating curve (Kouraev et al., 2004).

3.2.2 Observations-based discharge (OBD)

OBD will be our reference monthly discharge. It is computed using observations as follows:

- By default, OBD is defined as the monthly in situ discharge at the gauge station.
- For basins without available in situ data for the study period, OBD is the altimetry-based discharge (Sect. 3.2.1). It is computed from the CCI river monthly discharge product (Brahmaputra, Ganges, Limpopo, Po, Zambezi) or with the quantile method (Nile) (Sect. 3.2.1).
- For basins where the monthly in situ discharge has gaps and altimetry-based discharge is available, OBD is the monthly in situ discharge with gaps filled with monthly altimetry-based discharge using data from the CCI river discharge (Lena, Maroni, Mississippi, Niger, Ob, Colville) or computed for this study (Indus, Mekong, Volga, Yangtze, Yellow River). Since altimetry measurements do not always provide valid measurements, some discharges may still have gaps. We chose not to fill these gaps mathematically to have the most reliable data possible.

We calculated the percentage of missing discharge data. For 83% of the basins, we obtained more than 85% of valid data for the OBD over the period 2003-2019. Only two basins have less than 70% of valid data: Rio Grande 47% and Zambezi 65%. The use of altimetry-based discharge as a complement to in situ data allows obtaining discharge measurements for an average of 10 additional basins each month during the study period (Fig. S23).

3.2.3 Model-based discharge (MBD)

Monthly discharges were calculated from runoff data in two different ways. The first method is similar to the one used for P (Sect. 2.2.2) and ET (Sect. 2.2.3) and is performed by applying the basin mask (with the per-pixel weighting factor defined in Sect. 2.1) to the runoff datasets and then summing the time series of each pixel (i.e. without routing scheme). These discharges are labelled Total Runoff Discharge (TRD). Most studies that calculate WBE with runoff data rely on this approach (Lehmann et al., 2022; Li et al., 2026).

The second method, called Routed Runoff Discharge (RRD), uses, in addition to the basin mask and the per-pixel ratio, the TRIP routing model (Oki and Sud, 1998). For pixels that do not drain into the sub-basin mask (i.e., some border pixels with a



low ratio), the 1-degree EROS digital elevation model is used to find the adjacent pixel within the mask with the lowest elevation. This computation can, for some border pixels with very low ratio, lead to non-physical crossings in the routing (see Fig. 2b for lon=-77 and lat=-10), but without significant impact on the discharge. Propagation speed (S_{prop}) is assumed to be constant over the entire basin, as Nohara et al. (2006) did when they employed TRIP routing to assess the impact of climate change on discharges. Arora and Boer (1999) reported average speeds ranging from 0.14 m.s^{-1} to 2.13 m.s^{-1} using a comparable simplified routing model applied to 20 large basins (all of which are included in the present study). Allen et al. (2018) found a median propagation speed between 0.5 m.s^{-1} and 2.4 m.s^{-1} for 20 000 gauge stations in the United States. In this study the velocity is calculated separately for each sub-basin. The selected S_{prop} value is the one that maximizes the correlation between the RRD monthly climatology and the OBD monthly climatology constrained to the interval $[0.1, 4] \text{ m.s}^{-1}$. The methodology to compute RRD is detailed in Appendix I.

In practice, propagation speed is not uniform across a catchment. It depends on the slope as well as on the discharge. Moreover, the pixels are not independent of each other. Nevertheless, we consider that using a fixed speed and a linear approach is an acceptable approximation for a monthly calculation. Because the routing scheme does not account for meanders, S_{prop} is necessarily lower than the river's effective mean velocity.

Both TRD and RRD are referred to as Model-Based Discharge (MBD) in the rest of the study.

We therefore calculate 17 monthly discharges for each basin:

- 1 OBD calculated solely with direct observations (in situ and/or altimetry).
- 8 TRD calculated without routing as the sum of the runoff of each model.
- 8 RRD calculated with 1-degree routing and a constant propagation speed that optimizes the climatology.

Calculating the propagation speed that maximizes the correlation between RRD and OBD monthly climatologies reveals inconsistencies among models (Fig S5). ERA5-Land and GRUN tend to give a higher S_{prop} speed (around 1.2 m.s^{-1}) whereas GLDAS, ISBA, MERRA2, and JRA55, yield average values between 0.2 m.s^{-1} and 0.5 m.s^{-1} . Propagation speed results are detailed in the supplementary material (Fig. S6 to Fig. S12).

3.3 Discharge evaluation

We present in this section the method for evaluating the different discharges considered in this study. First, we aim to verify the quality of the in situ-based discharge (Sect. 3.3.1) and then of the altimetry-based discharge (Sect. 3.3.2). The goal is to check whether OBD can truly be considered a reliable reference (Sect. 3.3.3). Finally, we will present the metrics used to compare MBDs with OBD (Sect. 3.3.4).



3.3.1 In situ-based discharge

440 To our knowledge, very few studies have quantified the uncertainties associated with in-situ discharge measurements. Carter and Anderson (1963) estimated that the precision of river-discharge measurements obtained with the current-meter method is about 2 %. Clarke (1999) suggests taking into account uncertainties in the calculation of the rating curve coefficients. Shiklomanov et al. (2006) examined six major Siberian Arctic rivers (Ob, Lena, Yenisei, Kolyma, Severnaya Dvina, and Pechora) and found that rating-curve-based discharge estimates vary from 10–25 % for low flows to 2–5 % for high flows, 445 with the smallest errors occurring at average flow conditions. (Sahoo et al., 2011) considers an RMS error of 7%, while (Abolafia-Rosenzweig et al., 2021) proposed a discharge-dependent error that changes non-linearly: from 28.8 % for low flows to 2.3 % during flood periods.

Building on these findings, we adopt the following RMS-error specifications for our study: 20 % for the lowest discharges, 2 % for average discharges, and 5 % for the highest discharges. Inspired by the error curves presented in (Shiklomanov et al., 450 2006), we defined an RMS-error function that depends on the discharge. The function follows an exponential decay from 20 % at the lowest discharge to 2 % at the mean discharge, and then an exponential increase from 2 % at the mean discharge to 5 % at the highest discharge. The slope of the function is -2 at the low flow end and 0 at the mean flow point (see Fig. E1).

An RMS error signal consistent with this in situ uncertainty function is added to the OBD of each basin to generate a OBD_{error} time series. In situ measurement uncertainties are then assessed for each basin by computing the Nash-Sutcliffe Efficiency 455 (NSE) between OBD and OBD_{error} . The NSE values range from a minimum of 0.982 to a median of 0.998, indicating that uncertainties in OBD arising from in situ measurement errors are negligible

3.3.2 Altimetry-based discharge

The previous section shows that in situ-based discharges are highly reliable. We calculated NSE and KGE between these 460 discharges and the CCI river discharges for basins where gauge data are available (Amazon, Congo, Danube, Garonne, Irrawaddy, Lena, Mackenzie, Maroni, Mississippi, Niger, and Ob). For comparison, we also calculated NSE and KGE between in situ-based discharges and the MBDs over the same basins. For altimetry-based discharge the median NSE and KGE are 0.87 and 0.86, respectively, while the lowest NSE and KGE are 0.74 and 0.65, which is significantly better than the results obtained with any MBD that, even with routing, always yields a negative NSE for at least one basin (see Figure B1).

465 We computed the NSE between in situ discharges and the 16 discharges from spatial altimetry used in this study (CCI river discharge data with the addition of altimetry-based discharge on the Indus, Upper Indus, Mekong, Volga, and Yellow River computed in this study). The median NSE is 0.82 and the lowest NSE is 0.55 (Upper Indus). The median KGE is 0.79 and the lowest KGE is 0.3 (Indus). 6 basins obtain a NSE value above 0.9: Mekong (0.99), Amazon (0.97), Ob (0.95), Congo (0.93), Irrawaddy (0.91), Niger (0.9) and 6 basins obtain a KGE above 0.85: Amazon (0.97), Congo (0.92), Irrawaddy (0.92), Ob 470 (0.92), Mississippi (0.87), Danube (0.86).



These results show that the discharges from the ESA CCI River Discharge Precursor project and, more broadly, the altimetry-derived discharges, are of higher quality than the MDB discharges.

3.3.3 Observations-Based Discharge (OBD)

475 Section 3.2.1 shows that in situ discharges are highly reliable, while Sect. 3.2.2 shows that altimetric discharges, despite having lower performance, are significantly more reliable than MBD discharges. OBD is based on a combination of in situ and altimetric data (Sect. 3.1.2). Throughout this study, OBD uses in situ data 86% of the time (see Fig. S23). Thus, OBD can be considered a reference discharge and is far more reliable than MBD.

3.3.4 Model-Based Discharge and Observations-Based Discharge comparisons

480 Section 3.2.3 shows that OBD are very reliable. The performance of MBD will be evaluated by comparing it to OBD. Comparisons between OBD and MBD are carried out with correlation, NSE and KGE where the observation part corresponds to the OBD and the model part to the MBD (Sect. 3.1). The components of KGE related to correlation (Eq. (4)), mean (Eq. (5)), and standard deviation (Eq. (6)) are also examined to more precisely identify the nature of the differences between MBD and OBD.

485

3.4 Water budget evaluation

TWS monthly variation data are computed using the same basin mask as the P, ET, TRD, and RRD datasets. Nevertheless, 490 TWS variation is provided as monthly mean with a value centered on the middle of the month. The difference between two successive monthly TWS data points therefore indicates thus a variation in TWS between mid-months, while P, ET, and Q monthly data provide calculated volumes from the beginning to the end of the month. So deriving TWS data in the water budget equation (Eq. (10)) requires a specific processing.

$$\frac{dTWS(t)}{dt} = P(t) - ET(t) - Q(t) \quad (10)$$

495 The GRACE TWS data are processed by using the interpolation method recommended in the literature and presented in Eq. (11) and Eq. (12) (Landerer et al., 2010; Pellet et al., 2021). Justifications and limitations of this method can be found in Appendix A. The filtering applied to the right side of Eq. (11) is necessary because incorporating TWS data from the previous and following months requires adding P-ET-Q data from those same months to the equation, with appropriate weighting. For the rest of this paper, DTWS(n) refers to the left-hand term of Eq. (11), i.e., the derivative of TWS for month n.



500

$$DTWS(n) = \frac{TWS(n+1) - TWS(n-1)}{2} = \tilde{P}(n) - \tilde{ET}(n) - \tilde{Q}(n) \quad (11)$$

$$\tilde{F}(n) = \frac{1}{4}F(n-1) + \frac{1}{2}F(n) + \frac{1}{4}F(n+1) \quad (12)$$

505 The WBE is evaluated using the NSE (Sect. 3.1) where the observation part corresponds to DTWS(n) and the model part to the three-points filtered (P-ET-Q)(n) (see Eq. (13)). A positive NSE means that the budget reconstruction (P-E-Q)(n) is a better approximation of DTWS(n) than the DTWS long-term mean. KGE should not be used to calculate WBE because the average of DTWS can be close to 0 and thus create mathematical instability (see KGE definition in Eq. (2)).

$$NSE \left(\text{Obs} = DTWS(n), \text{Mod} = \tilde{P}(n) - \tilde{ET}(n) - \tilde{Q}(n) \right) \quad (13)$$

510 Considering the SAGSA TWS variation dataset, NSEs are specifically calculated for each combination of 20 DTWS, 12 P, 12 ET and for each one of the 17 discharges (OBD + 8 RRD + 8 TRD). This yields 2 880 NSE values for each discharge. NSE_{OBD} is the highest NSE among the 2880 results considering the OBD discharge. We compute in the same way 8 NSE_{RRD} and 8 NSE_{TRD} that are the highest NSE considering, respectively, the 8 RRD and the 8 TRD.

515 To our knowledge, no study has precisely evaluated the relevance of the three-point filtering (Eq. 12). Among recent studies, some use it (Jiao et al., 2025; Lehmann et al., 2022; Lorenz et al., 2014; Zhu et al., 2021), and others do not, particularly those seeking to enhance the closure of the water budget through methods such as PR, CKF, MCL or MSD techniques (Abolafia-Rosenzweig et al., 2021; Luo et al., 2026, 2023; Pan et al., 2015; Zhang et al., 2018).

520 To validate the method, we calculated NSE_{OBD} both with and without three-point filtering. The filtering provides an improvement in the WBE for each of the 82 basins (Figure A2). These computations obtain an average and median NSE_{OBD} values of 0.56 and 0.62 without filtering and an average and median NSE_{OBD} values of 0.81 and 0.84 with filtering. For some basins like the Amazon, Ob, and Niger, the improvement is modest (<0.05), whereas for other basins such as the Amur, Brahmaputra, or Magdalena River, the gain is substantial (>0.5). We strongly recommend that three-point filtering be used when computing the WBE, and it will be applied throughout the remainder of the study.

525 The P and ET datasets that allows obtaining the highest NSE are identified for each of the 17 discharges. P_{OBD} and ET_{OBD} are the precipitation and evapotranspiration datasets derived from the combination that yields the NSE_{OBD} . Similarly, P_{TRD} and ET_{TRD} on the one hand and P_{RRD} and ET_{RRD} on the other hand are the P and ET datasets derived from the combination that yields NSE_{RRD} and NSE_{TRD} .

530 ΔP_{RRD} is the percentage difference between the annual volumes of P_{RRD} , and the annual volume of P_{OBD} . ΔP_{TRD} is the percentage difference between the annual volumes of P_{TRD} , and the annual volume of P_{OBD} . ΔET_{RRD} and ΔET_{TRD} are, as well, the percentage difference between the annual volumes of ET_{RRD} and ET_{TRD} , and the annual volume of ET_{OBD} (Eq. (12), Eq. (13), Eq. (14), and Eq. (15)).



$$\Delta P_{RRD} = \frac{P_{RRD} - P_{OBD}}{P_{OBD}} \quad (14)$$

$$\Delta P_{TRD} = \frac{P_{TRD} - P_{OBD}}{P_{OBD}} \quad (15)$$

$$\Delta ET_{RRD} = \frac{ET_{RRD} - ET_{OBD}}{ET_{OBD}} \quad (16)$$

$$535 \quad \Delta ET_{TRD} = \frac{ET_{TRD} - ET_{OBD}}{ET_{OBD}} \quad (17)$$

144 combinations of 1 DTWS, 12 P, 12 ET and OBD have been also calculated for each TWS mascon product GSFC, CSR, and JPL. NSE_{GSFC} , NSE_{CSR} , and NSE_{JPL} are the highest NSE values considering these combinations.

540 A summary of the methodology is shown in Figure 3.

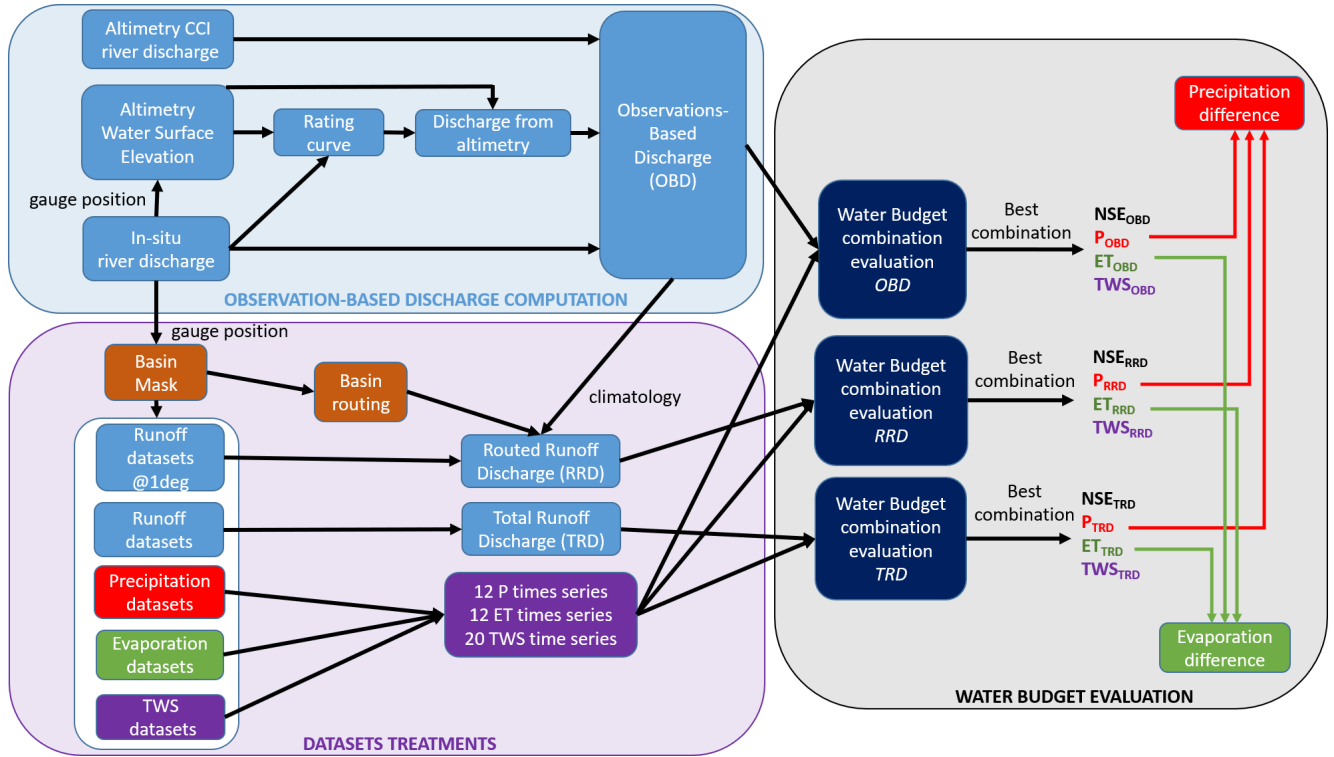


Figure 3: flowchart of the study methodology



3.5 Dispersion of each component of the water budget

The dispersions are estimated using the CRV (Sect. 3.1) applied to the monthly climatologies of each dataset for DTWS, P, ET, TRD, and RRD, with the monthly-climatology amplitude of DTWS serving as the normalization coefficient (see Eq. 18). This approach allows the dispersion of each component to be evaluated relative to the average annual water-stock variation of the basin. Unlike the CV, CRV values are directly comparable across variables: a CRV_P of 20% and CRV_{ET} of 20% for a $DTWS_{amp}$ of 100 km^3 correspond for example to the same monthly average standard deviation of 20 km^3 .

$$CRV_X = CRV_{DTWS}(C_m(X)) \quad (18)$$

With:

- $X = \text{DTWS, P, ET, TRD, RRD}$
- $C_m(X)$ the time-dependent vector that contains the monthly climatologies of all datasets associated with X .

555

The ratio of two CRVs can be expressed directly as the product of the ratio of the corresponding CVs and the ratio of their mean values (Sect. 3.1). Accordingly, for $X=P$ or ET , the CRV_X/CRV_{RRD} ratios can be compared to CV_X/CV_{RRD} multiplied by the ratio the mean volume of X and RRD (Eq. 9).

$$\frac{CRV_X}{CRV_{RRD}} = \frac{CV_m(X)}{CV_m(RRD)} \times \frac{\mu_X}{\mu_{RRD}} \quad (19)$$

560 With μ_X , the mean annual volume of the X .

4 Results

Before tackling the core of the study and the calculation of WBE, we will first precisely examine the performances and specificities of the different MBDs across the 82 basins for each climate type (Sect. 4.1). Exploiting the ensemble approach of the SAGSA product (whose performance will be evaluated preliminarily), we will then examine the results of the dispersion calculations for each component of the water budget (i.e. DTWS, P, ET, and Q) (Sect. 4.2). The objective of this second part is to identify whether the absolute dispersion of the MBDs is significant compared to that of the other components and can therefore have an impact on the conclusions of the WBE. Finally, we will present the results obtained from the WBE calculation for the 82 basins: the closure levels as well as any compensations and their impact on the evaluation of P and ET volumes (Sect. 4.3). These results will be shown at the basin scale, by climate type, and with a more global approach.



4.1 Model-based discharge evaluation

575 The use of runoff routing significantly increases the correlation between the MBD (i.e. both TRD and RRD) and the OBD, regardless of which MBD is considered (Fig. 4c). Routing also improves discharge estimates for each of the 8 MBDs (Fig. 4a and Fig. 4b). However, performance is highly variable from one MBD to another. The median NSE obtained between OBD and MDB is negative in most cases, meaning that for half of the basins MDB is less accurate than a discharge calculated from a simple average. Moreover, each MDB, whether routed or not, estimates discharges poorly for several basins, with NSE and
580 KGE values far below zero (Fig. S13 and Fig. S14). GRUN, the only non-physical model based on AI, is the model that gives the best results in term of TRD or RRD.

RRD results vary significantly depending on the climate types: GRUN and ERA5-Land are often among the best three products for each climate (Fig. S15 to Fig. S20). Tropical and subarctic basins are those for which the performances are the highest (NSE up to 0.74 and KGE up to 0.77 for GRUN RRD on tropical basins, see Fig. S22) and also those where the use of routing
585 improves results the most, particularly for the subarctic basins (Fig. S19). Very poor results are obtained on arid basins with or without routing, regardless of the MBD (Fig. S16).

RRDs are statistically well correlated with OBD (median value > 0.7). Six of the eight RRDs, although performing better than the TRD, estimate OBD less accurately than a simple average of OBD (median $NSE < 0$) and none of them attains a median NSE above 0.5, i.e., the threshold commonly used to deem a simulation behavioural. Only three RRDs achieve a median KGE
590 that meets the behavioural threshold of 0.3: MERRA-2 (0.3), ERA5-Land (0.37) and GRUN (0.44). The decomposition of the KGE helps to better understand the origin of these relatively low discharge performances (Fig. H1 and H2). Median KGE_{corr} is smaller than the sum of the median KGE_{mean} and of the median KGE_{std} when applying the routing, indicating that the largest part of the error is not due to a temporal shift issue. Applying routing has a modest impact on the discharge performance for most basins as the error seem more dominated by noise and bias.

595

The monthly RRD climatologies show that the differences between the RRD are mainly related to under or overestimation at the peak flow (Fig. G1, Fig. G2, Fig. G3). Low flow is often closer to in situ data. Moreover, the RRD are also very strongly correlated with each other across most basins (Fig. S24).

600 To better understand differences between OBD and RRD, we calculated for each basin the percentage between the average annual volume of MBD and the average annual volume of OBD. Note that average annual volume of RRD and TRD are mathematically very close. There is a global overestimation of discharges for the GLDAS VIC (+157%) and GLDAS Noah (+38%) models, a global underestimation for the MERRA2 model (-32%), and more balanced values for the other models: CLSM (-4.3%), ISBA (-9.3%), ERA5-Land (+8.8%), JRA-55 (+11%), and GRUN (-2%) (Fig. 5). There is a high
605 overestimation of discharges on most arid basins and often an underestimation on subarctic and polar basins (Fig. S22).



610 OBD discharge can be considered a much better estimate of real discharge than MBD (Sect. 3.2.4). The use of a simple routing (i.e. RRD) increases the correlation between MDB and OBD and improves in most basins the quality of the discharge computed from the modelled runoff. Thus, RRD will be preferred to TRD in the rest of this study. Nevertheless, even with routing, RRD performances remains low for many basins due to important differences with in situ data during high flow periods, especially for arid ones where discharge are often overestimated. These findings prompt us to examine in Sect. 4.3 how the choice of discharge source (OBD versus MDB) influences the conclusion of the WBE.



OBSERVATIONS-BASED DISCHARGE VS MODEL-BASED DISCHARGE

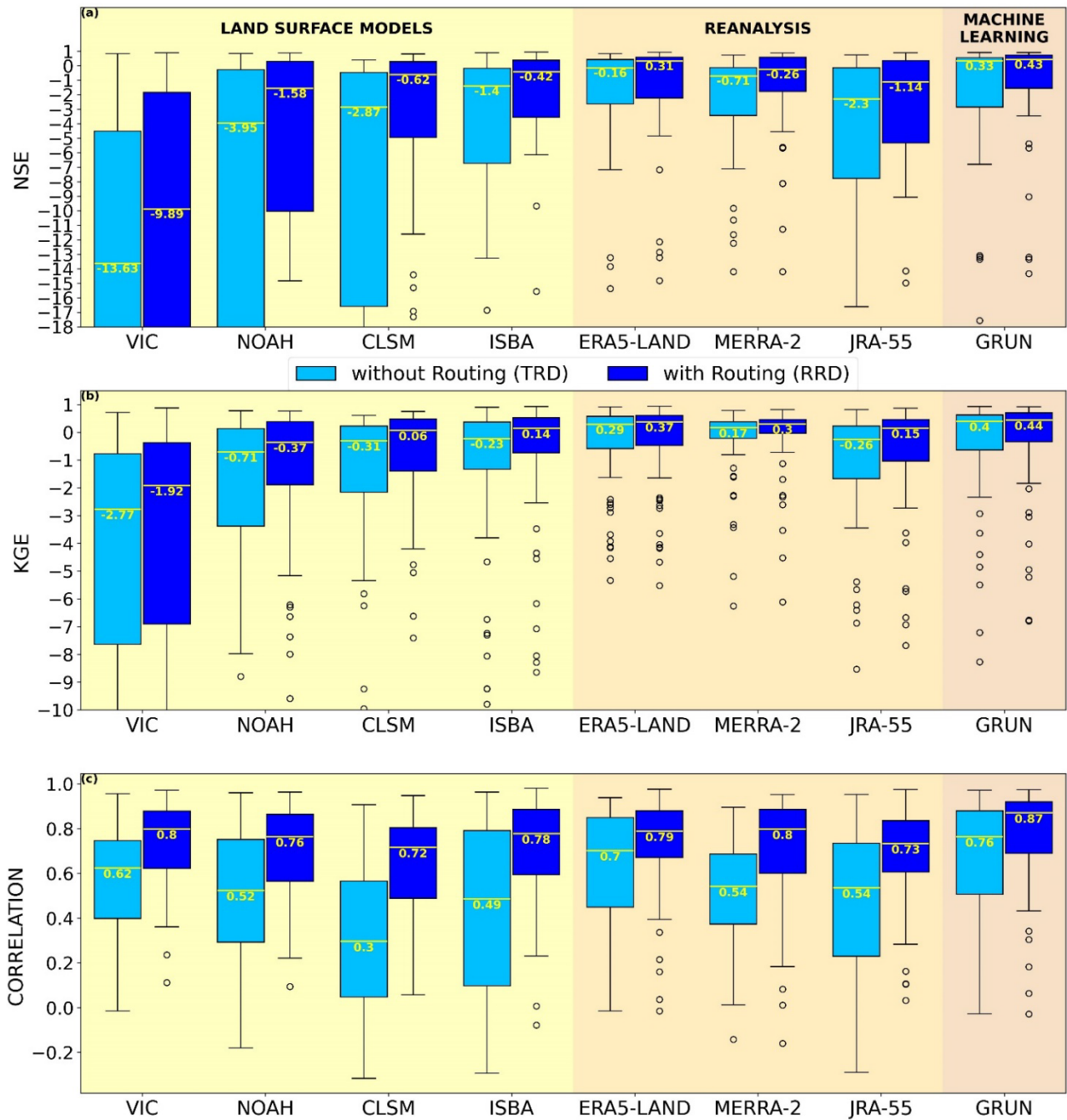


Figure 4: Evaluation of model-based discharge and effect of the routing on the discharge estimation. Boxplot of NSE (a), KGE (b) and correlations (c) for all 82 basins between the Observations-Based Discharge (OBD) and the two model-based discharge: TRD (light blue) and RRD (blue) discharges. Median values are indicated in yellow. There is a zoom on subfigure (a): minimum y axis value is -18 (same figure without zoom is available in supplementary materials). There is a zoom on the KGE: minimum y axis value is -10

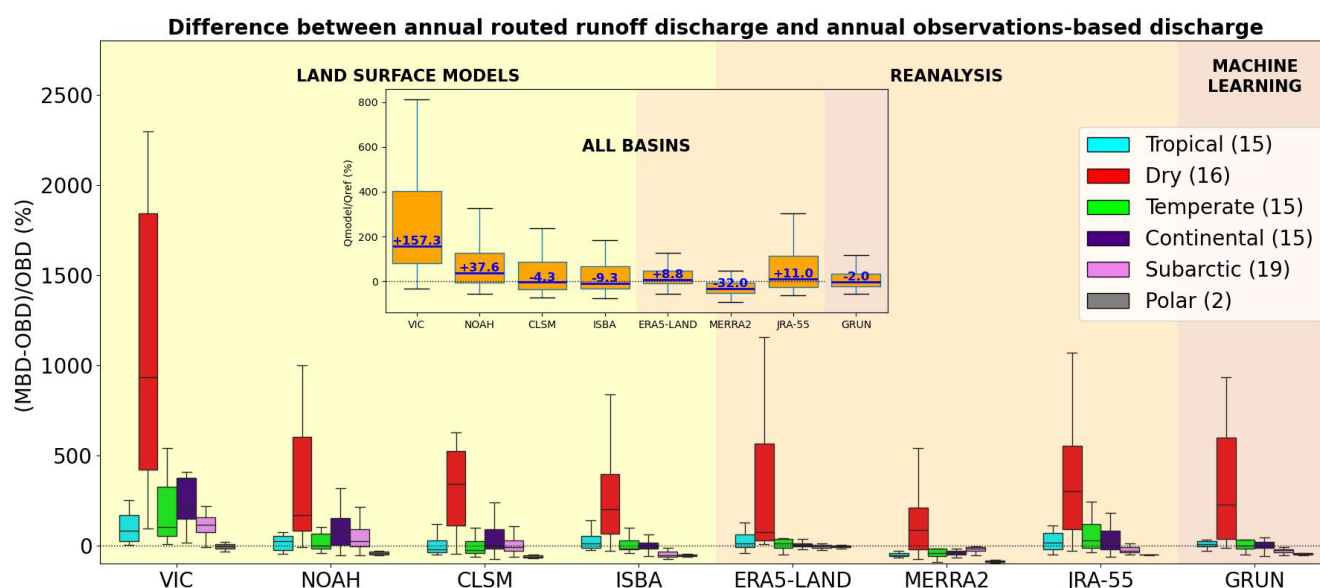


Figure 5: Percentage of the average annual volume of the RRD compared to that of the OBD: for each climate in the main figure and for all 82 basins in the secondary figure. Median values are indicated in blue on the secondary figure. The black number in the legend corresponds to the number of basins considered for each climate (same figure with zoom between -100% and +200% is available in supplementary material).

4.2 Datasets dispersion

Lehmann et al. (2022) indicates that compensation phenomena can occur during the calculation of the WBE. We consider that two components of the water budget with comparable dispersion levels can create compensation phenomena between them, especially if their datasets are correlated. We can also assume the existence of three- or four-term compensations, where the stronger variability of one term is offset by the combined variability of several other terms. A component with low dispersion, on the other hand, will play a minor role in this compensation issue. Previous section shows significant differences between RRD despite a high level of correlation with OBD. Thus, before computing WBE, we evaluate the dispersion of RRD and compare it with DTWS, P, and ET dispersions.

4.2.1 SAGSA ensemble product evaluation

The SAGSA product provides a set of TWS solutions (Sect. 2.2.1), allowing us to compute the DTWS dispersion and compare it directly with the dispersions derived from P, ET and RRD datasets. However, there is no published study evaluating the SAGSA TWS ensemble product. Thus, before proceeding to the dispersion comparisons between each part of the water budget equation, SAGSA TWS product performance must be evaluated. This evaluation will be performed by comparing the SAGSA



TWS product with GSFC, CSR, and JPL MASCON TWS products. Although, these latter have already been widely used in the literature (Aires and Pellet, 2024b; Lehmann et al., 2022; Luo et al., 2026) and are considered reference products.

Following the method proposed in Sect 3.4, we compared NSE_{OBD} (computed with SAGSA product) against NSE_{GSFC} , NSE_{CSR} , and NSE_{JPL} for each basin. WBE closure levels are quite similar, with NSE ranging from 0.80 (CSR and JPL) to 0.83 (SAGSA and GSFC) showing the maturity of gravimetry products (Fig. 6). Regarding the results by climate, the SAGSA product gives slightly better WBE performances except for temperate climates where median NSE_{GSFC} is slightly higher than median NSE_{OBD} (but with a lower minimum value). On the two polar sub-basins (Brahmaputra and Upper Indus), the improvement is more significant with a median NSE_{OBD} of 0.82 compared to 0.76, 0.63, and 0.46 for the three Mascon products. SAGSA obtains the highest minimum (Rio Colorado, 0.49), followed by GSFC (Rio Magdalena 0.38) and CSR (Hayes River 0.15), whereas JPL yields the lowest minimum (Hayes River -0.49) (see detailed results in Appendix F). Apart from polar climates where it is difficult to conclude due to the low number of basins, the four products show the same pattern by climate type. Tropical basins give a better level of closure with a median NSE around 0.92. Temperate, continental, and subarctic basins give a comparable level of closure with a median NSE around 0.8. Polar and arid basins are those that give the lowest level of closure with a median NSE around 0.7.

The good results of the SAGSA product can be attributed to the fact that it is an ensemble product that includes the solutions of the three centres, adding those of the TUGRAZ and CNES centres as well.

In conclusion, the SAGSA ensemble product yields WBE of a quality that is comparable to, or even slightly better than, the three reference products. Consequently, using this product to compute a dispersion and to identify compensation phenomena in the water budget equation is entirely legitimate.

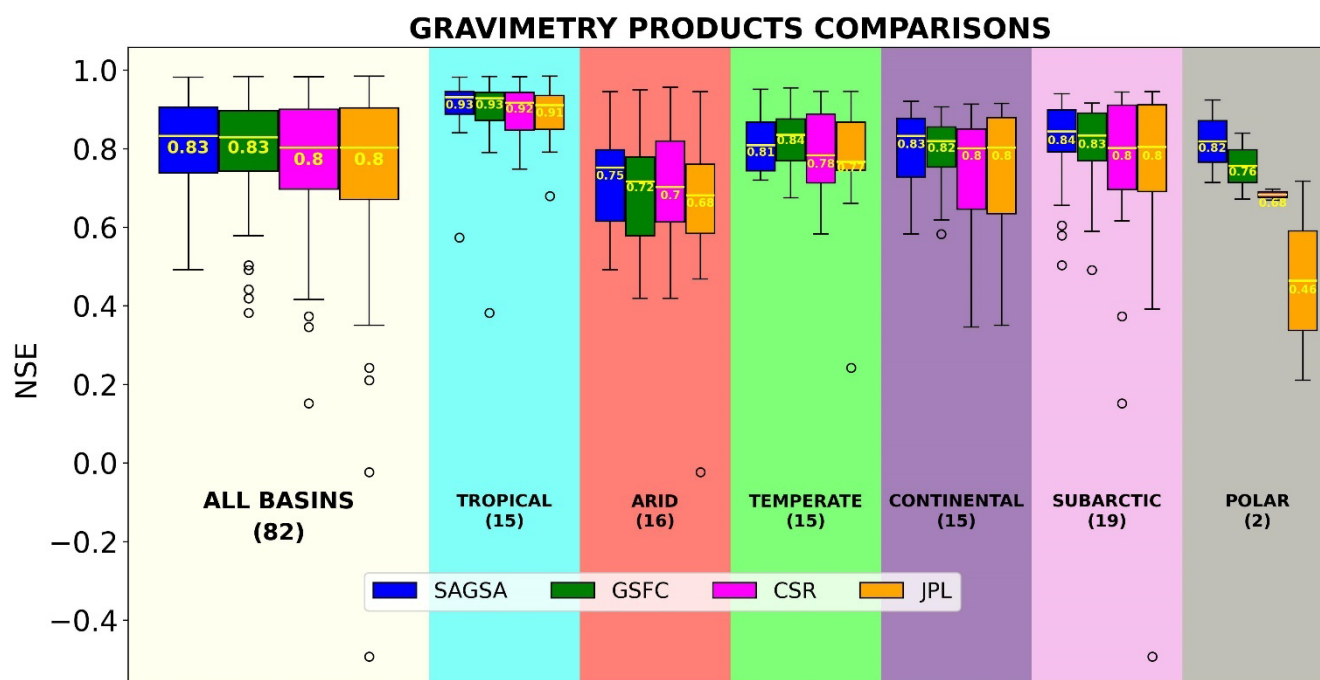


Figure 6: Boxplot of the highest NSEs obtained for each of the 4 GRACE products considered: for each of the 82 basins and also for each of the 6 climate types. The black number in parentheses corresponds to the number of basins considered in each case. The median values are indicated in yellow.

4.2.2 Dispersion of each component of the water budget (DTWS, P, ET, and RRD)

The dispersion of each component of the water budget equation reveals whether this component plays an important role in the overall budget. A component with low dispersion will have little effect on the achieved closure and, consequently, on the results, whereas a component with high dispersion can cause the water-budget closure to vary widely. Moreover, compensation phenomena are more likely to emerge when two, three, or even all four components exhibit large dispersions.

The 12 precipitation datasets exhibit the highest level of dispersion, with a median CRV_P of 43.5 % (Fig. 7). The 8 RRDs and the 12 evaporation datasets show comparatively similar dispersion, with a median CRV_{RRD} of 36 % and a median CRV_{ET} of 33.5 %. In contrast, the twenty DTWS derived from the SAGSA product display a significantly lower dispersion, with a median CRV_{DTWS} of 7.1 %. Recall here that using the CRV allows the calculation of comparable dispersion across the different components of the water budget (Sect 3.1 and Sect 3.5).

Thus, although discharge represents the smallest flux of the water cycle, its dispersion between datasets is slightly stronger than that of ET. It is even higher than that of P in continental basins (59 % vs 48 %) and sub-arctic basins (39 % vs 37 %). This indicates a particularly strong impact of the runoff datasets choice on these basins, as well as potential compensation phenomena.

The DTWS dispersion remains low compared to P, ET and RRD regardless of the climate. Thus, the choice of a particular TWS solution should not fundamentally change the conclusions of most WBE.



RRD dispersion is highly dependent on climate type. It is significantly higher in continental basins and lower on tropical and arid basins. This lower dispersion over arid basins may seem counterintuitive, since models tend to make larger errors in those basins (Fig. 5). This could be attributed to the fact that the proportion of discharge in the water budget is lower in arid basins (Addor et al., 2017; Oudin et al., 2008). Thus, although the RRD dispersion can be large relative to the actual discharge, it can still be small compared to the dispersions of P and ET. On the other hand, low dispersion between models does not necessarily imply low uncertainty. Models generally do not account well for anthropogenic alterations, so they may produce discharge estimates consistent with each other but nevertheless far from reality.

Median ET dispersion is slightly greater than P on arid basins, indicating also possible compensation phenomenon between P and ET. Median P and ET dispersion are lower on subarctic basins than on other basins. However, as seen previously for RRD on arid basin, this does not necessarily imply better reliability of P and ET datasets on these basins.

To the best of our knowledge, this way of calculating the dispersion of the four water budget equation components has never been proposed. For this reason, we compare our results, expressed as CRV, with the CVs reported by Jiao et al. (2025), using the relationship between CV and CRV described in Sect. 3.5 and Eq. 19.

Jiao et al. (2025) showed across 10 basins that runoff mean CV is generally twice as high as precipitation and evapotranspiration mean CVs. At the global scale, total annual discharge to ocean is between 3 and 3.2 times smaller than total annual precipitation over land and between 1.7 and 1.8 times smaller than total annual land evaporation (Dorigo et al., 2021). If the CV results of Jiao et al. (2025) could be generalised the ratio CRV_{RRD}/CRV_P would be between 0.62 and 0.66 while the ratio CRV_{RRD}/CRV_P would be between 1.11 and 1.17 (see Eq. 19). These results are consistent with our findings, i.e., runoff exhibits a slightly larger dispersion than evapotranspiration, and precipitation shows a greater dispersion than both runoff and evapotranspiration.

Our study finds a ratio of CRV_{RRD}/CRV_{ET} of 1.07 consistent with these former results. However our study finds a larger CRV_{RRD}/CRV_P (0.82 instead of 0.66), which could be explained by the differences in the datasets used and, above all, the type of basins considered by Jiao et al. (2025). For example, they did not consider subarctic basins, where, according to our results, the median CRV_P is lower (37%).

The results presented in this section show that, although the RRD signal has a smaller magnitude, its variability is comparable to that of P and ET, and for some basins it is even higher. RRD datasets are highly correlated with one another and their dispersion is primarily associated with under or overestimation of the peak flow (Sect. 4.1). As a consequence, compensations can readily emerge between P and RRD, as well as between ET and RRD, when computing water-budget estimates. Section 4.3 will investigate whether such compensations actually occur and, if so, how they influence the conclusions drawn from the WBE analysis.

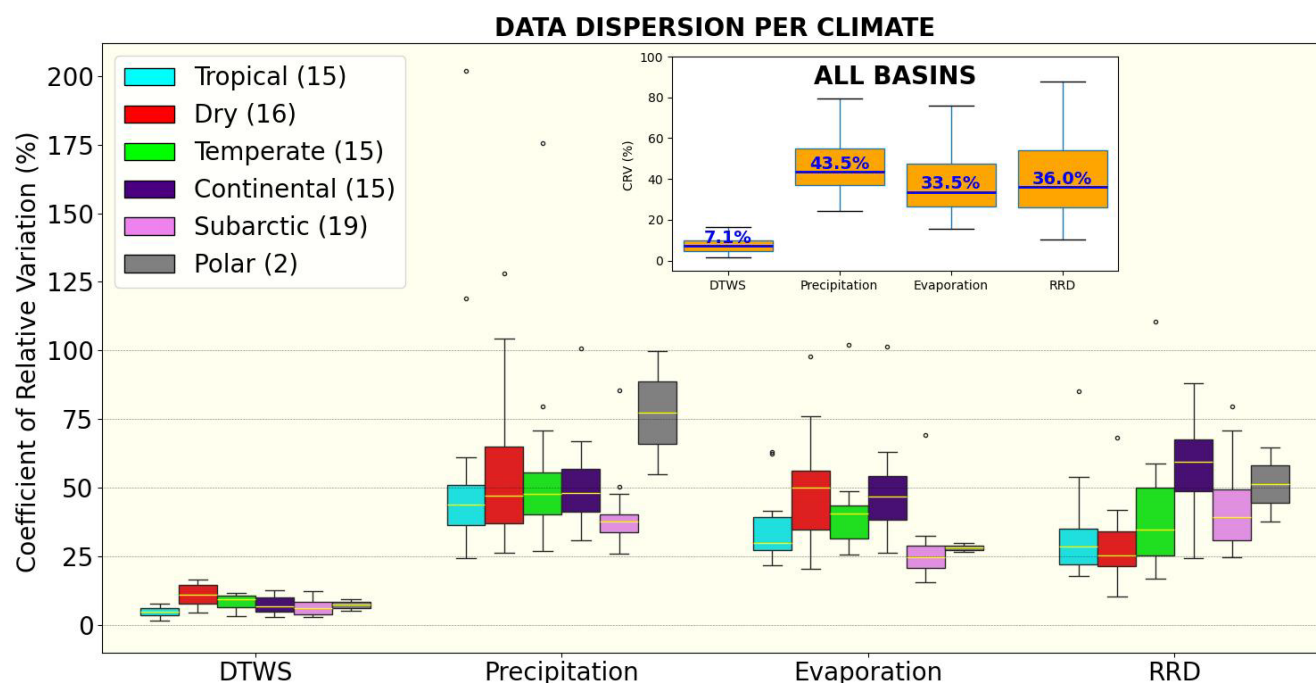


Figure 7: Coefficient of Relative Variation in % for each component of the water budget: for each climate (main figure) and across all basins (secondary figure). The coefficient of variation for discharges is calculated with RRD, i.e., the runoff summed using the routing method proposed in Sect. 3.2.3. Median values are indicated in yellow on the main figure. Median values are indicated in blue on the secondary figure. The black number in parentheses corresponds to the number of basins considered for each climate.

4.3 WBE results

4.3.1 WBE statistical results

This section presents the WBE obtained from OBD and from MDB, both with and without routing. Applying routing improves the water budget closure level for each MDB (Fig. 8). The WBE calculated with OBD are of high quality, exhibiting a median NSE_{OBD} of 0.83 and, most importantly, values above 0.5 for all basins, i.e., above the NSE behavioural threshold (Sect. 3.1).

The WBE obtained with MDB are of comparable quality, with median NSE_{RRD} exceeding 0.5. Moreover, the WBEs derived from ISBA, ERA5-LAND, MERRA-2, JRA-55 and GRUN give results that are very close to those obtained with OBD, showing a median NSE_{RRD} around 0.80, consistently positive NSE_{RRD} values, and only a few outliers below the 0.5 threshold. GLDAS VIC yields the poorest performance, although more than half of its WBEs still achieve $NSE_{RRD} > 0.5$.

The comparison between water budget closure levels (Fig. 8) and the assessment of TRD and RRD (Fig. 6) (presented on Fig. S21 with the same scale), raises a question: how can similar closure levels be obtained despite substantial differences in the discharge estimates? The most likely explanation is the existence of compensations (Lehmann et al., 2022). The datasets P,



ET, and RRD exhibit comparable dispersions (Sect. 4.2), while the correlation level of the RRD is very high (Sect. 4.1). These two observations support the hypothesis that compensation occurs during the WBE calculation. Comparing the WBE obtained with OBD and each MDB across a few basins could help confirm the existence and importance of these potential compensations.

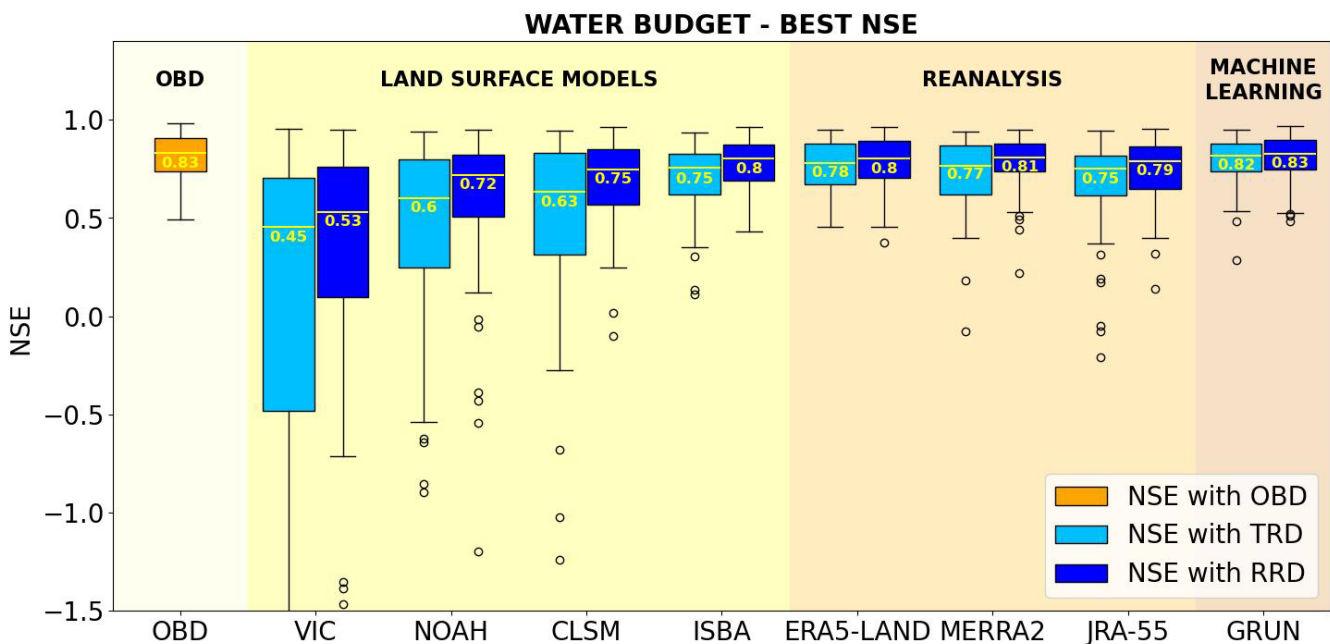


Figure 8: Boxplot of the highest NSEs obtained for each of the 82 basins and for each type of discharge: Observations-Based Discharge in orange, light blue for TRD discharges, and blue for RRD discharges. Median values are indicated in yellow. The figure is zoomed with a minimum y axis value of -1.5.

4.3.2 Examples of compensation in the WBE

We will focus on WBE for the Indus and Brahmaputra basins in order to investigate the compensation phenomenon (Amazon, Ob, Columbia, Irrawaddy, and Rio Bravo WBE available in Appendix C).

The Indus basin is classified as arid and is highly anthropogenically altered. Intensive groundwater pumping has caused a marked decline in total water storage (Rodell et al., 2018, 2009; Zhu et al., 2021). The Indo-Gangetic plain represents the world's largest irrigated region (Khan and Adams, 2019). We compare the average annual volumes of OBD, P_{OBD} , and ET_{OBD} with average annual volumes of, respectively TRD, P_{TRD} , and ET_{TRD} , and RRD, P_{RRD} , and ET_{RRD} (Fig. 9). Note that the average annual volume of TRD and RRD are similar. Except for MERRA2, MBD annual volume is overestimated. VIC RRD overestimates discharge by 752 % relative to OBD (i.e., an excess of $255 \text{ km}^3 \text{ yr}^{-1}$). VIC P_{RRD} volume is +13 % ($+63 \text{ km}^3 \text{ yr}^{-1}$) above P_{OBD} , whereas VIC ET_{RRD} volume is -39 % ($-156 \text{ km}^3 \text{ yr}^{-1}$) under ET_{OBD} . The overestimation of precipitation together



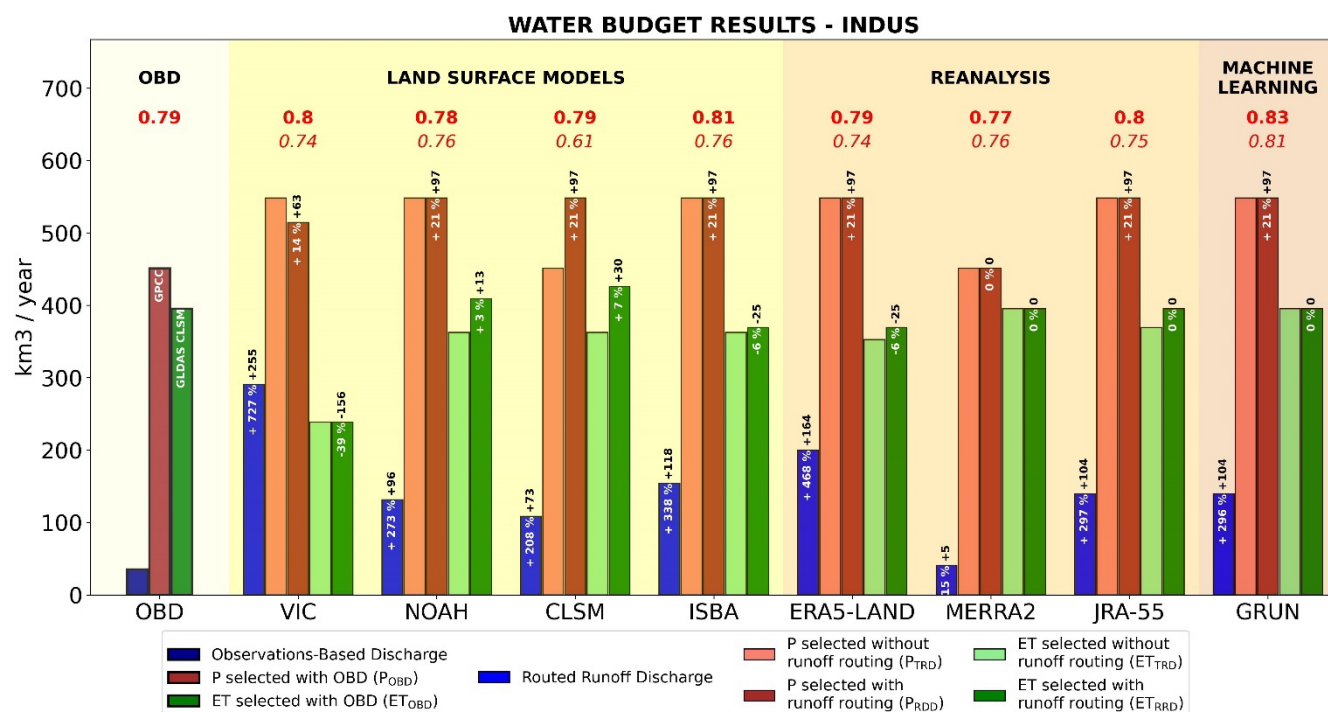
with the underestimation of evapotranspiration compensates the discharge overestimation, resulting in a high level of water budget closure ($NSE_{RRD} = 0.80$), which is comparable to the closure obtained with OBD ($NSE_{OBD} = 0.79$).

755 Similar behaviour is observed with other models: RRD overestimation is compensated by selecting parameter combinations that overestimate precipitation and/or underestimate evapotranspiration. Moreover, applying routing does not fundamentally alter the results. When routing is applied to GLDAS-CLSM, performance actually degrades: the precipitation excess relative to RRD rises to +21 % (versus 0 % for TRD).

760 The opposite pattern emerges in the Brahmaputra basin (Figure C1). Here, all MDBs, except ERA5-Land, underestimate discharge, for example by 71% for the CLSM RRD, yet with a higher value of water budget closure ($NSE_{RRD} = 0.74$ whereas $NSE_{OBD} = 0.71$). In this case the compensation stems from a substantial underestimation of both precipitation (-52 %) and evapotranspiration (-27 %)

Comparable cases with high WBE closure level despite marked discrepancies in the estimates of precipitation, evapotranspiration, and discharge are found for other basins (see Appendix C).

765 These findings underscore that a high level of water budget closure alone does not guarantee accurate estimation of the individual flux components. WBE computation with low-quality discharge data can lead to misleading inferences about the annual volume of precipitation and evapotranspiration.



770 **Figure 9: Average annual volume of discharge (blue), precipitation (red), and evapotranspiration (green) that, when combined together, provide the highest NSE for OBD and each MBD. Light red and light green correspond to the precipitation and evapotranspiration volume for TRD, while red and green to the precipitation and evapotranspiration volume for RRD. The corresponding NSE is indicated above results in bold red for OBD and RRD discharges and in italic red for TRD discharges.**

775 4.3.3 Difference in P and ET estimation

Calculating a WBE for a basin allows us to identify which datasets perform better for that basin. It is a valuable tool for refining our understanding of the water cycle at the basin scale, especially for estimating the annual volumes of P and ET. The compensation phenomena described in the previous section demonstrate that substantial differences can exist in the estimated annual volumes of P and ET depending on the discharge dataset that is selected. We calculate in this section the differences in annual volume estimates of P and ET as a function of the MDBs and climate types

We computed ΔP_{RRD} , ΔP_{TRD} , ΔET_{RRD} , and ΔET_{TRD} for each basin. Before discussing the results, let us recall that TRD is the discharge commonly used by WBE studies based on runoff instead of a discharge. This study calculated an RRD to verify if the differences could be attributed to the lack of routing. As mentioned in Sect 4.1, the use of propagation primarily benefits certain large tropical and subarctic basins. Moreover, the difference between RRD and OBD is more a matter of high flows estimation than temporal offset.

The use of routing does not significantly reduce the errors in evaluating precipitation and evapotranspiration (Fig. D1). The use of GLDAS VIC and Noah MBD tends to overestimate P and underestimate ET, while the use of MERRA2 runoff tends



slightly to underestimate P. No significant trend is observed for the results obtained with the other MDBs, and each MDB leads to either an overestimation or an underestimation of ET and P, depending on the basin. All of the MDBs also produce substantial errors for several basins. ΔP_{RRD} outliers are above +30% and under -38% for each RRD while ΔET_{RRD} outliers are above +37% and under -37% for each RRD. GRUN RRD gets the best results but its ΔP_{RRD} and ΔET_{RRD} outliers are still around +/-40%.

ΔET_{TRD} and ΔET_{RRD} are generally greater than ΔP_{TRD} and ΔP_{RRD} . Median absolute values of ΔP_{RRD} vary between 13.4% and 5.4% depending on the RRD with a mean value of 8.5% (Table D1) while median absolute values of ΔET_{RRD} vary between 33.1% and 0% with a mean value of 11.9% (Table D2). The compensation phenomenon depends on the degree of dispersion among the RRDs, and this dispersion is fixed for a given basin. That absolute dispersion among the RRDs can offset errors either in P or ET. Since the annual ET volume is on average about two-thirds of the annual P volume (Dorigo et al., 2021), this explains why the compensation effect is proportionally stronger for ET than for P. These results are consistent with Luo et al. (2026) who found mean error values of 6% for P and of 7% for ET using reinforcement methods in WBE computation. The slightly higher mean values observed in our study can be explained by the fact that we considered twice as many P, ET and MDB datasets, which increases dispersion and compensation effects.

For a given RRD, both the sign and magnitude of ΔP_{RRD} and ΔET_{RRD} depend on climate type (Fig. 10). Median ΔET_{RRD} are frequently negative for continental, subarctic and polar basins. ΔP_{RRD} and ΔET_{RRD} are particularly large in the two polar basins. Although discharge errors are larger in arid basins (Sect. 4.1; Fig. 5), the resulting ΔP_{RRD} and ΔET_{RRD} are not markedly greater, likely because the CRV_{RRD} is smaller in those regions (Sect. 4.2). Classifying basins into two climate types (humid or arid), Luo et al. (2026) found lower error for arid basins. However, using six different climate types and twice as many MDBs, our analysis does not reveal a consistently lower ΔP_{RRD} and ΔET_{RRD} in arid basins regardless of the MDB, especially when compared with tropical and temperate basins (Table 2). For example, when the MERRA-2 and GLDAS NOAH RRD are employed, ΔP_{RRD} and ΔET_{RRD} are indeed reduced in arid basins. Conversely, with the GLDAS CLSM and GRUN RRD, ΔP_{RRD} becomes markedly larger in arid basins than in tropical or temperate ones. Our results show moderate ΔP_{RRD} and ΔET_{RRD} for the tropical, arid, and temperate basins (6%), while the continental, sub-arctic, and polar basins exhibit larger errors (15%) (Table 2). This can be explained by the fact that the dispersion of the RRDs is greater in continental, sub-arctic, and polar basins (Sect 4.2.2; Fig. 8). In the continental and sub-arctic basins, the spread of the RRDs even exceeds that of P and ET. Reanalysis generally yield lower ΔP_{RRD} and ΔET_{RRD} than LSMs. GRUN RRD obtains the best results on most regions except on the two polar basins. ERA5-Land is the only RRD that delivers moderate ΔP_{RRD} and ΔET_{RRD} on the two polar basins. VIC RRD yields the poorest results in most situations, with a median absolute values of ΔET_{RRD} exceeding 30% for continental, sub-arctic, and polar basins.



- 820 The compensation phenomenon is not an isolated case; it affects many basins regardless of the MDB considered (with or without a routing scheme) and can lead to significant estimation errors in P and ET volumes, especially over continental and sub-arctic basins.



825 **Table 2.** Median absolute values (%) of ΔP_{RRD} and ΔET_{RRD} for each RRD and each climate. The colour bar used for each cell is provided below the table.

CLIMAT	VIC		NOAH		CLSM		ISBA		ERA5-LAND		MERRA-2		JRA-55		GRUN	
	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}	ΔP_{RRD}	ΔET_{RRD}
Tropical	7.9	17.8	4.8	6.6	0.0	10.5	4.5	5.5	7.8	4.5	7.8	14.3	4.1	5.7	0.0	0
Dry	10.1	19.1	5.2	1.2	12.4	5.5	1.4	3.4	4.8	6.2	0.6	0	7.6	6.7	6.9	1.2
Temperate	13.0	14.8	6.6	8.5	0.0	12.2	0.0	8.5	1.9	6.6	12.5	11.2	3.0	8.8	1.9	0
Continental	15.4	34.4	11.7	27.5	12.7	27.5	8.3	19.3	8.3	1.4	8.3	14.8	8.3	18.4	0.4	0
Subarctic	17.8	44.5	11.7	28.6	13.3	24.1	22.4	21.4	13.3	14.9	14.6	17.1	12.6	11.5	10.2	5.7
Polar	26.9	63.3	30.3	20.6	36.1	28.7	30.3	29.4	6.4	10.5	54.8	20.6	28.6	11.2	30.3	5.7

05101520253035404550556065

colour bar (%)

830

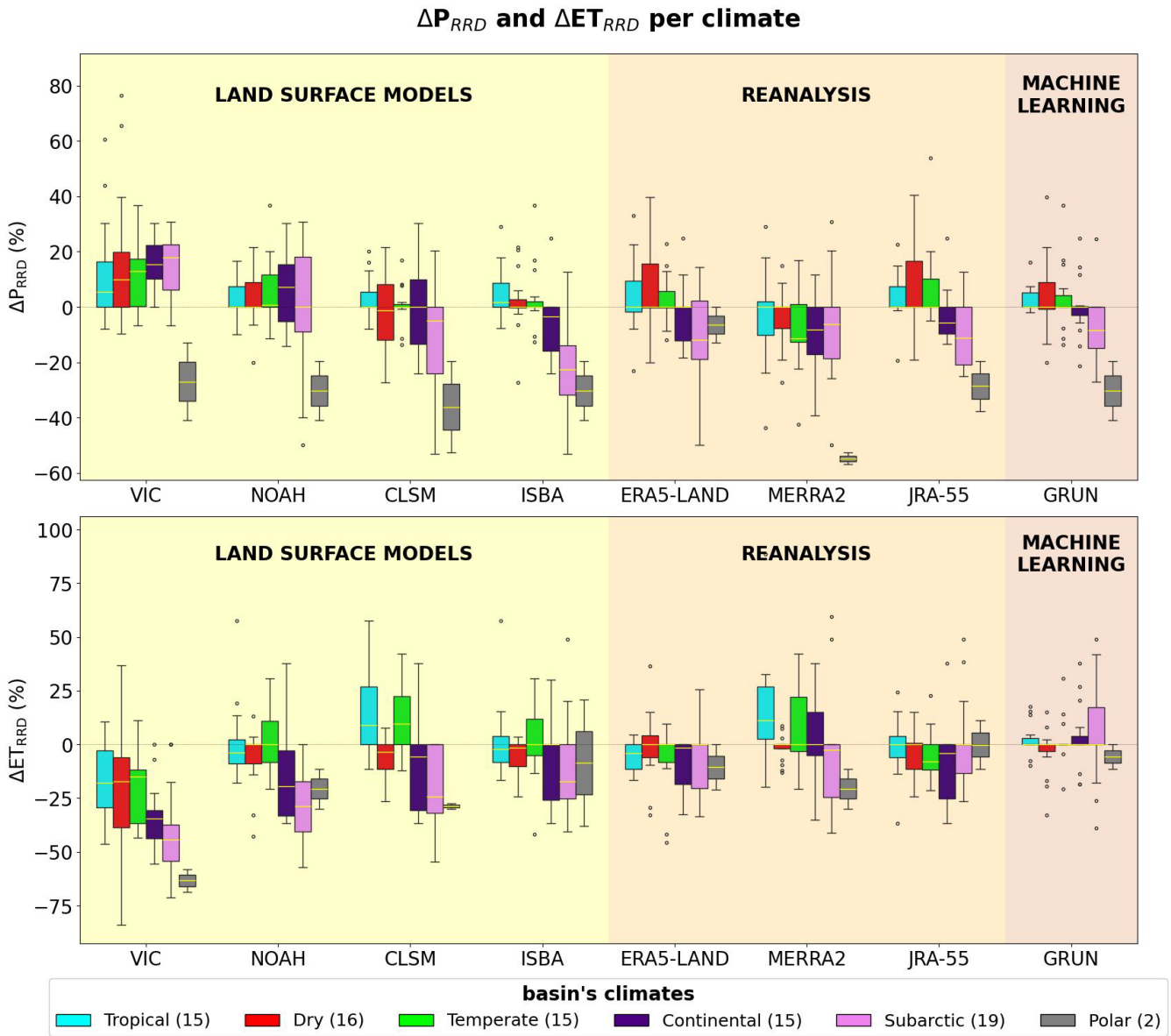


Figure 10 : ΔP_{RRD} and ΔET_{RRD} obtained for each model-based discharge and each basin's climate type. Median values are indicated in yellow.



4.3.4 Cumulated results

Estimating water volume of each component of the WBE at the global scale remains a major scientific question. Many uncertainties still exist regarding the global estimation of P (Koutsoyiannis, 2020; Rodell et al., 2015), ET (Miralles et al., 2016) and Q (Ghiggi et al., 2019; Müller Schmied et al., 2021). The cumulative OBD (Q_{OBDc}) for the 82 basins used in this study represents a mean discharge of 18,073 km³ per year, that is to say almost half of the total annual discharge to the ocean (Q_{REFc}), currently estimated between approximately 38,500 and 39,800 km³ per year (Ghiggi et al., 2019; Müller Schmied et al., 2021). Thus, some of the cumulative results from this study can be considered to approach those that would be obtained using a global approach, especially regarding the ratio P/Q and ET/Q.

The global annual P volume (P_{REFc}) is between 116,500 km³ and 123,300 km³ per year (Koutsoyiannis, 2020; Rodell et al., 2015), while the global annual ET volume (ET_{REFc}) is of between 62,200 km³ and 76,200 km³ per year (Miralles et al., 2016). It gives a P_{REFc}/Q_{REFc} ratio of between 2.9 and 3.2 and a ET_{REFc}/Q_{REFc} ratio of between 1.5 and 1.9.

When summing the results of this study for all 82 basins with OBD, we obtain a P_{OBDc} cumulated value of 55,970 km³ per year (P_{OBDc}) and a ET_{OBDc} cumulated value of 37,860 km³ per year (ET_{OBDc}) (Fig 11). It means a P_{OBDc}/Q_{OBDc} ratio of 3.09 consistent with P_{REFc}/Q_{REFc} uncertainties box and a ET_{OBDc}/Q_{OBDc} ratio of 2.09 slightly above the ET_{REFc}/Q_{REFc} uncertainties box. Global evaporation over land estimation uncertainties (10%) are more than twice the one of total precipitation over land (4.4%) and river discharge (4%) (Dorigo et al., 2021). The larger uncertainties associated with evapotranspiration estimation may help explain the differences we observe in the ET / Q ratios.

Despite these differences, the cumulative results of WBE can help improve or reduce the uncertainties of the estimation of P, ET and Q global annual volumes. Thus, the same question that arises at the basin scale also applies at global scales: What type of error can be expected when calculating the annual global volumes of P, ET, and Q by summing the results obtained from the WBE of all the world's basins?

Using the GLDAS VIC RRD in the WBEs leads to an overestimation of Q and P and an underestimation of ET, whereas the MERRA-2 RRD produces the opposite: underestimating Q and P and slightly overestimating ET (Fig. 11). For the remaining RRDs, the cumulative P and ET estimates are very close to those obtained with OBD. The JRA-55 RRD, in particular, delivers excellent overall results, with cumulative estimates of Q, P, and ET virtually identical to those derived from OBD.

VIC RRD overestimates the annual cumulative discharge volume significantly more than other RRD models (+83%). However, it provides an annual cumulative volume very close to that of OBD at the Obidos station on the Amazon, almost identical to the estimate by ERA5-Land RRD, which, in contrast, overestimates only slightly the cumulative volume (+10%). As mentioned in Sect. 4.1, the performance of RRD models varies from one basin to another and results at the basin level should not be generalized.



870 The P / Q and ET / Q ratios derived from the cumulative WBE results based on OBD agree quite well with values reported in
the literature. For a given MBD, an overestimation of P and ET for some basins and underestimation for others during the
calculation of the WBE are observed (Sect. 4.3.3). At the global scale, these estimation errors tend to average out, which allows
for a global WBE quite close to that obtained using OBD. The use of MBD at the global scale therefore still appears relatively
pertinent, but care must be taken not to apply global conclusions at the basin scale. Finally conducting WBE with OBD with
875 an exhaustive basin-by-basin approach may provide a reliable estimate of P and ET volume at global scale.

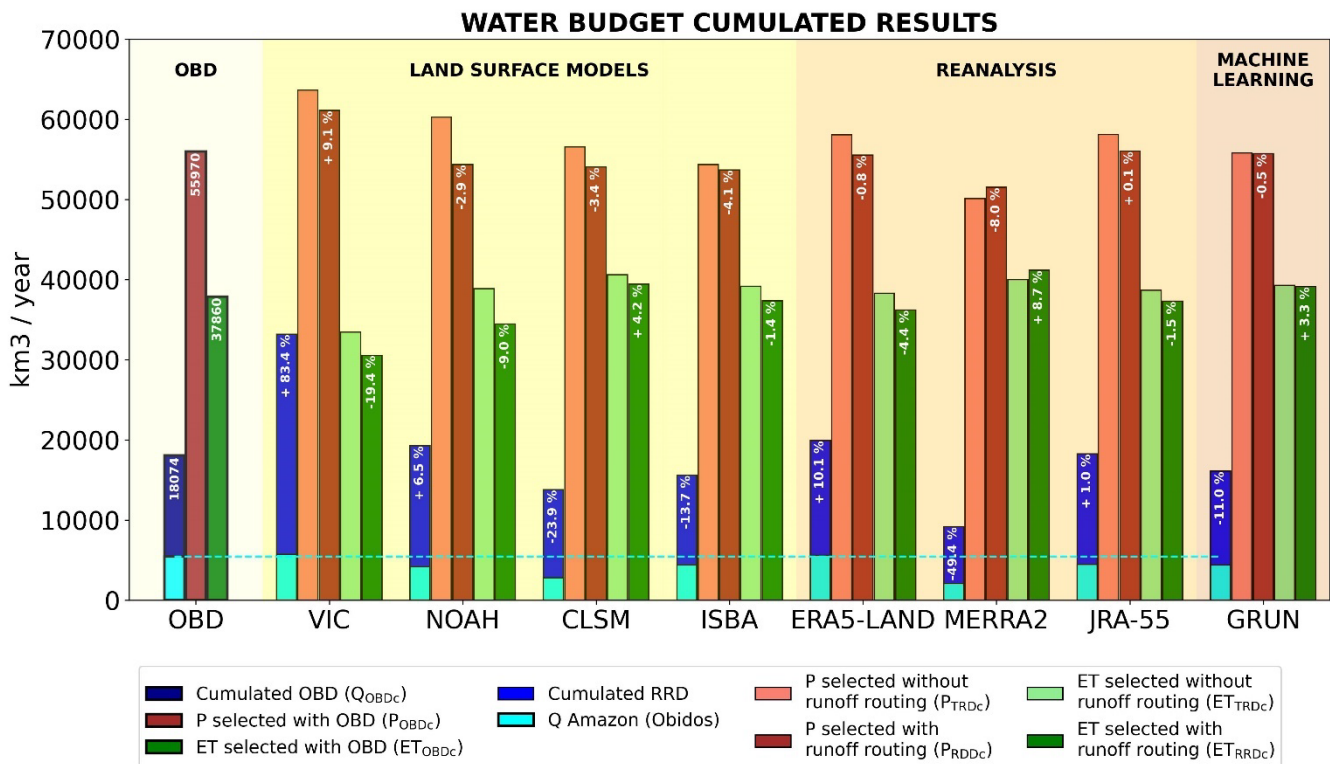


Figure 11: 82 basins cumulated average annual volume of discharge (blue), precipitation (red), and evapotranspiration (green) for, when combined together, provide the highest NSE for OBD and each MBD. Light red and light green correspond to the cumulated precipitation and evapotranspiration volume for TRD, while red and green to the cumulated precipitation and evapotranspiration volume for RRD. Light blue is the average estimated annual volume of discharge for the Amazon River at Obidos gauge station.

5 Conclusions and perspectives

This study aimed to determine how the source of discharge can affect the conclusions drawn from a water-budget assessment at basin scale, and to identify the types of basins in which the corresponding errors are greatest.

A previous study used reinforcement methods to evaluate the errors made in WBE by substituting Q with runoff (Luo et al., 2026). In the present work we instead adopt the three-point filtering method, whose importance have been demonstrated in Section 3.4. By leveraging satellite altimetry, we also propose a technique that enables the computation of WBE for basins where in-situ discharge measurements are unavailable or too sparse. This allowed us to include major basins such as the Ganges, Brahmaputra, Nile, Mekong, Indus, or Volga, and thanks to the large number of selected basins, to perform a statistical analysis for six distinct climate types. Moreover, this study is built on an innovative ensemble product derived from GRACE and GRACE-FO TWS measurements.



Our results show that great caution is needed when discharges are calculated simply by summing the runoff produced by a model. Such an approach can lead to a substantial overestimation of discharge, especially in highly anthropogenically altered arid basins (e.g., Indus, Rio Bravo, or Nile), or to an underestimation, as observed in high-altitude basins (e.g., Brahmaputra and Upper Indus).

900 The comparable relative variability of precipitation, evapotranspiration, and model-based discharge datasets can produce compensations errors within the water-budget equation. Although discharge is the smallest component of the balance, the variability between datasets obtained by substituting it with runoff is, on average, only slightly lower in magnitude than that of precipitation datasets and higher than that of evapotranspiration datasets. In continental and sub-arctic basins, the variability of runoff datasets can even exceed that of precipitation. Consequently, a discharge that is overestimated by a model can be
905 easily offset in the water-budget equation by using datasets that overestimate precipitation and/or underestimate evapotranspiration.

We have shown that a high level of water-budget closure does not, by itself, guarantee that each of its components is reliably estimated. Similar high closure levels can be achieved despite significant differences in the estimated annual volumes of the individual components.

910 The use of routing helps reduce discharge estimation errors for some basins, particularly large sub-arctic and tropical basins, but not for most basins. In many cases, the error appears to be more related to the estimation of high flow than to the modelling of its propagation, especially in arid and continental basins. In general, substituting discharge with runoff should be evaluated carefully before drawing conclusions from a water budget assessment.

Employing discharge derived from direct observations, whether from in situ data alone or from in situ data supplemented by
915 satellite observations, such as radar altimetry, helps reduce uncertainties and better constrain the water budget, thereby increasing confidence in the estimation of each of its components.

In the future, we aim to generalize and improve discharge estimation using satellite altimetry at basin outlets. The goal is a global, basin-by-basin water-budget calculation with the highest possible precision in discharge. These results could improve
920 the overall accuracy of estimates of land precipitation, land evaporation, and ocean discharge.

In this study we were, however, unable to derive water-surface elevations from satellite altimetry for the Pearl River (China), the Krishna River (India) and the Rio Bravo. For the first two rivers the problem most likely stems from a multipath phenomenon that produces multiple peaks in the altimeter waveform. In such cases, using retracers (i.e. algorithm to retrieve WSE from altimetry waveforms) of handling multi-peak waveforms could overcome this issue. For the Rio Bravo the difficulty
925 arises from the river's narrow width at its outlet, making it unlikely that even more sophisticated processing could derive a reliable WSE from satellite altimetry. In the most difficult situations, narrow rivers, nearby lakes, or lack of a satellite nadir altimeter observation, other satellite-based discharge products could be used to complement in situ data (Cerbelaud et al., 2025). For rivers without gauge measurements we plan to employ discharge from the Surface Water and Ocean Topography



(SWOT) mission (Andreadis et al., 2025; Durand et al., 2023). These discharges will be complemented with data from
930 conventional radar altimetry satellites in the same way that we complement in situ discharge data in this paper.

To go further, we also plan to expand the range of precipitation datasets we use. The FROGS database contains several products
(e.g., CHIRPS, 3B42, GSMAP, CMORPH, PERSIANN-CDR, TAMSAT, ARC) that were omitted from the present analysis
because they either lack high-latitude coverage or are only available regionally. Adding these datasets will improve the
robustness of the water-budget closure, providing important elements to understand the boundaries regarding regional
935 uncertainties.

Our results confirm that three-point filtering is necessary in the water-budget equation. Developing a reinforcement technique
similar to Constrained Kalman Filtering (CKF) (Abolafia-Rosenzweig et al., 2021; Pan et al., 2012; Pan and Wood, 2006),
Multiple Colocation (MCL) (Pan et al., 2015) or Minimized Series Deviation method (MSD) (Luo et al., 2023), that explicitly
940 accounts for three-point filtering, (which is not the case for the moment), should enable a more rigorous estimation of the
different components of the water budget equation.

Most water-budget studies, including the present one, do not account for groundwater discharge. Yet up to 25 % of estuaries
can receive groundwater inputs greater than 25% of the river discharge (Luijendijk et al., 2020). Future assessments should
945 therefore incorporate this flux, at least for basins where it is known to be significant.

Finally, although using observed discharges reduces compensation errors, it does not eliminate them. Even with reliable
discharge data, precipitation and evapotranspiration datasets can still compensate for each other. Defining a method to assess
these compensations and to minimise them would be particularly valuable.

950



Appendix A: Monthly Water budget Computation and results

Most TWS products derived from the GRACE and GRACE-FO missions are provided at a monthly resolution. Therefore, in order to calculate a WBE we must cast the water-budget equation (Eq. A1) in monthly form.

$$\frac{dTWS}{dt} = P(t) - ET(t) - Q(t) \quad (A1)$$

TWS(t), P(t), ET(t) and Q(t) refer in Eq. A1 to the daily data of respectively, TWS, precipitation, evapotranspiration and discharge. When we move to a monthly framework, the symbols P_n , ET_n , and Q_n represent the total volumes of water that enter (P) or leave (ET and Q) the basin between the last day of the month and the first day of the month (Eq. A2, Eq. A3 and Eq. A4).

$$P_n = \sum_{t=\text{month } n \text{ first day}}^{\text{month } n \text{ last day}} P(t) \quad (A2)$$

$$ET_n = \sum_{t=\text{month } n \text{ first day}}^{\text{month } n \text{ last day}} ET(t) \quad (A3)$$

$$Q_n = \sum_{t=\text{month } n \text{ first day}}^{\text{month } n \text{ last day}} Q(t) \quad (A4)$$

Equation A2 is the monthly version of Eq. A1 and the time increment dt in Eq. A1 can be taken as one month, i.e., $dt=1$

$$TWS(t = \text{month } n \text{ last day}) - TWS(t = \text{month } n \text{ first day}) = P_n - ET_n - Q_n \quad (A5)$$

TWS data derived from the GRACE and GRACE-FO missions for month n (TWS_n) can be interpreted as the mean TWS over that month (Eq. A6)

$$TWS_n = \frac{1}{\text{month } n \text{ number of days}} \sum_{t=\text{first day month } n}^{\text{last day month } n} TWS(t) \quad (A6)$$

970

Because TWS is a cumulative quantity, the most reliable point estimate is the value at the middle of the month (Eq. A7). In fact, the gravimetric records released by the various processing centers (JPL, GSFC, CSR, ...) are dated to the month-midpoint. This midpoint assumption is justified when the rate of change of TWS is roughly constant throughout the month, or at least when it does not vary dramatically. It breaks down, however, if TWS reaches a maximum or minimum near the middle of the month; in such cases using the mid-month value would lead to a systematic over- or under-estimate (see Fig. A1).

975

$$TWS_n \approx TWS(t=\text{middle of month } n) \quad (A7)$$

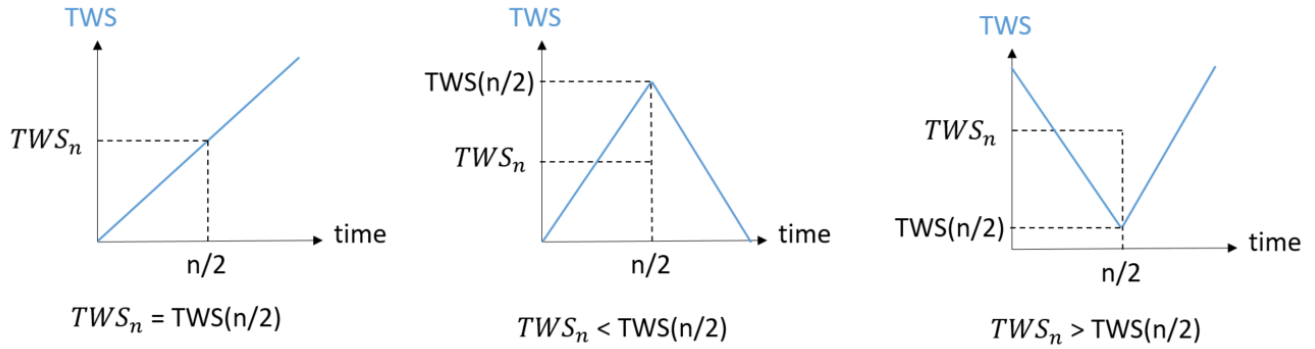


Figure A1: Comparison of monthly average TWS value with the TWS value at the middle of the month n for three scenario: (a) linear increase of TWS during the month, (b) TWS reaches a maximum at the mid-month, (c) TWS reaches a minimum at mid-month

$TWS(t = \text{month } n \text{ last day})$ and $TWS(t = \text{month } n \text{ first day})$ from Eq. (A2) can therefore be estimated by performing a linear interpolation between the data for the current month and the data for the following month and the previous month, respectively.

$$TWS(t = \text{month } n \text{ last day}) \approx \frac{TWS(t=\text{middle } n) + TWS(t=\text{middle } n+1)}{2} = \frac{TWS_{n+1} + TWS_n}{2} \quad (\text{A8})$$

$$TWS(t = \text{month } n \text{ first day}) \approx \frac{TWS(t=\text{middle } n-1) + TWS(t=\text{middle } n)}{2} = \frac{TWS_{n-1} + TWS_n}{2} \quad (\text{A9})$$

We finally obtain the quantities for the left-hand side of Eq. (A1), as is done in many papers and as recommended by (Landerer et al., 2010)

$$TWS(\text{month } n \text{ last day}) - TWS(\text{month } n \text{ first day}) = \frac{TWS_{n+1} - TWS_{n-1}}{2} \quad (\text{A10})$$

The interpolation presented in Eq. (A10) reveals data from months $n-1$ and $n+1$. It is therefore necessary to account for these in the right-hand side of Eq. (A1). We define $PEQ(t)$ and PEQ_n as follow:

$$PEQ(t) = P(t) - ET(t) - Q(t) \quad (\text{A11})$$

$$PEQ_n = \sum_{t=\text{first day month } n}^{\text{last day month } n} PEQ(t) \quad (\text{A12})$$

Since TWS_n can be considered the best estimate of the mid-month TWS, we can therefore rewrite:

$$TWS_{n+1} - TWS_{n-1} \approx \sum_{t=\text{middle of month } n-1}^{\text{middle of month } n+1} PEQ(t) \quad (\text{A13})$$

$$TWS_{n+1} - TWS_{n-1} \approx \sum_{t=\text{middle day of month } n-1}^{\text{last day of month } n-1} PEQ(t) + \sum_{t=\text{first day of month } n}^{\text{last day of month } n} PEQ(t) + \sum_{t=\text{first day of month } n+1}^{\text{middle day of month } n+1} PEQ(t) \quad (\text{A14})$$

$$TWS_{n+1} - TWS_{n-1} \approx \sum_{t=\text{middle day of month } n-1}^{\text{last day of month } n-1} PEQ(t) + PEQ_n + \sum_{t=\text{first day of month } n+1}^{\text{middle day of month } n+1} PEQ(t) \quad (\text{A15})$$



If the PEQ values are continuous throughout the month—i.e., the differences in precipitation, evapotranspiration, and discharge volumes between the beginning and the end of the month are negligible—then we can write:

$$\sum_{t=\text{middle day of month } n-1}^{\text{last day of month } n-1} \text{PEQ}(t) \approx \frac{1}{2} \sum_{t=\text{first day of month } n-1}^{\text{last day of month } n-1} \text{PEQ}(t) = \frac{1}{2} \text{PEQ}_{n-1} \quad (\text{A16})$$

$$\sum_{t=\text{first day of month } n+1}^{\text{middle day of month } n+1} \text{PEQ}(t) \approx \frac{1}{2} \sum_{t=\text{first day of month } n+1}^{\text{last day of month } n+1} \text{PEQ}(t) = \frac{1}{2} \text{PEQ}_{n+1} \quad (\text{A17})$$

And then obtain Eq. (A18)

$$\frac{\text{TWS}_{n+1} - \text{TWS}_{n-1}}{2} = \frac{1}{4} \text{PEQ}_{n-1} + \frac{1}{2} \text{PEQ}_n + \frac{1}{4} \text{PEQ}_{n+1} \quad (\text{A18})$$

We then find the filtering proposal found in (Landerer et al., 2010) and apply in (Lehmann et al., 2022; Lorenz et al., 2014).

$$\frac{\text{TWS}(n+1) - \text{TWS}(n-1)}{2} = \tilde{P}(n) - \tilde{ET}(n) - \tilde{Q}(n) \quad (\text{A19})$$

$$\tilde{F}(n) = \frac{1}{4} F(n-1) + \frac{1}{2} F(n) + \frac{1}{4} F(n+1) \quad (\text{A20})$$

To conclude Eq. A18 can be considered as valid only if:

- TWS does not reach a maximum or a minimum around the middle of the month $n-1$, n and $n+1$
- The daily PEQ data are relatively uniform for the months $n-1$ and $n+1$.

Note: The three-point filtering method has the disadvantage of requiring discharge data for three consecutive months. If a single month is missing in the discharge data, TWS cannot be estimated not only for this specific month, but also for the previous and the following months.

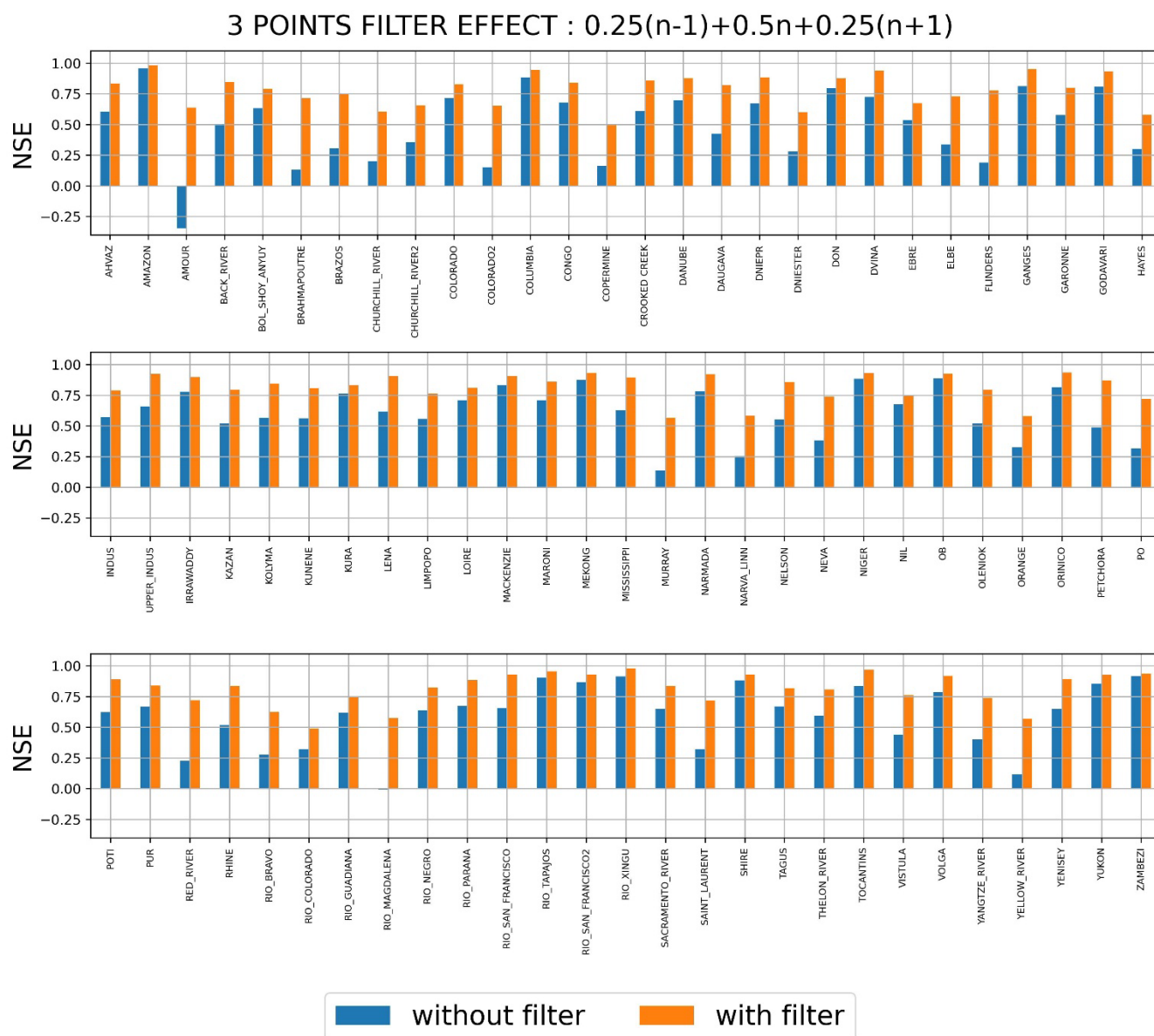


Figure A2 : Highest NSE_{0BD} obtained with and without the 3 points filter for each basin



Appendix B: CCI discharge evaluation

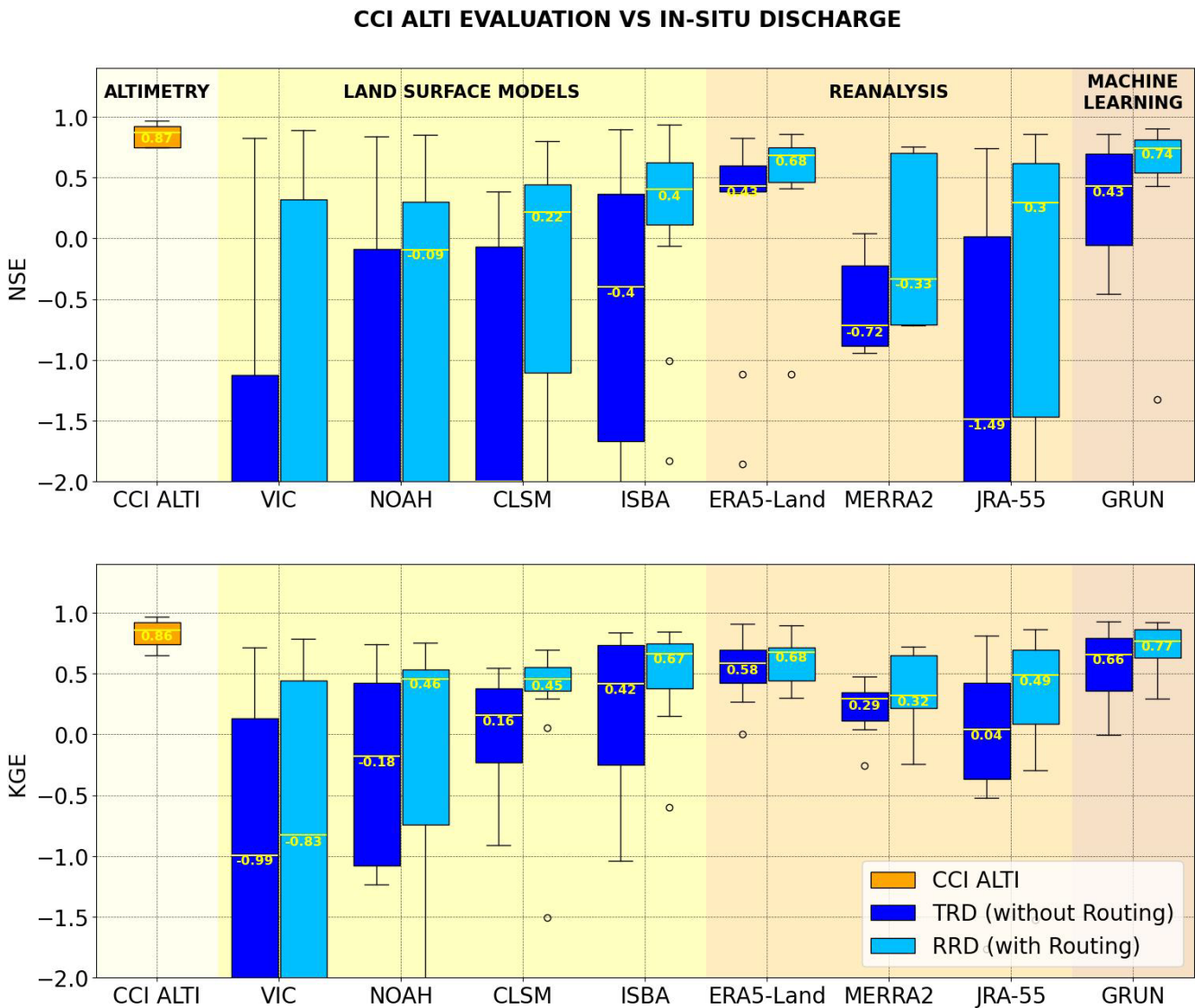


Figure B1 : Boxplot of NSEs (top) and KGEs (down) for Amazon, Congo, Danube, Garonne, Irrawaddy, Lena, Mackenzie, Maroni, Mississippi, Niger, and Ob between in situ discharge and the Altimetry CCI discharge (orange), TRD (light blue) and RRD (blue). Median values are indicated in yellow.). There is a zoom on both figure: minimum y axis value is -2

1030



1035 **Appendix C: Water budget**

On the Amazon, Ob, and Columbia (Fig. C2 to C4), we observe a strong impact of considering routing, which ameliorates the level of closure and changes the conclusions on precipitation and evapotranspiration for most models.

1040 The Rio Bravo, classified as arid, is the basin on which we obtained the highest discharge overestimation, up to almost +10,000% with GLDAS VIC. This error paradoxically has limited impact on the WBE conclusions (Fig. C6), as even when overestimated, the different RRD volume remain significantly lower than precipitation and evapotranspiration volume.

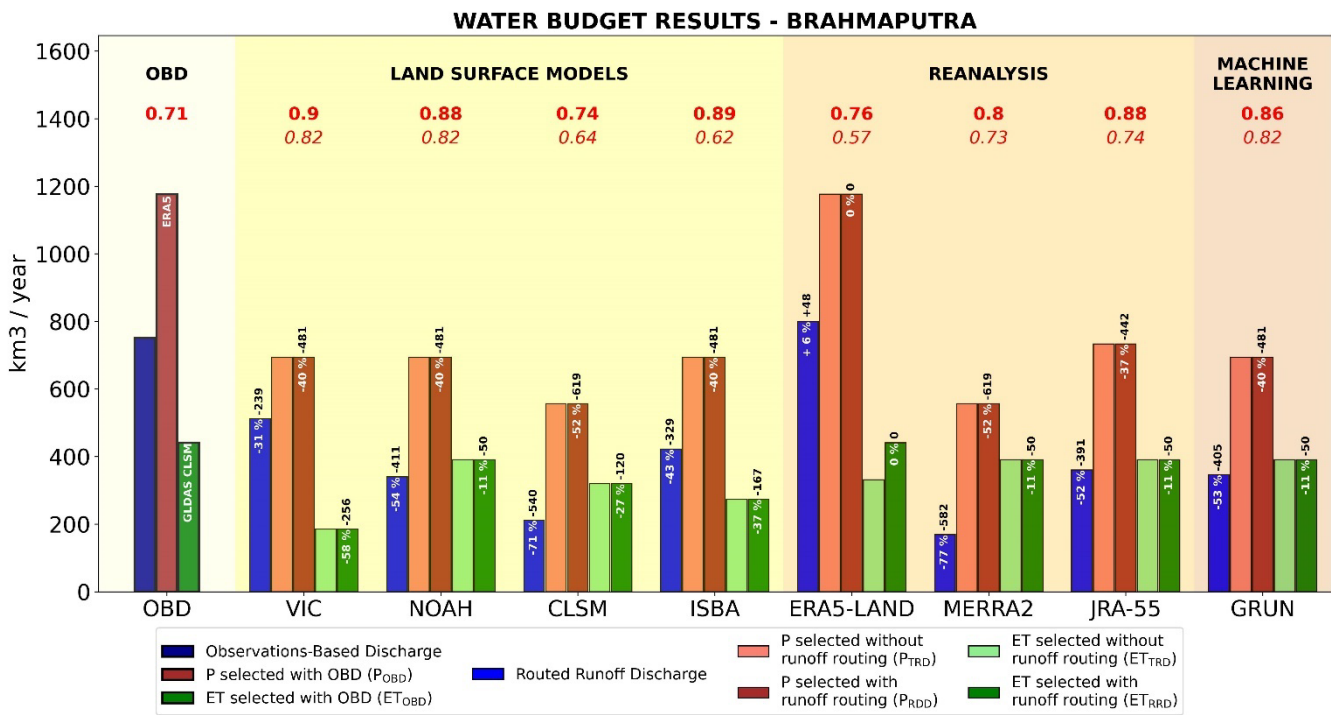


Figure C1: Brahmaputra WBEs

1045

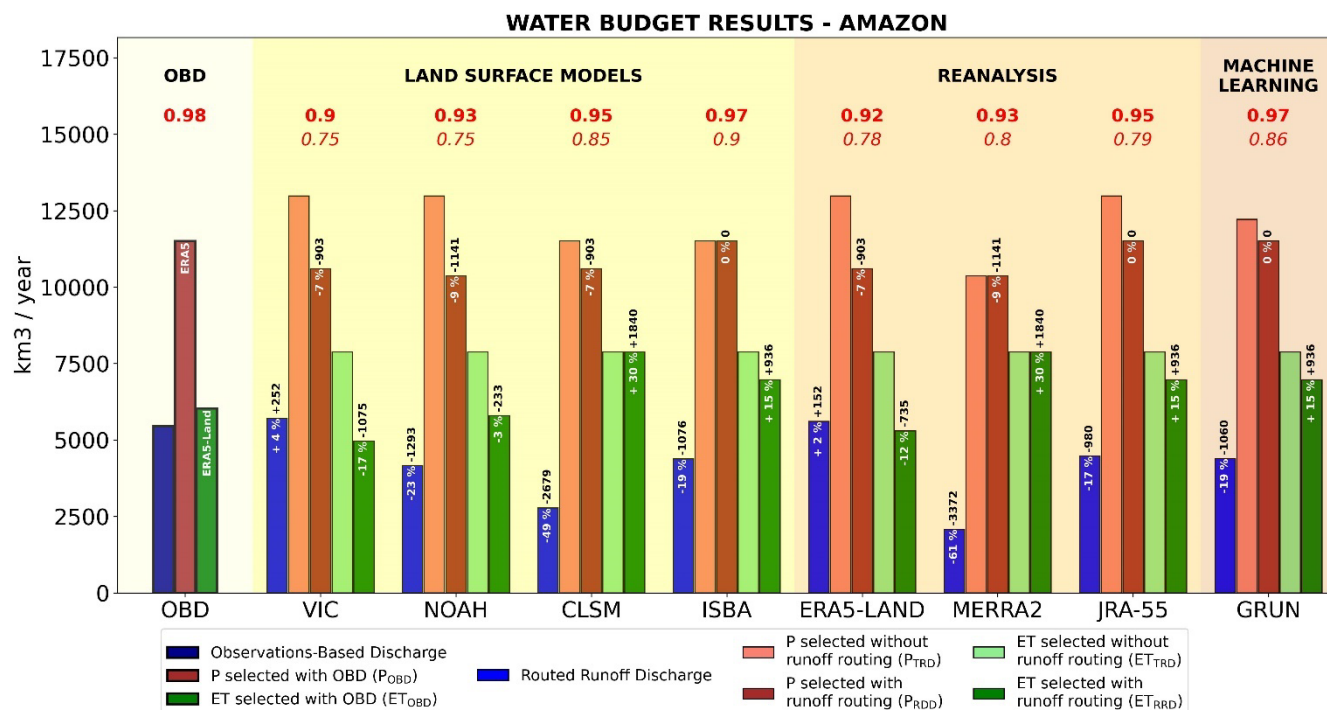


Figure C2: Amazon WBEs

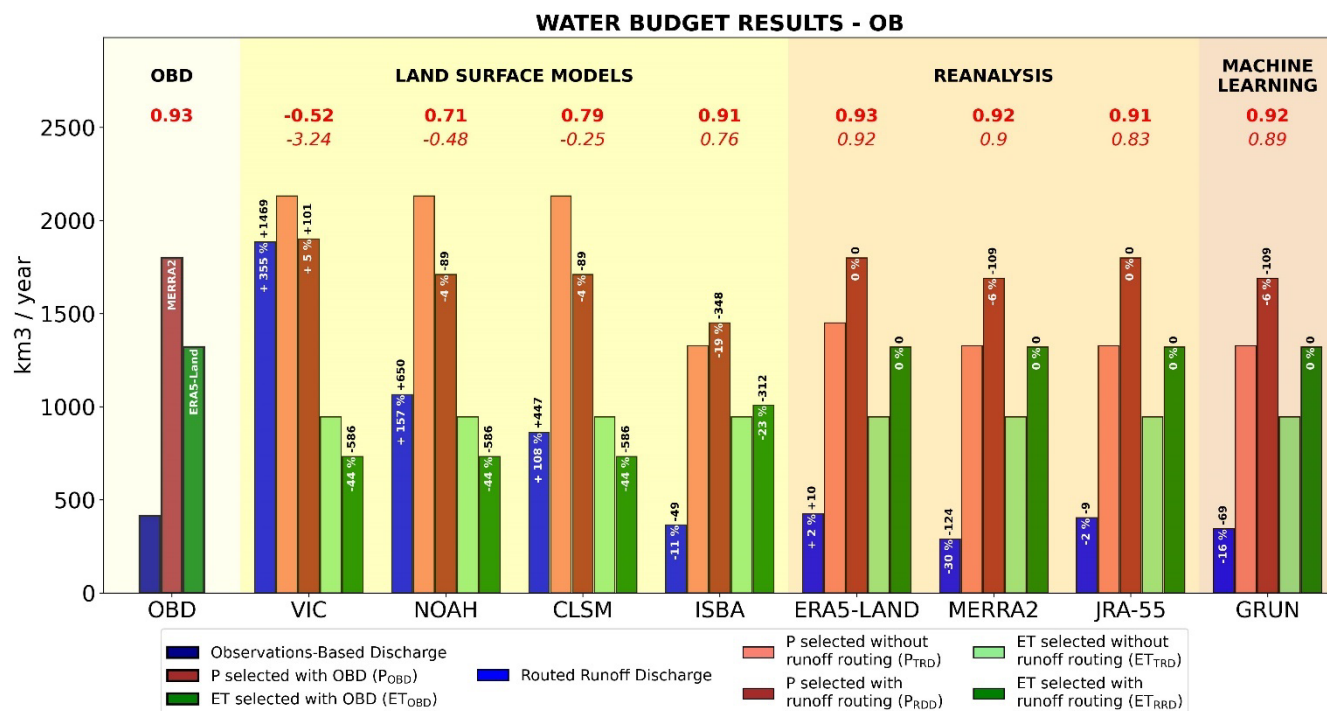


Figure C3: Ob WBEs



1050

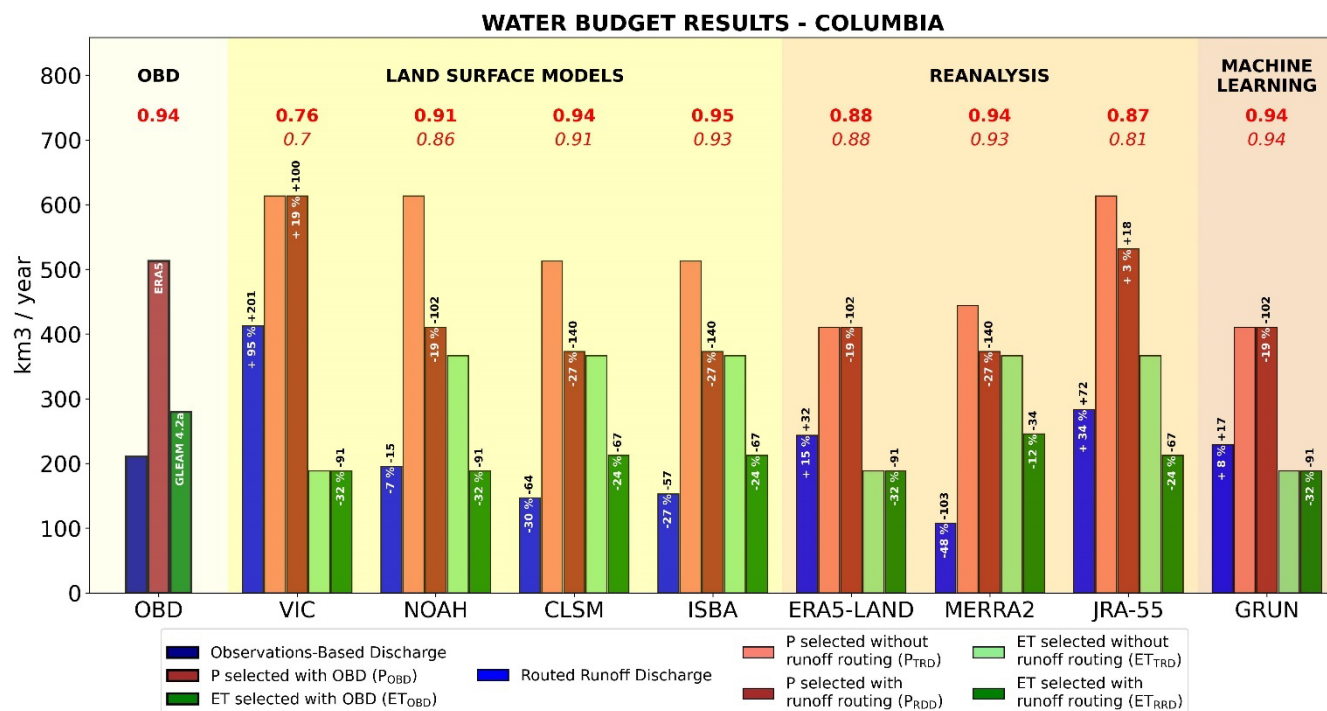


Figure C4: Columbia WBEs

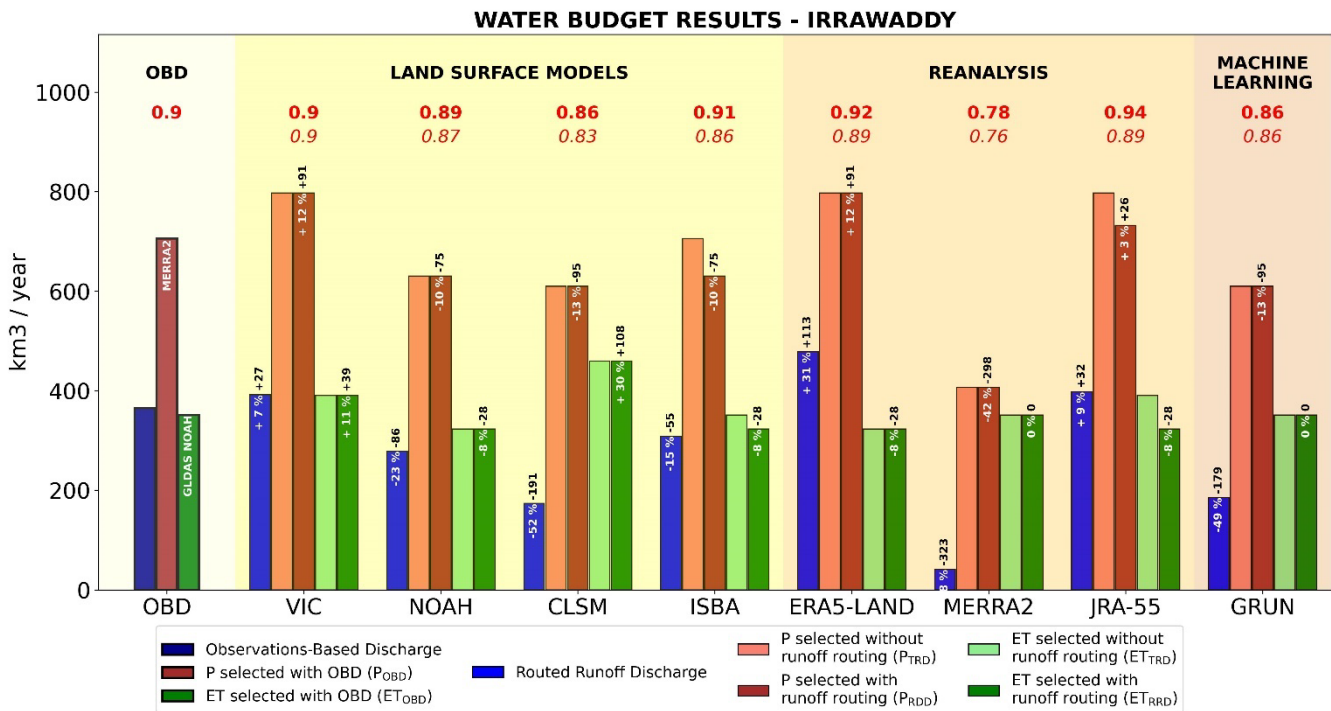


Figure C5: Irrawaddy WBEs

1055

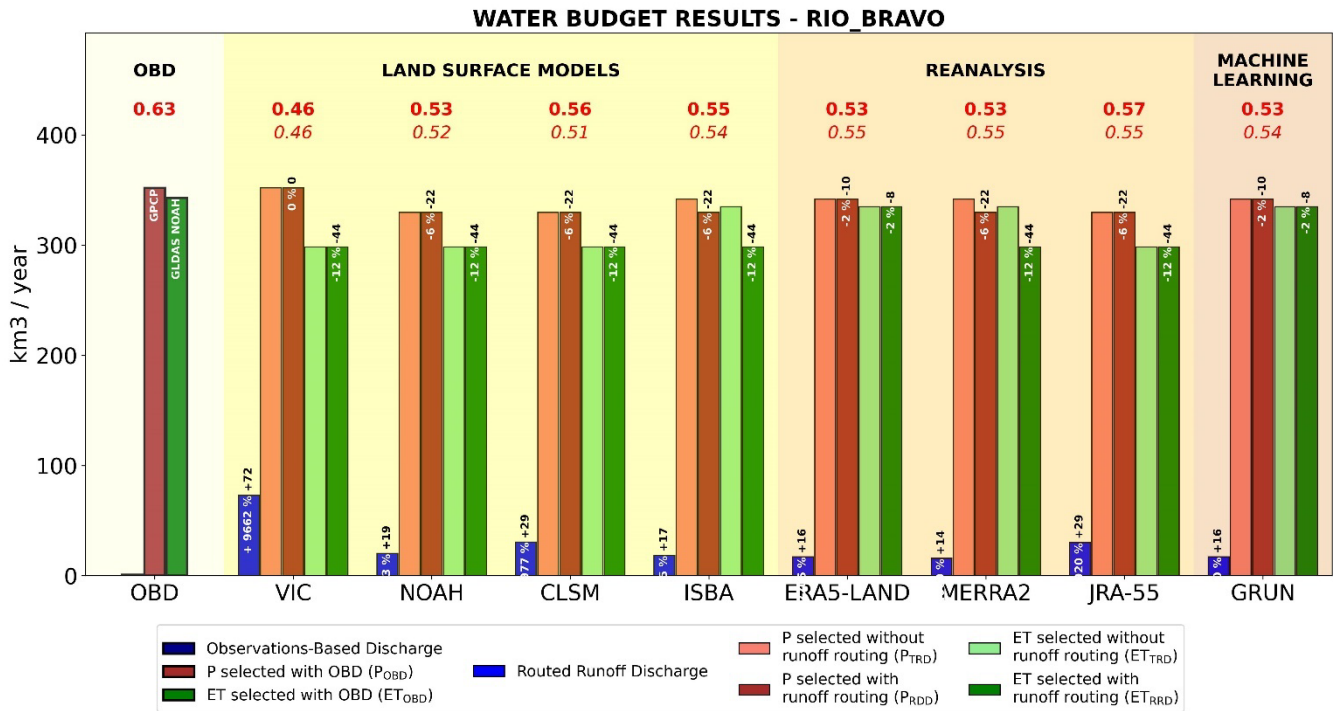


Figure C6: Rio Bravo WBEs



1065 **Appendix D: Error on P and ET**

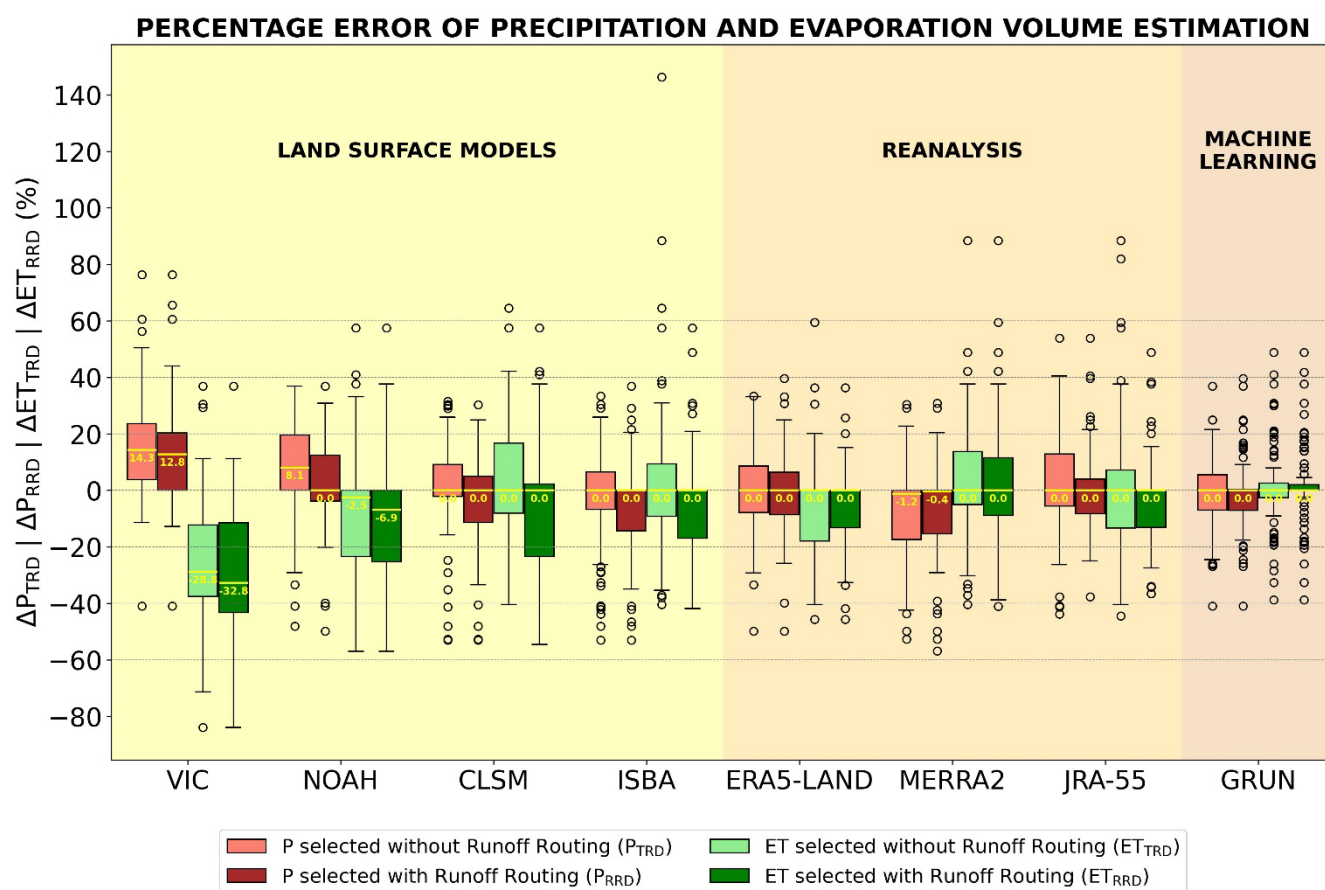


Figure D1 : Boxplot of ΔP_{TRD} , ΔP_{RRD} , ΔET_{TRD} , and ΔET_{RRD} values for all 82 basins. Median values are indicated in yellow.



1070

Table D1 : Quartile, median, extreme values, mean absolute value (MAV), and median absolute value (MDAV) of ΔP_{RRD} , and ΔP_{TRD} . The last two columns represent the mean values of the 8 TRD and RRD.

	VIC		NOAH		CLSM		ISBA		ERA5-Land		MERRA2		JRA-55		GRUN		MEAN	
	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD
min	-41.0	-41.0	-48.1	-49.8	-53.0	-53.0	-53.0	-53.0	-49.8	-49.8	-52.7	-56.9	-43.8	-37.7	-41.0	-41.0	-47.8	-47.8
Q1	3.8	0.0	0.0	-3.8	-2.2	-11.3	-6.8	-14.3	-7.8	-8.7	-17.3	-15.4	-5.6	-8.3	-7.0	-7.1	-5.4	-8.6
med	14.3	12.8	8.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.2	-0.4	0.0	0.0	0.0	0.0	2.6	1.5
Q3	23.6	20.3	19.5	12.5	9.1	5.0	6.5	0.0	8.6	6.4	0.0	0.0	12.9	4.0	5.5	0.3	10.7	6.1
max	76.4	76.4	36.8	36.8	31.5	30.3	33.4	36.8	33.4	39.7	30.4	30.9	53.9	53.9	36.8	39.7	41.6	43.1
MAV	17.0	15.7	14.4	10.6	11.5	11.1	13.2	11.8	10.3	9.8	12.9	12.4	13.4	10.0	8.9	8.2	12.7	11.2
MDAV	14.5	13.2	11.7	8.4	7.3	7.4	7.2	7.6	8.8	8.1	11.7	11.4	9.7	6.3	6.0	5.4	9.6	8.5

1075

Table D2 : Quartile, median, extreme values, mean absolute value (MAV), and median absolute value (MDAV) of ΔET_{RRD} , and ΔET_{TRD} . The last two columns represent the mean values of the 8 TRD and RRD.

	VIC		NOAH		CLSM		ISBA		ERA5-Land		MERRA2		JRA-55		GRUN		MEAN	
	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD
min	-84.0	-84.0	-57.0	-57.0	-40.4	-54.6	-40.4	-41.8	-45.7	-45.7	-40.4	-41.1	-44.5	-36.6	-38.8	-38.8	-48.9	-49.9
Q1	-37.5	-43.3	-23.4	-25.3	-8.0	-23.4	-9.2	-17.0	-17.9	-13.2	-4.9	-8.9	-13.3	-13.1	-2.6	0.0	-14.6	-18.0
med	-28.8	-32.8	-2.5	-6.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-3.9	-5.0
Q3	-12.3	-11.5	0.0	0.0	16.7	2.2	9.4	0.0	0.0	0.0	13.8	11.5	7.2	0.0	2.6	2.1	4.7	0.5
max	36.9	36.9	57.5	57.5	64.5	57.5	146.3	57.5	59.5	36.4	88.4	88.4	88.4	48.8	48.8	48.8	73.8	54.0
MAV	28.2	29.6	16.4	16.5	17.0	15.9	15.8	13.3	11.8	9.9	14.5	14.4	15.8	11.0	9.3	8.0	16.1	14.8
MDAV	30.4	33.3	14.1	13.8	14.5	11.8	9.6	10.4	7.7	6.6	11.3	10.6	11.2	8.9	2.8	0.0	12.7	11.9



1080

Table D3 : median absolute value (MDAV) of ΔP_{RRD} , and ΔP_{TRD} for each climate and each MDB. The last two columns represent the mean values of the 8 TRD and RRD.

CLIMAT	VIC		NOAH		CLSM		ISBA		ERA5-Land		MERRA2		JRA-55		GRUN		MEAN	
	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD
Tropical	5.7	7.9	5.4	4.8	1.3	0	3.8	4.5	5.7	7.8	4.1	7.8	5.7	4.1	0	0	4.0	4.6
Dry	8.6	10.1	8	5.2	6.8	12.4	0.5	1.4	7.1	4.8	0	0.6	10.7	7.6	8.8	6.9	6.3	6.1
Temperate	13.4	13	6.6	6.6	1.9	0	7.6	0	5	1.9	15.5	12.5	4.4	3	7.6	1.9	7.8	4.9
Continental	17.5	15.4	17.5	11.7	8.3	12.7	10.3	8.3	5.2	8.3	11.7	8.3	9.2	8.3	5.8	0.4	10.7	9.2
Subarctic	20.2	17.8	24	11.7	19.9	13.3	27.1	22.4	13.5	13.3	16.5	14.6	24	12.6	13.8	10.2	19.9	14.5
Polar	20.9	26.9	30.3	30.3	26.7	36.1	30.3	30.3	0.4	6.4	36.1	54.8	28.6	28.6	20.5	30.3	24.2	30.5



colour bar (%)

Table D4 : median absolute value (MDAV) of ΔET_{RRD} , and ΔET_{TRD} for each climate and each MDB. The last two columns represent the mean values of the 8 TRD and RRD.

CLIMAT	VIC		NOAH		CLSM		ISBA		ERA5-Land		MERRA2		JRA-55		GRUN		MEAN	
	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD	TRD	RRD
Tropical	26.1	17.8	13.4	6.6	4.8	10.5	5.5	5.5	9.9	4.5	14	14.3	9	5.7	2.6	0	10.7	8.1
Dry	26	19.1	5.3	1.2	7.8	5.5	1.9	3.4	6.6	6.2	0	0	7.5	6.7	3.3	1.2	7.3	5.4
Temperate	14.6	14.8	8.1	8.5	17.3	12.2	8.5	8.5	0	6.6	13.2	11.2	11.2	8.8	0	0	9.1	8.8
Continental	33.8	34.4	28.6	27.5	5.7	27.5	1.5	19.3	6.7	1.4	11.6	14.8	0	18.4	0	0	11.0	17.9
Subarctic	33.2	44.5	28.3	28.6	30.2	24.1	28.4	21.4	19.2	14.9	26.6	17.1	28.4	11.5	17.3	5.7	26.5	21.0
Polar	47	63.3	20.6	20.6	24.2	28.7	24.5	29.4	23	10.5	16.1	20.6	20.6	11.2	16.1	5.7	24.0	23.8



colour bar (%)



1085 Appendix E: In situ discharge error computation

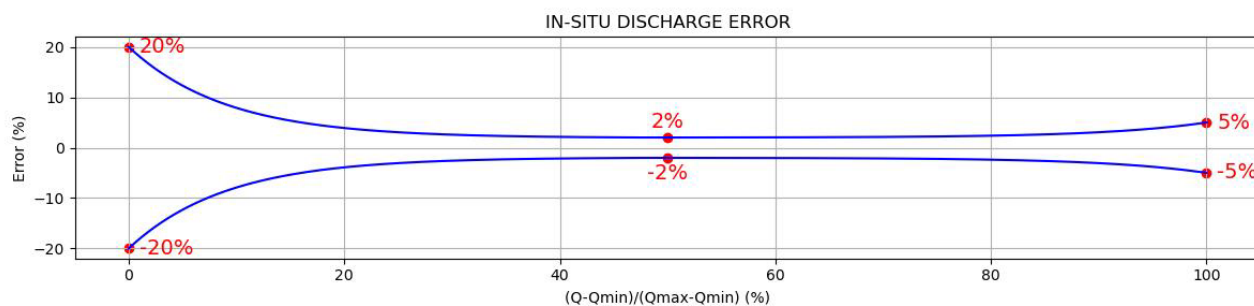


Figure E1 : In situ discharge percentage error used in this study



1090 **Appendix F: NSE_{OBD} , NSE_{GSFC} , NSE_{CSR} , and NSE_{JPL} for each basin**

	NSE_{OBD} (SAGSA)	NSE_{GSFC}	NSE_{CSR}	NSE_{JPL}	
AHVAZ	0.83	0.78	0.80	0.82	1.00
AMAZON	0.98	0.98	0.98	0.98	0.90
AMUR	0.64	0.62	0.42	0.56	0.80
BACK RIVER	0.85	0.88	0.80	0.79	0.70
BOL SHOY ANYUY	0.79	0.76	0.76	0.80	0.60
BRAHMAPOUTRE	0.71	0.67	0.67	0.72	0.50
BRAZOS	0.75	0.75	0.72	0.74	0.40
CHURCHILL RIVER (Manitoba)	0.60	0.59	0.37	0.39	0.30
CHURCHILL RIVER (Quebec)	0.66	0.66	0.68	0.60	0.20
COLORADO (Arizona-Nevada)	0.83	0.86	0.84	0.85	0.10
COLORADO (Texas)	0.65	0.63	0.65	0.65	0.00
COLUMBIA	0.95	0.95	0.96	0.94	-0.10
CONGO	0.84	0.79	0.79	0.79	-0.20
COPERMINE	0.50	0.63	0.62	0.64	-0.30
CROOKED CREEK	0.86	0.81	0.68	0.73	-0.40
DANUBE	0.88	0.89	0.88	0.89	-0.50
DAUGAVA	0.82	0.75	0.61	0.52	
DNIEPR	0.88	0.84	0.81	0.84	
DNIESTER	0.60	0.63	0.54	0.50	
DON	0.88	0.87	0.88	0.91	
DVINA	0.94	0.90	0.94	0.94	
EBRE	0.67	0.74	0.68	0.59	
ELBE	0.73	0.80	0.78	0.75	
FLINDERS	0.78	0.63	0.71	0.70	
GANGES	0.95	0.96	0.95	0.95	
GARONNE	0.80	0.84	0.68	0.66	
GODAVARI	0.93	0.94	0.93	0.92	
HAYES	0.58	0.49	0.15	-0.49	
INDUS	0.79	0.77	0.77	0.73	
UPPER INDUS	0.92	0.84	0.70	0.21	
IRRAWADDY	0.90	0.89	0.90	0.87	

NSE colour bar



KAZAN	0.80	0.79	0.71	0.66
KOLYMA	0.84	0.85	0.84	0.82
KUNENE	0.81	0.81	0.58	0.24
KURA	0.83	0.82	0.82	0.80
LENA	0.91	0.86	0.89	0.88
LIMPOPO	0.76	0.72	0.72	0.67
LOIRE	0.81	0.85	0.78	0.75
MACKENZIE	0.91	0.91	0.91	0.91
MARONI	0.86	0.86	0.85	0.81
MEKONG	0.93	0.94	0.92	0.93
MISSISSIPPI	0.89	0.91	0.89	0.90
MURRAY	0.57	0.58	0.53	0.57
NARMADA	0.92	0.93	0.93	0.91
NARVA LINN	0.58	0.58	0.35	0.35
NELSON	0.86	0.83	0.82	0.87
NEVA	0.74	0.76	0.68	0.71
NIGER	0.93	0.92	0.93	0.95
NIL	0.75	0.71	0.70	0.73
OB	0.93	0.92	0.92	0.93
OLENIOK	0.79	0.78	0.78	0.80
ORANGE	0.58	0.58	0.52	-0.02
ORINICO	0.93	0.93	0.92	0.92
PETCHORA	0.87	0.90	0.92	0.92
PO	0.72	0.72	0.74	0.76
POTI	0.89	0.85	0.81	0.84
PUR	0.84	0.83	0.80	0.82
RED RIVER (US Oklahoma)	0.72	0.68	0.70	0.77
RHINE	0.84	0.84	0.83	0.86
RIO BRAVO	0.63	0.42	0.63	0.69
RIO COLORADO (Argentina)	0.49	0.50	0.42	0.56
RIO GUADIANA	0.75	0.77	0.81	0.64
RIO MAGDALENA	0.57	0.38	0.75	0.68
RIO NEGRO (Argentina)	0.82	0.81	0.86	0.87
RIO PARANA	0.88	0.88	0.87	0.88
RIO SAN FRANCISCO	0.93	0.90	0.84	0.89
RIO TAPAJOS	0.96	0.94	0.95	0.94
RIO SAN FRANCISCO2	0.93	0.92	0.90	0.86



RIO XINGU	0.98	0.97	0.97	0.95
SACRAMENTO RIVER	0.84	0.86	0.88	0.86
SAINT LAURENT	0.72	0.79	0.77	0.76
SHIRE	0.93	0.93	0.90	0.91
TAGUS	0.82	0.78	0.81	0.67
THELON RIVER	0.81	0.83	0.74	0.72
TOCANTINS	0.97	0.95	0.96	0.96
VISTULA	0.76	0.83	0.78	0.80
VOLGA	0.92	0.89	0.91	0.91
YANGTZE RIVER	0.74	0.75	0.71	0.77
YELLOW RIVER	0.57	0.44	0.57	0.45
YENISEY	0.89	0.88	0.92	0.92
YUKON	0.93	0.90	0.92	0.95
ZAMBEZI	0.94	0.91	0.92	0.92



Appendix G: RRD and OBD climatology

DISCHARGES CLIMATOLOGY

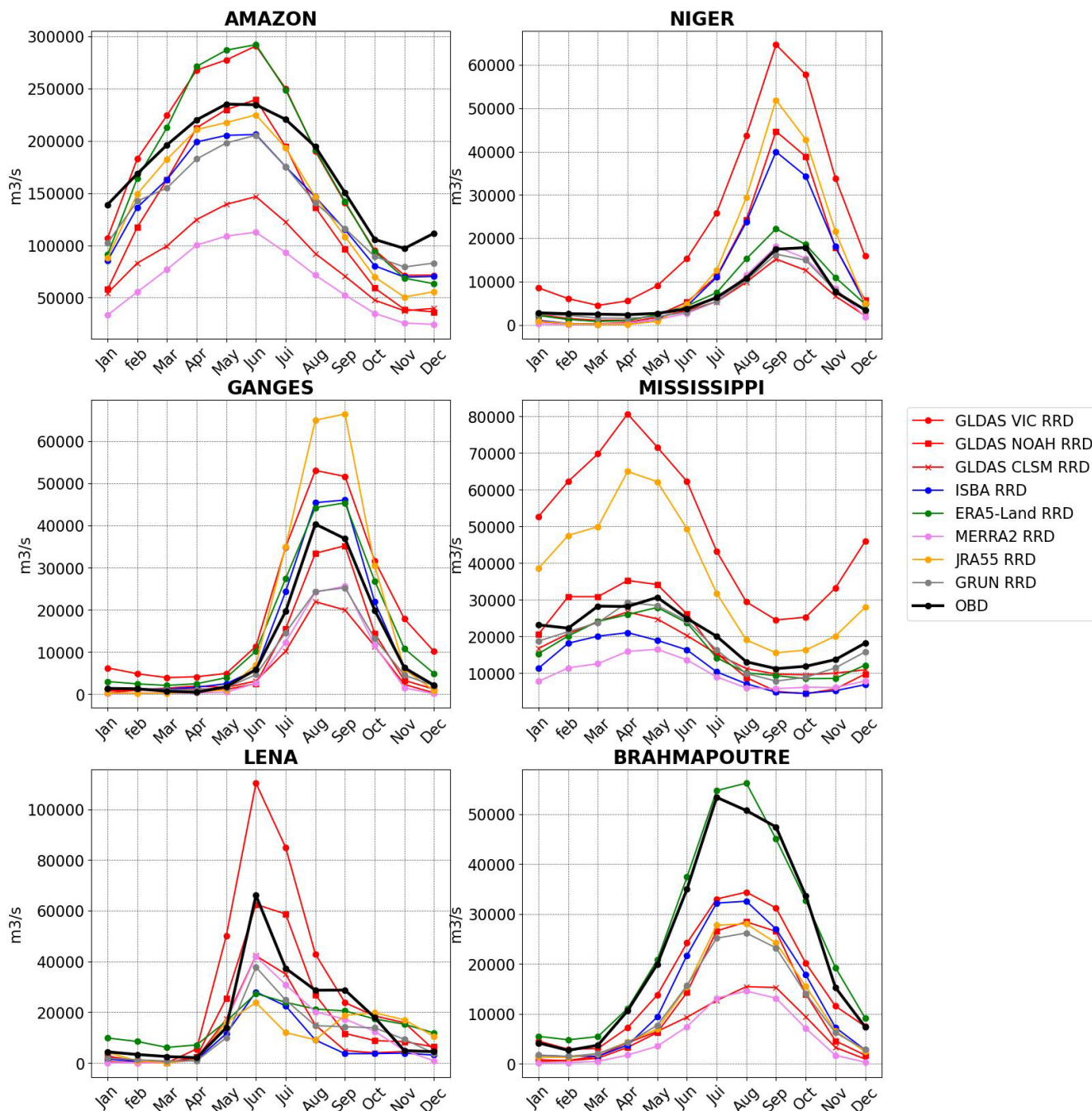


Figure G1 : RRD and OBD monthly climatology for Amazon, Niger, Ganges, Mississippi, Lena and Brahmaputra.



DISCHARGES CLIMATOLOGY

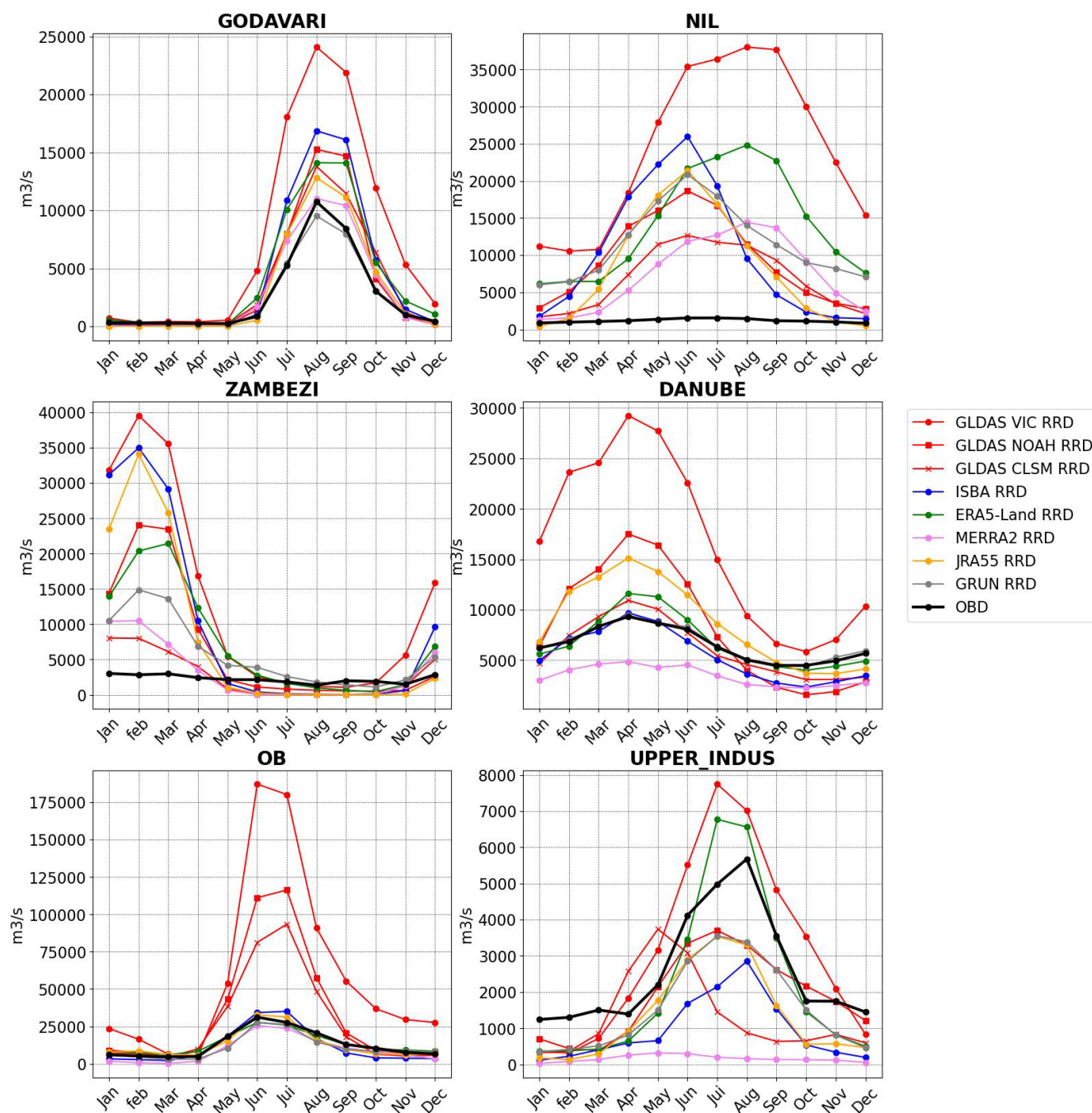


Figure G2 : RRD and OBD monthly climatology for Godavari, Nile, Zambezi, Danube, Ob and Upper Indus.



DISCHARGES CLIMATOLOGY

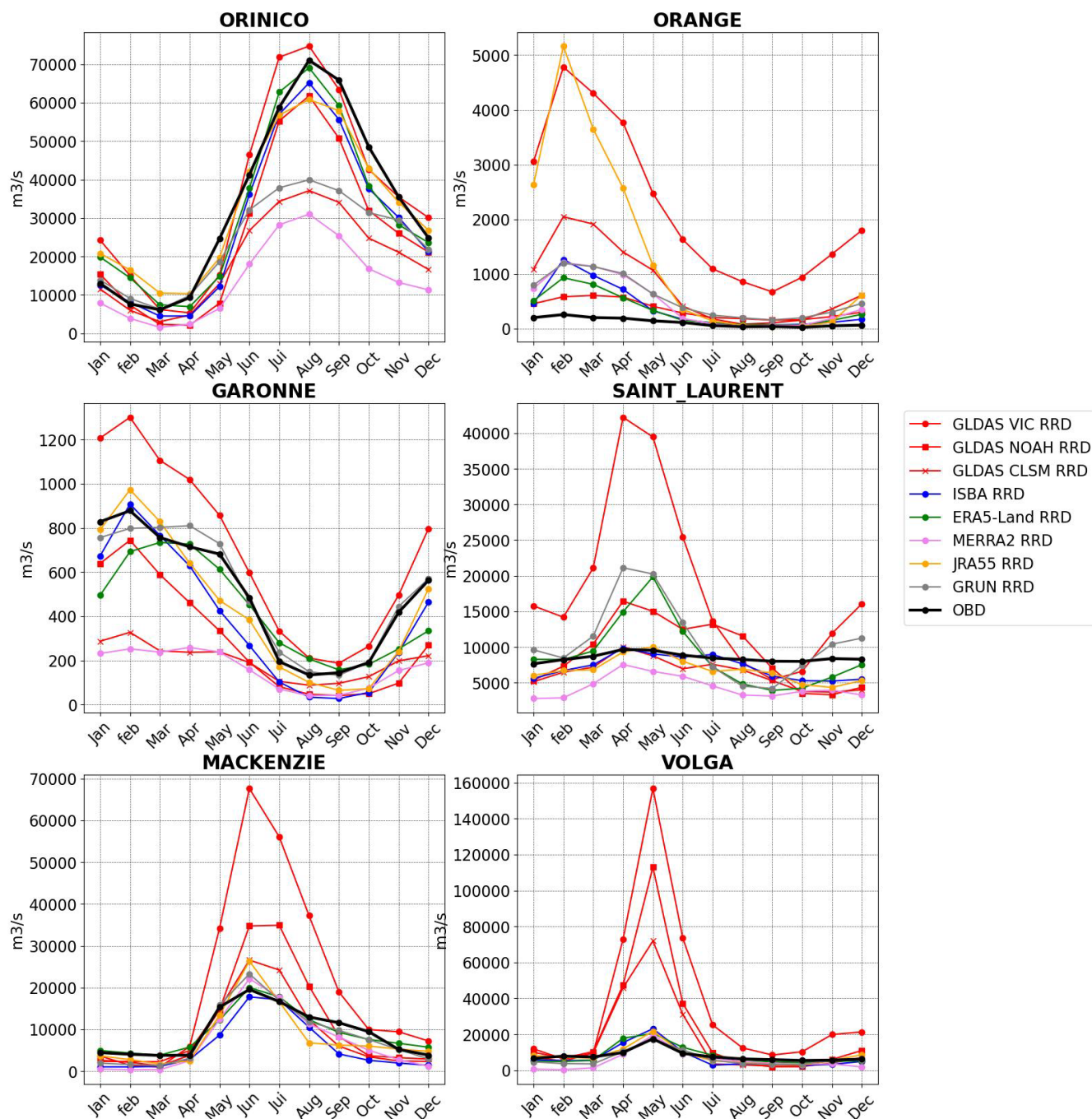
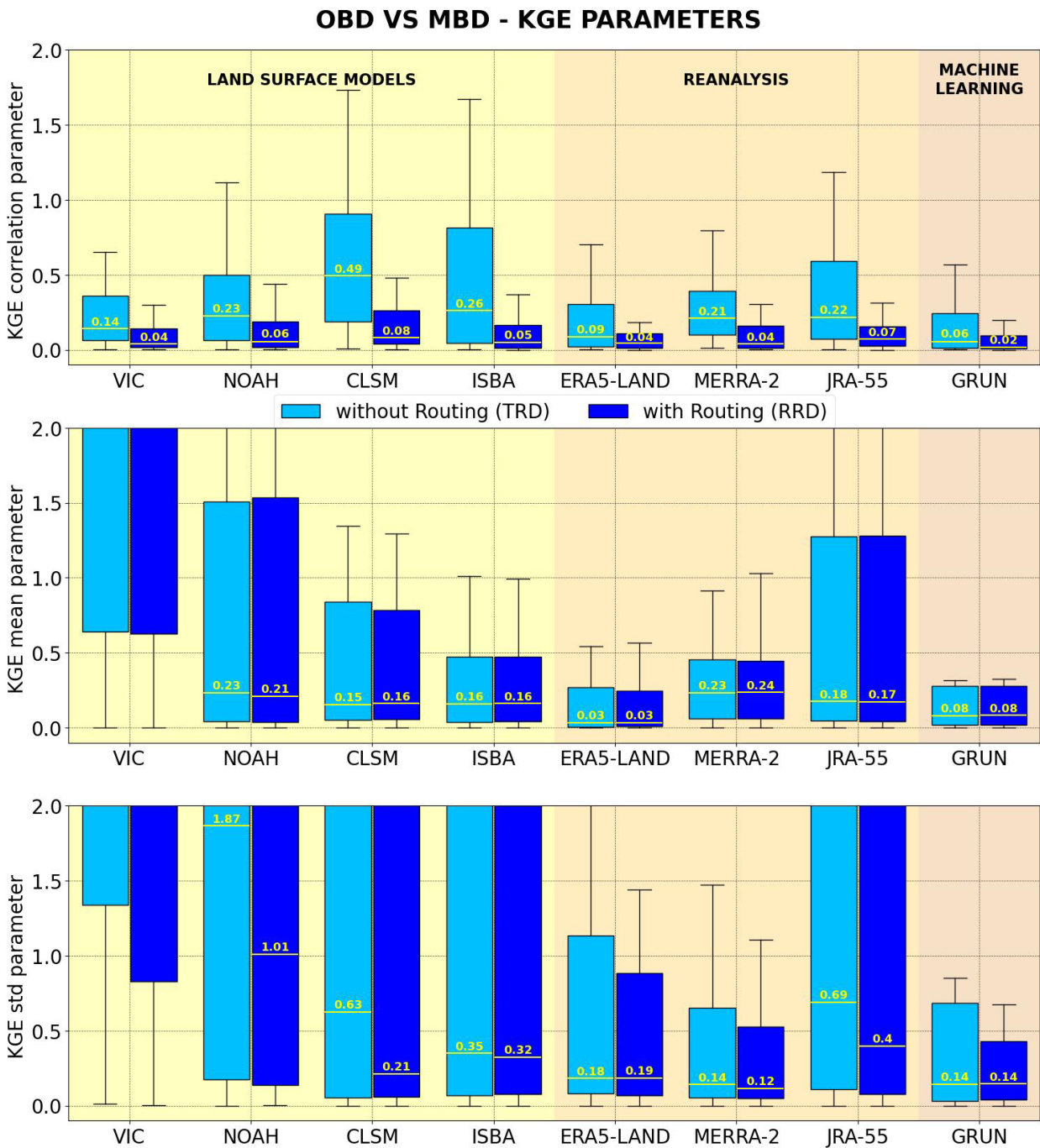


Figure G3 : RRD and OBD monthly climatology for Orinico, Orange, Garonne, Saint Laurent, Mackenzie and Volga River.



Appendix H: KGE(OBD,MBD) decomposition (as described in Sect 3.1)



1105 Figure H1 : KGE decomposition when comparing OBD and MBD. Top: KGE_{corr} boxplot, middle: KGE_{mean} boxplot, bottom: KGE_{std} boxplot (see Eq. (3) to Eq. (6)). Outliers are not plotted; y axis is limited to 2.

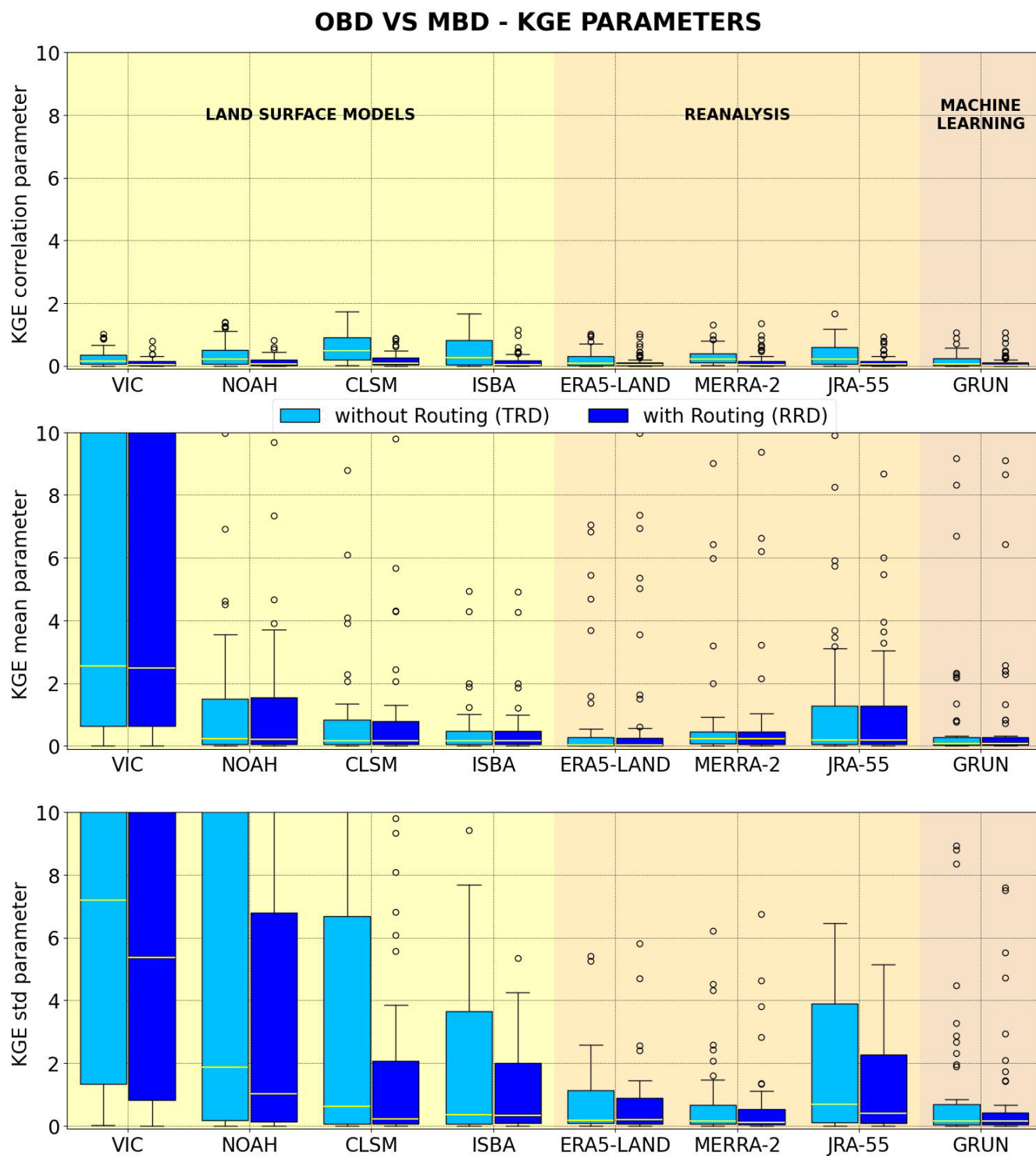


Figure H2 : KGE decomposition when comparing OBD and MBD. Top: KGE_{corr} boxplot, middle: KGE_{mean} boxplot, bottom: KGE_{std} boxplot (see Eq. (3) to Eq. (6)). Y axis is limited to 10.



1110

Appendix I: Routed River Discharge (RRD) computation

The detailed computation of the RRD is presented in this section.

As described in Sect. 3.2.3, the path between each pixel included in the $1^\circ \times 1^\circ$ sub-basin mask and the gauge station is computed with the TRIP routing model (Oki and Sud, 1998). The distance between two consecutive pixels along the path is computing using the haversine formula between the centres of each pixel. The total distance from a given pixel to the gauge station is the sum of the distances between all successive pixel pairs along that path. Since the routing scheme does not account for meanders, the effective length of rivers is potentially significantly greater (TRIP meandering ratio has not been used, see discussion in Oki and Sud (1998)). Each $1^\circ \times 1^\circ$ pixel is then processed independently to calculate its runoff contribution to discharge as a function of its distance to the gauge station and of the considered speed.

1120 Consider:

- $P(\text{lon}, \text{lat})$, a pixel of the basin,
- $R_P(n)$ the runoff of pixel P in month n ,
- $Q_{P,S}(n)$, the discharge that pixel P delivers to the gauge station in month n considering S , the propagation speed,
- X_S , the quotient of the distance from the pixel to the station (km) divided by S (km / month)

1125 $X_{S,P}$ is a delay; it is the number of months required for $R_P(n)$ to reach the station. If X_S is an integer, then $Q_P(n)$ is equal to the total runoff of a previous month, ie $R_P(n-X_S)$. If X_S is not an integer, $Q_P(n)$ is derived from the runoff of the pixel over two consecutive months:

$$Q_{S,P}(n) = (1 - \{X_{S,P}\}) \times R_P(n - [X_{S,P}]) + \{X_{S,P}\} \times R_P(n - 1 - [X_{S,P}]) \quad (2)$$

With:

- 1130
- $[X_{S,P}]$: the integer part of X_S ,
 - $\{X_{S,P}\} = X_{S,P} - [X_{S,P}]$: the fractional part of X_S .

$Q_S(n)$, the total basin discharge for month n considering the propagation speed S , is then:

$$Q_S(n) = \sum_P Q_{S,P}(n) \quad (20)$$

1135 Cor_S , the correlation between the monthly climatology of Q_S and the monthly climatology of OBD is then computed for each S values, i.e., for S ranging from 0.1 m s^{-1} to 4 m s^{-1} (see Fig. I1 with ERA5-Land). The RRD is defined as the time series that attains the highest Cor_S value.



1140

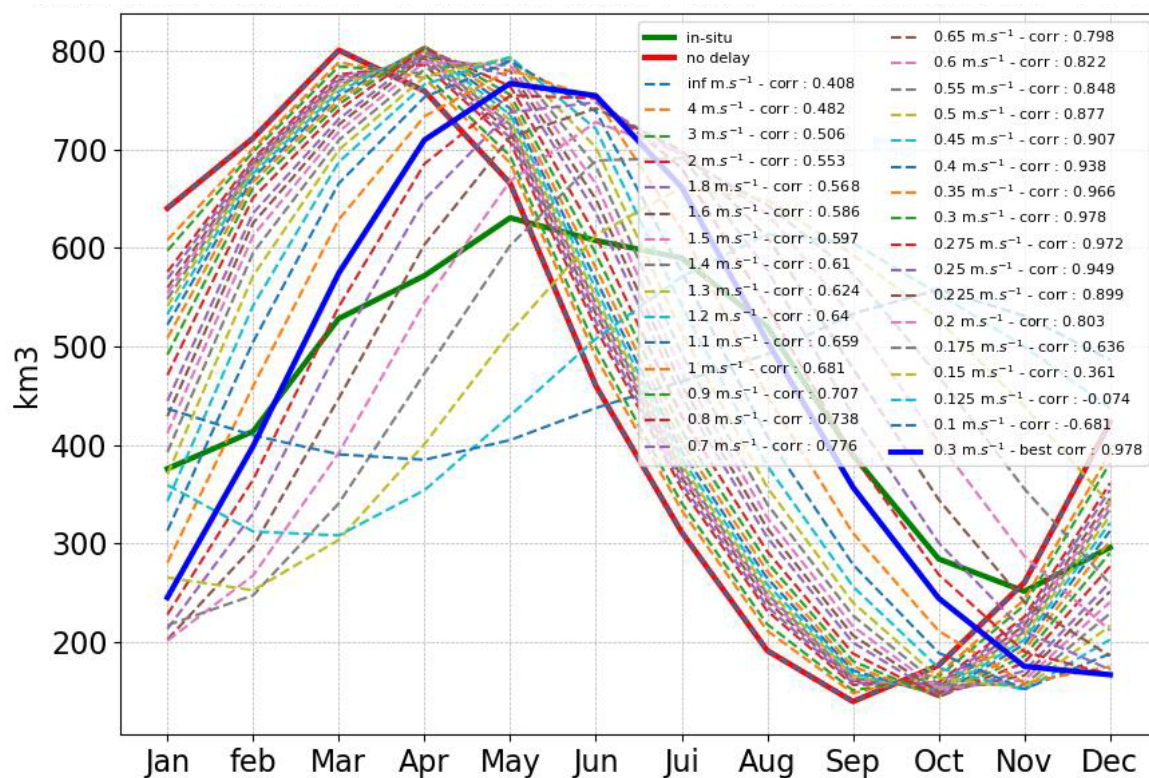


Figure I1 : ERA5-Land RRD propagation speed computation for the Amazon basin



1145 **Appendix J: Acronyms used in this study (acronyms in bold correspond to the ones defined in this study and never used in previous ones)**

ADCP	Acoustic Doppler Current Profiler
BESS	Breathing Earth System Simulator
CCI	Climate Change Initiative
CFSR	Climate Forecast System Reanalysis
CKF	Constrained Kalman Filtering
CLSM	Catchment Land Surface Model
CNES	Centre National d'Etudes Spatiales (French space agency)
CPC	Global Unified Gauge-Based Analysis Climate Prediction Center
CRV	Coefficient of Relative Variation
CSR	Center for Space Research
CTOH	Centre de Topographie de l'Océan et de l'Hydrosphère
CV	Coefficient of Variation
DDK	Denoising and Decorrelation Kernel
DTWS	Derivative Of Total Water Storage
ENVISAT	ENVironment SATellite
ERA5	fifth generation of the European Centre for Medium-Range Weather Forecasts atmospheric reanalysis
EROS	Earth Resources Observation and Science
ERS	European Remote-Sensing Satellite
ESA	European Space Agency
ET	EvapoTranspiration
FLDAS	Famine Early Warning Systems Network
FROGS	Frequent Rainfall Observations on GridS
GIA	Global Isostatic Adjustment
GIRAFE	Global Interpolated RAInFall Estimation
GLDAS	Global Land Data Assimilation System
GLEAM	Global Land Evaporation Amsterdam Model
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
GRACE	Gravity Recovery And Climate Experiment
GRDC	Global Runoff Data Centre
GRUN	Global Runoff Reconstruction
GSFC	Goddard Space Flight Center
HYBAM	HYdro-geochemistry



IMERG	NASA Integrated Multi-satellitE Retrievals for GPM
ISBA	Interaction Sol-Biosphère-Atmosphère
JPL	Jet Propulsion Laboratory
JRA-3Q	Japanese Reanalysis for Three Quarters of the Century
JRA-55	Japanese 55-year reanalysis
KGE	Kling–Gupta efficiency
LSM	Land Surface Model
MASCON	Mass Concentration
MBD	Model-Based Discharge
MCL	Multiple Colocation
MERRA-2	second version of Modern-Era Retrospective Analysis for Research and Applications
MSD	Minimized Series Deviation
NSE	Nash–Sutcliffe Efficiency
OBD	Observations-Based Discharge
P	Precipitation
PML	Penman Monteih Leuning
PR	Proportional Redistribution
Q	Discharge
RRD	Routed-Runoff Discharge
SAGSA	Service d’Activités en Gravimétrie Spatiale et Applications
SSEBop	Simplified Surface Energy Balance
SWOT	Surface Water and Ocean Topography
TRD	Total Runoff Discharge
TRIP	Total Runoff Integrating Pathways
TUGRAZ	Technical University of Graz
TWS	Total Water Storage
VIC	Variable Infiltration Capacity
WBE	Water Budget Evaluation
WSE	Water Surface Elevation

Data availability

The codes can be made available upon request by email to the first author.

- 1150 SAGSA TWS can be retrieved from the hydroweb.next website: <https://hydroweb.next.theia-land.fr/>.
 GSFC MASCON TWS were downloaded from <https://earth.gsfc.nasa.gov/geo/data/grace-mascons>
 CSR MASCON TWS were downloaded from <http://www2.csr.utexas.edu/grace>
 JPL MASCON TWS were downloaded from <https://podaac.jpl.nasa.gov/dataset/>



- FROGS P datasets were downloaded from the IPSL ftp server (<ftp://ftp.climserv.ipsl.polytechnique.fr/FROGs>)
- 1155 ERA5-Land ET and Runoff data were downloaded from <https://cds.climate.copernicus.eu/datasets/>
MERRA2 ET and Runoff data were downloaded from <https://disc.gsfc.nasa.gov/datasets/>
JRA-55 ET and Runoff data were downloaded from <https://gdex.ucar.edu/datasets/>
GLDAS VIC, Noah, and CLSM ET and Runoff data were downloaded from <https://disc.gsfc.nasa.gov/>
GLEAM ET datasets were downloaded from GLEAM ftp server, see <https://www.gleam.eu/#downloads>
- 1160 BESS V2 ET data were downloaded from
<https://drive.google.com/drive/folders/1i03DkZ0vV3pAnrhSrUYmexmUtQqMILDs>
SSEBop V5 ET data were downloaded from
<https://edcintl.cr.usgs.gov/downloads/sciweb1/shared/fews/web/global/monthly/etav5/downloads/>
PML V2 ET data were downloaded from
- 1165 https://code.earthengine.google.com/?asset=projects/pml_evapotranspiration/PML/OUTPUT/PML_V2_8day_v017_ARC_061
Synthesized ET data were downloaded from
<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZGOUED>
ISBA runoff data are developed at CNRM and available upon request to the authors
- 1170 GRUN runoff data were downloaded from https://figshare.com/articles/dataset/G-RUN_ENSEMBLE/12794075
GRDC in situ discharge data were downloaded from the GRDC website (<https://portal.grdc.bafg.de/>).
HYBAM in situ discharge data were downloaded from the HYBAM website (<https://hybam.obs-mip.fr/>).
Indus in situ river discharge comes from annual flood report documents downloaded from the Federal Flood Commission website (<https://www.ffc.gov.pk/>).
- 1175 Mekong in situ river discharge cannot be deposited publicly because of commercial constraints as they have been purchased from the Mekong River Commission (<https://www.mrcmekong.org/>)
CCI river discharge are available on the CCI website: <https://climate.esa.int/fr/data/>
Altimetry WSE data were downloaded from the CTOH website: <https://www.legos.omp.eu/ctoh/catalogue/>
- 1180 **Supplement link**
The file is called: Supplementary_Material.docx



Author contributions

JL, SB, and AB have designed the experiment. JL implemented the code and wrote the paper with support from all co-authors. AB has provided the SAGSA data. SM has provided the ISBA data. LG has computed the rating curve and discharge for the Nile and the Yangtze. RO has computed the calibration of GIRAFE precipitation data and validate the choice of precipitation datasets. JE has validated the choices of evapotranspiration datasets. JL, SB, AB, and SM have analysed the results.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

We thank the Centre National d'Etudes Spatiales (CNES) for funding Julien Lefebvre's position and for providing access to the TREX server where the calculations were performed. We thank NASA and DLR for providing data from the GRACE and GRACE-FO missions, as well as the various data production centres: CSR, JPL, TUGRAZ, CNES, and GFZ. We thank the CCI river discharge for providing the altimetry-based discharge.

We thank the various centres that produced the precipitation, evapotranspiration, and runoff datasets, in particular:

- the Institut Pierre-Simon Laplace (IPSL) for producing the FROGS database
- the Physical Sciences Laboratory (PSL) of the National Oceanic and Atmospheric Administration (NOAA) for producing the GPCP, CPC and GPCC precipitation datasets.
- the NOAA National Center Environmental Prediction (NCEP) for producing the CFSR reanalysis.
- NASA for producing the IMERG precipitation dataset, the MERRA-2 reanalysis and the VIC, Noah and CLSM GLDAS variants.
- The European Organization For the Exploitation of Meteorological Satellites (EUMETSAT) for producing the GIRAFE precipitation dataset and the Copernicus Climate Change Service (C3S) for providing it.
- The European Centre For Medium-Range Weather Forecasts (ECMWF) for producing the ERA5-Land reanalysis and C3S for providing it.



1210 - The Japan Meteorological Agency (JMA) for producing the JRA-55 and JRA-3Q reanalysis

We thank the centres that provide the available in situ discharge data (GRDC, HYBAM, Federal Flood Commission of Pakistan, Mekong River Commission).

1215 Finally, we thank CTOH for its ALTis software (<https://ctoh-docs.sedoo.fr/ALTis-NG/docs/wiki/ALTis-NG-Wiki-ia.html>) and its altimetry measurement database.

Financial support

The salary of J Lefebvre was supported by the CNES.

Review statement

1220 The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

- 1225 Abbott, B.W., Bishop, K., Zarnetske, J.P., Minaudo, C., Chapin, F.S., Krause, S., Hannah, D.M., Conner, L., Ellison, D., Godsey, S.E., Plont, S., Marçais, J., Kolbe, T., Huebner, A., Frei, R.J., Hampton, T., Gu, S., Buhman, M., Sara Sayedi, S., Ursache, O., Chapin, M., Henderson, K.D., Pinay, G., 2019. Human domination of the global water cycle absent from depictions and perceptions. *Nat. Geosci.* 12, 533–540. <https://doi.org/10.1038/s41561-019-0374-y>
- 1230 Abolafia-Rosenzweig, R., Pan, M., Zeng, J.L., Livneh, B., 2021. Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques. *Remote Sensing of Environment* 252, 112191. <https://doi.org/10.1016/j.rse.2020.112191>
- Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci.* 21, 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Aires, F., Pellet, V., 2024a. Synergizing AI and Physical Expertise to Close the Water Budget from Satellite Data. *Journal of Hydrometeorology* 25, 1303–1321. <https://doi.org/10.1175/JHM-D-23-0001.1>
- 1235 Aires, F., Pellet, V., 2024b. Synergizing AI and Physical Expertise to Close the Water Budget from Satellite Data. *Journal of Hydrometeorology* 25, 1303–1321. <https://doi.org/10.1175/JHM-D-23-0001.1>
- Allen, G.H., David, C.H., Andreadis, K.M., Hossain, F., Famiglietti, J.S., 2018. Global Estimates of River Flow Wave Travel Times and Implications for Low-Latency Satellite Data. *Geophysical Research Letters* 45, 7551–7560. <https://doi.org/10.1029/2018GL077914>
- 1240 Alsdorf, D.E., Rodríguez, E., Lettenmaier, D.P., 2007. Measuring surface water from space. *Reviews of Geophysics* 45, 2006RG000197. <https://doi.org/10.1029/2006RG000197>
- Althoff, D., Rodrigues, L.N., 2021. Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *Journal of Hydrology* 600, 126674. <https://doi.org/10.1016/j.jhydrol.2021.126674>



- 1245 Andreadis, K.M., Coss, S.P., Durand, M., Gleason, C.J., Simmons, T.T., Tebaldi, N., Bjerklie, D.M., Brinkerhoff, C., Dudley, R.W., Gejadze, I., Larnier, K., Malaterre, P., Oubanas, H., Allen, G.H., Bates, P.D., David, C.H., Domeneghetti, A., Elmi, O., Marc, L.F., Prata de Moraes Frasson, R., Friedmann, E., Garambois, P., Gehring, J., Getirana, A., Hughes, M., Lee, J., Matte, P., Minear, J.T., Monnier, J., Muhebwa, A., Tourian, M.J., Pavelsky, T.M., Riggs, R.M., Rodríguez, E., Sikder, M.S., Smith, L.C., Stuurman, C., Taneja, J., Tarpanelli, A., Wang, J., Williams, B.A., Yadav, B., 2025. A First Look at River Discharge Estimation From SWOT Satellite Observations. *Geophysical Research Letters* 52, e2024GL114185. <https://doi.org/10.1029/2024GL114185>
- 1250 Arora, Boer, 1999. A variable velocity flow routing algorithm for GCMs. *J. Geophys. Res.* 104, 30965–30979. <https://doi.org/10.1029/1999JD900905>
- Bai, H., Ming, Z., Zhong, Y., Zhong, M., Kong, D., Ji, B., 2022. Evaluation of evapotranspiration for exorheic basins in China using an improved estimate of terrestrial water storage change. *Journal of Hydrology* 610, 127885. <https://doi.org/10.1016/j.jhydrol.2022.127885>
- 1255 Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of Hydrology* 212–213, 198–212. [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4)
- Beaudoing, H., 2020. GLDAS VIC Land Surface Model L4 monthly 1.0 x 1.0 degree V2.1. <https://doi.org/10.5067/VWTH7S6218SG>
- 1260 Beaudoing, H., Rodell, M., NASA/GSFC/HSL, 2020. GLDAS Noah Land Surface Model L4 monthly 1.0 x 1.0 degree, Version 2.1. <https://doi.org/10.5067/LWTYSMP3VM5Z>
- Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A.I.J.M., Miralles, D.G., 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Sci Data* 10, 724. <https://doi.org/10.1038/s41597-023-02549-6>
- 1265 Biancamaria, S., 2019. Total water storage variability from GRACE mission and hydrological models for a 50,000 km² temperate watershed_ the Garonne River basin (France). *Journal of Hydrology*.
- Birkett, C.M., 1998. Contribution of the TOPEX NASA Radar Altimeter to the global monitoring of large rivers and wetlands. *Water Resources Research* 34, 1223–1239. <https://doi.org/10.1029/98WR00124>
- 1270 Birkinshaw, S.J., Moore, P., Kilsby, C. g., O'Donnell, G.M., Hardy, A. j., Berry, P. a. M., 2014. Daily discharge estimation at ungauged river sites using remote sensing. *Hydrological Processes* 28, 1043–1054. <https://doi.org/10.1002/hyp.9647>
- Bjerklie, D.M., Lawrence Dingman, S., Vorosmarty, C.J., Bolster, C.H., Congalton, R.G., 2003. Evaluating the potential for measuring river discharge from space. *Journal of Hydrology* 278, 17–38. [https://doi.org/10.1016/S0022-1694\(03\)00129-X](https://doi.org/10.1016/S0022-1694(03)00129-X)
- 1275 Blazquez, A., Meyssignac, B., Lemoine, J., Berthier, E., Ribes, A., Cazenave, A., 2018. Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets. *Geophysical Journal International* 215, 415–430. <https://doi.org/10.1093/gji/ggy293>
- Bloeschl, G., Bierkens, M.F.P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H.G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Szeles, B., Viglione, A., Aksoy, H., Allen, S.T., Amin, A., Andreassian, V., Arheimer, B., Aryal, S.K., Baker, V., Bardsley, E., Barendrecht, M.H., Bartosova, A., Batelaan, O., Berghuijs, W.R., Beven, K., Blume, T., Bogaard, T., de Amorim, P.B., Boettcher, M.E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Yangbo, Chen, Yuanfang, Chiffard, P., Claps, P., Clark, M.P., Collins, A.L., Croke, B., Dathe, A., David, P.C., de Barros, F.P.J., de Rooij, G., Di Baldassarre, G., Driscoll, J.M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W.H., Feicabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Bevacqua, A.G., Gonzalez-Dugo, M.P., Grimaldi, S., Gupta, A.B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlavacikova, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T.H., Inam, A., Innocente, C., Istanbuloglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnova, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M.L.R., Lindquist, E., Link, T., Liu, J., Loucks, D.P., Luce, C., Mahe, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B.D., Montanari, A., Mueller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C.,



- 1295 Nesterova, N., Nurtaev, B., Odongo, V.O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M.J., Post, D., Prieto, C., Ramos, M.-H., Renner, M., Reynolds, J.E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D.E., Rosso, R., Roy, T., Sa, J.H.M., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R.C., Skaugen, T., Smith, H., Spiessl, S.M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R.J., van der Ploeg, M., Van Loon, A.F., van
- 1300 Meerveld, I., van Nooijen, R., van Oel, P.R., Vidal, J.-P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A.J., Ward, P., Westerberg, I.K., White, C., Wood, E.F., Woods, R., Xu, Z., Yilmaz, K.K., Zhang, Y., 2019. Twenty-three unsolved problems in hydrology (UPH) - a community perspective. *Hydrol. Sci. J.* 64, 1141–1158. <https://doi.org/10.1080/02626667.2019.1620507>
- 1305 Bonfils, C.J.W., Santer, B.D., Fyfe, J.C., Marvel, K., Phillips, T.J., Zimmerman, S.R.H., 2020. Human influence on joint changes in temperature, rainfall and continental aridity. *Nat. Clim. Chang.* 10, 726–731. <https://doi.org/10.1038/s41558-020-0821-1>
- Brinkerhoff, C.B., Gleason, C.J., Ostendorf, D.W., 2019. Reconciling at-a-Station and at-Many-Stations Hydraulic Geometry Through River-Wide Geomorphology. *Geophysical Research Letters* 46, 9637–9647. <https://doi.org/10.1029/2019GL084529>
- 1310 C3S, 2018. ERA5 hourly data on single levels from 1940 to present. <https://doi.org/10.24381/CDS.ADBB2D47>
- Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate* 28, 6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- 1315 Carter, R.W., Anderson, I.E., 1963. Accuracy of Current Meter Measurements. *J. Hydr. Div.* 89, 105–115. <https://doi.org/10.1061/JYCEAJ.0000898>
- Cerbelaud, A., David, C.H., Pavelsky, T., Biancamaria, S., Garambois, P. -A., Kittel, C., Papa, F., Bates, P., Tourian, M.J., Durand, M., Frasson, R., Oubanas, H., Simard, M., Schumann, G., Allen, G., Tarpanelli, A., Wongchuig, S., Wade, J., Tom, M., Malaterre, P. -O., Gleason, C.J., Crétaux, J. -F., Nielsen, K., Elmi, O., Schwatke, C., Scherer, D., Gal, L., Paris, A., Dasgupta, A., Munier, S., Feng, D., Bauer-Gottwein, P., Hossain, F., Yamazaki, D., Kitambo, B.,
- 1320 Mischel, S., Bonnema, M., Verma, K., Andreadis, K., Picot, N., Benveniste, J., 2025. Satellite Requirements to Capture Water Propagation in Earth's Rivers. *Reviews of Geophysics* 63, e2024RG000871. <https://doi.org/10.1029/2024RG000871>
- Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H., Koren, V., Duan, Q.Y., Ek, M., Betts, A., 1996. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.* 101, 7251–7268. <https://doi.org/10.1029/95JD02165>
- 1325 Clarke, R.T., 1999. Uncertainty in the estimation of mean annual flood due to rating-curve indefiniton. *Journal of Hydrology* 222, 185–190. [https://doi.org/10.1016/S0022-1694\(99\)00097-9](https://doi.org/10.1016/S0022-1694(99)00097-9)
- Coe, M.T., Birkett, C.M., 2004. Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad basin. *Water Resources Research* 40, 2003WR002543. <https://doi.org/10.1029/2003WR002543>
- 1330 Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., Arheimer, B., 2020. Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal* 65, 699–711. <https://doi.org/10.1080/02626667.2019.1659509>
- 1335 Crowley, J.W., Mitrovica, J.X., Bailey, R.C., Tamisiea, M.E., Davis, J.L., 2006. Land water storage within the Congo Basin inferred from GRACE satellite gravity data. *Geophysical Research Letters* 33, 2006GL027070. <https://doi.org/10.1029/2006GL027070>
- Dorigo, W., Dietrich, S., Aires, F., Brocca, L., Carter, S., Cretaux, J.-F., Dunkerley, D., Enomoto, H., Forsberg, R., Güntner, A., Hegglin, M.I., Hollmann, R., Hurst, D.F., Johannessen, J.A., Kummerow, C., Lee, T., Luoju, K., Looser, U., Miralles, D.G., Pellet, V., Recknagel, T., Vargas, C.R., Schneider, U., Schoeneich, P., Schröder, M., Tapper, N., Vuglinsky, V., Wagner, W., Yu, L., Zappa, L., Zemp, M., Aich, V., 2021. Closing the Water Cycle from Observations across Scales: Where Do We Stand? *Bulletin of the American Meteorological Society* 102, E1897–E1935. <https://doi.org/10.1175/BAMS-D-19-0316.1>
- 1340



- Dunne, T., Black, R.D., 1970. An Experimental Investigation of Runoff Production in Permeable Soils. *Water Resources Research* 6, 478–490. <https://doi.org/10.1029/WR006i002p00478>
- 1345 Durand, M., Gleason, C.J., Pavelsky, T.M., Prata de Moraes Frasson, R., Turmon, M., David, C.H., Altenau, E.H., Tebaldi, N., Larnier, K., Monnier, J., Malaterre, P.O., Oubanas, H., Allen, G.H., Astifan, B., Brinkerhoff, C., Bates, P.D., Bjerklie, D., Coss, S., Dudley, R., Fenoglio, L., Garambois, P., Getirana, A., Lin, P., Margulis, S.A., Matte, P., Minear, J.T., Muhebwa, A., Pan, M., Peters, D., Riggs, R., Sikder, M.S., Simmons, T., Stuurman, C., Taneja, J., Tarpanelli, A., Schulze, K., Tourian, M.J., Wang, J., 2023. A Framework for Estimating Global River Discharge From the Surface
- 1350 Water and Ocean Topography Satellite Mission. *Water Resources Research* 59, e2021WR031614. <https://doi.org/10.1029/2021WR031614>
- Earth Resources Observation and Science (EROS) Center, 2010. Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). <https://doi.org/10.5066/F7J38R2N>
- 1355 Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.* 108, 2002JD003296. <https://doi.org/10.1029/2002JD003296>
- Elmi, O., Tourian, M.J., Bardossy, A., Sneeuw, N., 2021. Spaceborne River Discharge From a Nonparametric Stochastic Quantile Mapping Function. *Water Resour. Res.* 57, e2021WR030277. <https://doi.org/10.1029/2021WR030277>
- 1360 Elnashar, A., Wang, L., Wu, B., Zhu, W., Zeng, H., 2021. Synthesis of global actual evapotranspiration from 1982 to 2019. *Earth System Science Data* 13, 447–480. <https://doi.org/10.5194/essd-13-447-2021>
- Eom, J., Seo, K.-W., Ryu, D., 2017. Estimation of Amazon River discharge based on EOF analysis of GRACE gravity data. *Remote Sensing of Environment* 191, 55–66. <https://doi.org/10.1016/j.rse.2017.01.011>
- ESA Climate Change Initiative River Discharge Precursor, 2025. River Discharge from altimeters, river width, multispectral images and data combination Algorithm Theoretical Basis Document (ATBD).
- 1365 Fekete, B.M., Robarts, R.D., Kumagai, M., Nachtnebel, H.-P., Odada, E., Zhulidov, A.V., 2015. Time for in situ renaissance. *Science* 349, 685–686. <https://doi.org/10.1126/science.aac7358>
- Filippucci, P., Brocca, L., Bonafoni, S., Saltalippi, C., Wagner, W., Tarpanelli, A., 2022. Sentinel-2 high-resolution data for river discharge monitoring. *Remote Sensing of Environment* 281, 113255. <https://doi.org/10.1016/j.rse.2022.113255>
- 1370 Frappart, F., Blarel, F., Fayad, I., Bergé-Nguyen, M., Crétaux, J.-F., Shu, S., Schregengberger, J., Baghdadi, N., 2021. Evaluation of the Performances of Radar and Lidar Altimetry Missions for Water Level Retrievals in Mountainous Environment: The Case of the Swiss Lakes. *Remote Sensing* 13, 2196. <https://doi.org/10.3390/rs13112196>
- Frappart, F., Papa, F., Marieu, V., Malbeteau, Y., Jordy, F., Calmant, S., Durand, F., Bala, S., 2015. Preliminary Assessment of SARAL/AltiKa Observations over the Ganges-Brahmaputra and Irrawaddy Rivers. *Marine Geodesy* 38, 568–580. <https://doi.org/10.1080/01490419.2014.990591>
- 1375 Gal, L., Zakharova, E., Biancamaria, S., Paris, A., Kitambo, B., Lefebvre, J., Boussaroque, M., 2024. ESA River Discharge Climate Change Initiative (RD_cci): Altimetry-based River Discharge product, v1.0. <https://doi.org/10.5285/44C930E1388F40728884FBD7E28C109>
- Galelli, S., Turner, S.W.D., Pokhrel, Y., Yi Ng, J., Castelletti, A., Bierkens, M.F.P., Pianosi, F., Biemans, H., 2025. Advancing the Representation of Human Actions in Large-Scale Hydrological Models: Challenges and Future Research
- 1380 Directions. *Water Resources Research* 61, e2024WR039486. <https://doi.org/10.1029/2024WR039486>
- Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H.S., Yu, Q., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. *Ecohydrology* 11, e1974. <https://doi.org/10.1002/eco.1974>
- 1385 Gao, H., Tang, Q., Ferguson, C.R., Wood, E.F., Lettenmaier, D.P., 2010. Estimating the water budget of major US river basins via remote sensing. *International Journal of Remote Sensing* 31, 3955–3978. <https://doi.org/10.1080/01431161.2010.483488>
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and
- 1390 Applications, Version 2 (MERRA-2). *J. Climate* 30, 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>



- Ghiggi, G., Humphrey, V., Seneviratne, S.I., Gudmundsson, L., 2021. G-RUN ENSEMBLE: A Multi-Forcing Observation-Based Global Runoff Reanalysis. *Water Resources Research* 57, e2020WR028787. <https://doi.org/10.1029/2020WR028787>
- 1395 Ghiggi, G., Humphrey, V., Seneviratne, S.I., Gudmundsson, L., 2019. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data* 11, 1655–1674. <https://doi.org/10.5194/essd-11-1655-2019>
- Gleason, C.J., Smith, L.C., Lee, J., 2014. Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: Sensitivity to river form and optimization parameters. *Water Resources Research* 50, 9604–9619. <https://doi.org/10.1002/2014WR016109>
- 1400 Global Modeling And Assimilation Office, Pawson, S., 2015. MERRA-2 tavgM_2d_Ind_Nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4. <https://doi.org/10.5067/8S35XF81C28F>
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- 1405 Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility. *Geosci. Model Dev.* 11, 3481–3496. <https://doi.org/10.5194/gmd-11-3481-2018>
- Hanasaki, N., Kanae, S., Oki, T., 2006. A reservoir operation scheme for global river routing models. *Journal of Hydrology* 327, 22–41. <https://doi.org/10.1016/j.jhydrol.2005.11.011>
- 1410 Hannah, D.M., Demuth, S., van Lanen, H.A.J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L.M., 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes* 25, 1191–1200. <https://doi.org/10.1002/hyp.7794>
- Horton, R.E., 1945. EROSIONAL DEVELOPMENT OF STREAMS AND THEIR DRAINAGE BASINS; HYDROPHYSICAL APPROACH TO QUANTITATIVE MORPHOLOGY. *Geol Soc America Bull* 56, 275. [https://doi.org/10.1130/0016-7606\(1945\)56\[275:EDOSAT\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2)
- 1415 Huang, Q., Long, D., Han, Z., Han, P., 2022. High-resolution satellite images combined with hydrological modeling derive river discharge for headwaters: A step toward discharge estimation in ungauged basins. *Remote Sensing of Environment* 277, 113030. <https://doi.org/10.1016/j.rse.2022.113030>
- 1420 Huffman, G.J., Adler, R.F., Behrangi, A., Bolvin, D.T., Nelkin, E.J., Gu, G., Ehsani, M.R., 2023. The New Version 3.2 Global Precipitation Climatology Project (GPCP) Monthly and Daily Precipitation Products. *Journal of Climate* 36, 7635–7655. <https://doi.org/10.1175/JCLI-D-23-0123.1>
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K.-L., Joyce, R.J., Kidd, C., Nelkin, E.J., Sorooshian, S., Stocker, E.F., Tan, J., Wolff, D.B., Xie, P., 2020. Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG), in: Levizzani, V., Kidd, C., Kirschbaum, D.B., Kummerow, C.D., Nakamura, K., Turk, F.J. (Eds.), *Satellite Precipitation Measurement, Advances in Global Change Research*. Springer International Publishing, Cham, pp. 343–353. https://doi.org/10.1007/978-3-030-24568-9_19
- 1425 Hussain, D., Kao, H.-C., Khan, A.A., Lan, W.-H., Imani, M., Lee, C.-M., Kuo, C.-Y., 2020. Spatial and Temporal Variations of Terrestrial Water Storage in Upper Indus Basin Using GRACE and Altimetry Data. *IEEE Access* 8, 65327–65339. <https://doi.org/10.1109/ACCESS.2020.2984794>
- 1430 Ishikawa, Y., Yamazaki, D., Yang, Y., 2025. Evaluation of a Width-Based Satellite Discharge Algorithm for Detecting Longitudinal Flow Changes in a Human-Regulated Continental River Basin. *Geophysical Research Letters* 52, e2024GL114191. <https://doi.org/10.1029/2024GL114191>
- Jiao, J., Pan, Y., Cui, X., Mohasseb, H.A., Ding, H., 2025. Evaluation of runoff variability in transboundary basins over High Mountain Asia: Multi-dataset merging based on satellite gravimetry constraint. *Remote Sensing of Environment* 316, 114493. <https://doi.org/10.1016/j.rse.2024.114493>
- 1435 Khan, Adams, 2019. Indus River Basin - 1st Edition [WWW Document]. URL <https://shop.elsevier.com/books/indus-river-basin/khan/978-0-12-812782-7> (accessed 1.18.24).
- Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society* 98, 69–78. <https://doi.org/10.1175/BAMS-D-14-00283.1>
- 1440



- Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.* 23, 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan* 93, 5–48. <https://doi.org/10.2151/jmsj.2015-001>
- 1445 Koblinsky, C.J., Clarke, R.T., Brenner, A.C., Frey, H., 1993. Measurement of river level variations with satellite altimetry. *Water Resources Research* 29, 1839–1848. <https://doi.org/10.1029/93WR00542>
- Konrad, H., Roca, R., Niedorf, A., Finkensieper, S., Schröder, M., Cloché, S., Panegrossi, G., Sanò, P., Kidd, C., Jucá Oliveira, R.A., Fennig, K., Sikorski, T., Lemoine, M., Hollmann, R., 2025. GIRAFE v1: a global climate data record for precipitation accompanied by a daily sampling uncertainty. <https://doi.org/10.5194/essd-2024-568>
- 1450 Koren, V., Schaake, J., Mitchell, K., Duan, Q. -Y., Chen, F., Baker, J.M., 1999. A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.* 104, 19569–19585. <https://doi.org/10.1029/1999JD900232>
- Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T.Y., Tokuhito, T., Sato, Y., Matsushita, Y., Onogi, K., 2024. The JRA-3Q Reanalysis. *Journal of the Meteorological Society of Japan* 102, 49–109. <https://doi.org/10.2151/jmsj.2024-004>
- 1455 Koster, R.D., Suarez, M.J., Ducharme, A., Stieglitz, M., Kumar, P., 2000. A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.* 105, 24809–24822. <https://doi.org/10.1029/2000JD900327>
- 1460 Kouraev, A.V., Zakharova, E.A., Samain, O., Mognard, N.M., Cazenave, A., 2004. Ob’ river discharge from TOPEX/Poseidon satellite altimetry (1992–2002). *Remote Sensing of Environment* 93, 238–245. <https://doi.org/10.1016/j.rse.2004.07.007>
- Koutsoyiannis, D., 2020. Revisiting the global hydrological cycle: is it intensifying? *Hydrol. Earth Syst. Sci.* 24, 3899–3932. <https://doi.org/10.5194/hess-24-3899-2020>
- 1465 Landerer, F.W., Dickey, J.O., Güntner, A., 2010. Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *Journal of Geophysical Research: Atmospheres* 115. <https://doi.org/10.1029/2010JD014584>
- Landerer, F.W., Flechtner, F.M., Save, H., Webb, F.H., Bandikova, T., Bertiger, W.I., Bettadpur, S.V., Byun, S.H., Dahle, C., Dobsław, H., Fahnestock, E., Harvey, N., Kang, Z., Kruizinga, G.L.H., Loomis, B.D., McCullough, C., Murböck, M., Nagel, P., Paik, M., Pie, N., Poole, S., Strekalov, D., Tamisiea, M.E., Wang, F., Watkins, M.M., Wen, H., Wiese, D.N., Yuan, D., 2020. Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance. *Geophysical Research Letters* 47, e2020GL088306. <https://doi.org/10.1029/2020GL088306>
- 1470 Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research* 48, 2011WR011453. <https://doi.org/10.1029/2011WR011453>
- 1475 Lehmann, F., Vishwakarma, B.D., Bamber, J., 2022. How well are we able to close the water budget at the global scale? *Hydrology and Earth System Sciences* 26, 35–54. <https://doi.org/10.5194/hess-26-35-2022>
- Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world’s large river systems. *Hydrological Processes* 27, 2171–2186. <https://doi.org/10.1002/hyp.9740>
- 1480 Leuning, R., Zhang, Y.Q., Rajaud, A., Cleugh, H., Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research* 44, 2007WR006562. <https://doi.org/10.1029/2007WR006562>
- Li, B., 2020. GLDAS Catchment Land Surface Model L4 monthly 1.0 x 1.0 degree V2.1. <https://doi.org/10.5067/FOUXNLXFAZNY>
- 1485 Li, B., Ryu, Y., Jiang, C., Dechant, B., Liu, J., Yan, Y., Li, X., 2023. BESSv2.0: A satellite-based and coupled-process model for quantifying long-term global land–atmosphere fluxes. *Remote Sensing of Environment* 295, 113696. <https://doi.org/10.1016/j.rse.2023.113696>
- Li, L., Zou, X., Liu, B., 2026. Investigating long-term terrestrial water storage change in the Volga River basin based on the extended triple collocation weighted fusion model. *Science of Remote Sensing* 13, 100422. <https://doi.org/10.1016/j.srs.2026.100422>
- 1490



- Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* 99, 14415–14428. <https://doi.org/10.1029/94JD00483>
- Lin, H., Cheng, X., Liu, J., Shi, Q., Li, T., Zheng, L., Hou, X., Du, J., 2024. Estimating river discharge across scales with a novel regional gauging method driven by Sentinel satellite data. *Remote Sensing of Environment* 311, 114266. <https://doi.org/10.1016/j.rse.2024.114266>
- 1495 Liu, S., Lu, P., Liu, D., Jin, P., Wang, W., 2009. Pinpointing the sources and measuring the lengths of the principal rivers of the world. *International Journal of Digital Earth* 2, 80–87. <https://doi.org/10.1080/17538940902746082>
- Loomis, B.D., Luthcke, S.B., Sabaka, T.J., 2019. Regularization and error characterization of GRACE mascons. *J Geod* 93, 1381–1398. <https://doi.org/10.1007/s00190-019-01252-y>
- 1500 Lorenz, C., Kunstmann, H., Devaraju, B., Tourian, M., Sneeuw, N., Riegger, J., 2014. Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances. *Journal of Hydrometeorology* 15. <https://doi.org/10.1175/JHM-D-13-0157.1>
- Luijendijk, E., Gleeson, T., Moosdorf, N., 2020. Fresh groundwater discharge insignificant for the world's oceans but important for coastal ecosystems. *Nat Commun* 11, 1260. <https://doi.org/10.1038/s41467-020-15064-8>
- 1505 Luo, Z., Li, H., Zhang, S., Wang, Lunche, Wang, S., Wang, Lizhe, 2023. A Novel Two-Step Method for Enforcing Water Budget Closure and an Intercomparison of Budget Closure Correction Methods Based on Satellite Hydrological Products. *Water Resources Research* 59, e2022WR032176. <https://doi.org/10.1029/2022WR032176>
- Luo, Z., Zhang, S., Ding, X., Wang, L., Shao, Q., Huang, H., Chen, X., Li, H., 2026. A framework for assessing the impact trends of neglecting water surface evaporation and substituting streamflow on water budget closure. *Journal of Hydrology* 664, 134391. <https://doi.org/10.1016/j.jhydrol.2025.134391>
- 1510 Ma, N., Zhang, Y., Szilagyi, J., 2024. Water-balance-based evapotranspiration for 56 large river basins: A benchmarking dataset for global terrestrial evapotranspiration modeling. *Journal of Hydrology* 630, 130607. <https://doi.org/10.1016/j.jhydrol.2024.130607>
- Maggioni, V., Meyers, P.C., Robinson, M.D., 2016. A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *Journal of Hydrometeorology* 17, 1101–1117. <https://doi.org/10.1175/JHM-D-15-0190.1>
- 1515 Mahar, G.A., Zaigham, N.A., Azam, M., Hussain, S.I., 2022. Morphological evolution of Indus shelf region under the influence of hydrodynamic conditions in 20th century. *Journal of Sea Research* 184, 102216. <https://doi.org/10.1016/j.seares.2022.102216>
- 1520 Marvel, K., Cook, B.I., Bonfils, C.J.W., Durack, P.J., Smerdon, J.E., Williams, A.P., 2019. Twentieth-century hydroclimate changes consistent with human influence. *Nature* 569, 59–65. <https://doi.org/10.1038/s41586-019-1149-8>
- Masunaga, H., Schröder, M., Furuzawa, F.A., Kummerow, C., Rustemeier, E., Schneider, U., 2019. Inter-product biases in global precipitation extremes. *Environ. Res. Lett.* 14, 125016. <https://doi.org/10.1088/1748-9326/ab5da9>
- 1525 McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., Wood, E.F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. *Geosci. Model Dev.* 9, 283–305. <https://doi.org/10.5194/gmd-9-283-2016>
- Miao, Y., Wang, A., 2020. Evaluation of Routed-Runoff from Land Surface Models and Reanalyses Using Observed Streamflow in Chinese River Basins. *J Meteorol Res* 34, 73–87. <https://doi.org/10.1007/s13351-020-9120-z>
- 1530 Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E.C., Haghdoust, S., 2025. GLEAM4: global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. *Sci Data* 12, 416. <https://doi.org/10.1038/s41597-025-04610-y>
- Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M.F., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrol. Earth Syst. Sci.* 20, 823–842. <https://doi.org/10.5194/hess-20-823-2016>
- 1535 Monteith, Lennox, 1965. Evaporation and environment. *Symposia of the Society for Experimental Biology*.
- Morrow, E., Mitrovica, J.X., Fotopoulos, G., 2011. Water Storage, Net Precipitation, and Evapotranspiration in the Mackenzie River Basin from October 2002 to September 2009 Inferred from GRACE Satellite Gravity Data. *Journal of Hydrometeorology* 12, 467–473. <https://doi.org/10.1175/2010JHM1278.1>



- 1540 Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T.A., Popat, E., Portmann, F.T.,
Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., Döll, P., 2021. The global water resources
and use model WaterGAP v2.2d: model description and evaluation. *Geoscientific Model Development* 14, 1037–
1079. <https://doi.org/10.5194/gmd-14-1037-2021>
- 1545 Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., Pan, M., 2014. Combining data sets of satellite-
retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction
model. *JGR Atmospheres* 119. <https://doi.org/10.1002/2014JD021953>
- Munier, S., Decharme, B., 2022. River network and hydro-geomorphological parameters at 1/12° resolution for global
hydrological and climate studies. *Earth Syst. Sci. Data* 14, 2239–2258. <https://doi.org/10.5194/essd-14-2239-2022>
- 1550 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M.,
Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo,
C., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System
Science Data* 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- NASA/JPL, 2021. CSR TELLUS GRACE Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly
Release 6.0 version 04 in netCDF/ASCII/GeoTIFF Formats. <https://doi.org/10.5067/TELND-3AC64>
- 1555 Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal
of Hydrology* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nie, N., Zhang, W., Zhang, Z., Guo, H., Ishwaran, N., 2016. Reconstructed Terrestrial Water Storage Change (Δ TWS) from
1948 to 2012 over the Amazon Basin with the Latest GRACE and GLDAS Products. *Water Resour Manage* 30, 279–
294. <https://doi.org/10.1007/s11269-015-1161-1>
- 1560 Nohara, D., Kitoh, A., Hosaka, M., Oki, T., 2006. Impact of Climate Change on River Discharge Projected by Multimodel
Ensemble. *Journal of Hydrometeorology* 7, 1076–1089. <https://doi.org/10.1175/JHM531.1>
- Oki, T., Sud, Y.C., 1998. Design of Total Runoff Integrating Pathways (TRIP)—A Global River Channel Network. *Earth
Interact.* 2, 1–37. [https://doi.org/10.1175/1087-3562\(1998\)002<0001:DOTRIP>2.3.CO;2](https://doi.org/10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2)
- 1565 Oudin, L., Andréassian, V., Lerat, J., Michel, C., 2008. Has land cover a significant impact on mean annual streamflow? An
international assessment using 1508 catchments. *Journal of Hydrology* 357, 303–316.
<https://doi.org/10.1016/j.jhydrol.2008.05.021>
- Padrón, R.S., Gudmundsson, L., Decharme, B., Lawrence, D.M., Mao, J., Peano, D., Krinner, G., Kim, H.,
Seneviratne, S.I., 2020. Observed changes in dry-season water availability attributed to human-induced climate
change. *Nat. Geosci.* 13, 477–481. <https://doi.org/10.1038/s41561-020-0594-1>
- 1570 Pan, M., Fisher, C.K., Chaney, N.W., Zhan, W., Crow, W.T., Aires, F., Entekhabi, D., Wood, E.F., 2015. Triple collocation:
Beyond three estimates and separation of structural/non-structural errors. *Remote Sensing of Environment* 171, 299–
310. <https://doi.org/10.1016/j.rse.2015.10.028>
- Pan, M., Sahoo, A.K., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2012. Multisource Estimation of Long-Term
Terrestrial Water Budget for Major Global River Basins. *J. Climate* 25, 3191–3206. [https://doi.org/10.1175/JCLI-D-
11-00300.1](https://doi.org/10.1175/JCLI-D-
11-00300.1)
- 1575 Pan, M., Wood, E.F., 2006. Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble
Kalman Filter. *Journal of Hydrometeorology* 7, 534–547. <https://doi.org/10.1175/JHM495.1>
- Papa, F., Durand, F., Rossow, W.B., Rahman, A., Bala, S.K., 2010. Satellite altimeter-derived monthly discharge of the Ganga-
Brahmaputra River and its seasonal to interannual variations from 1993 to 2008. *J. Geophys. Res.* 115,
2009JC006075. <https://doi.org/10.1029/2009JC006075>
- 1580 Paris, A., 2015. Utilisation conjointe de données d'altimétrie satellitaire et de modélisation pour le calcul des débits distribués
dans le bassin amazonien (phd). Université de Toulouse, Université Toulouse III - Paul Sabatier.
- Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P.-A., Collischonn, W., Bonnet,
M.-P., Seyler, F., 2016. Stage-discharge rating curves based on satellite altimetry and modeled discharge in the
Amazon basin. *Water Resources Research* 52, 3787–3814. <https://doi.org/10.1002/2014WR016618>
- 1585 Pellet, victor, Aires, F., Yamazaki, D., 2021. Coherent Satellite Monitoring of the Water Cycle Over the Amazon. Part 2:
Total Water Storage Change and River Discharge Estimation. *Water Resources Research* 57, e2020WR028648.
<https://doi.org/10.1029/2020WR028648>



- 1590 Pellet, V., Aires, F., Yamazaki, D., Papa, F., 2021. Coherent Satellite Monitoring of the Water Cycle Over the Amazon. Part 1: Methodology and Initial Evaluation. *Water Resources Research* 57, e2020WR028647. <https://doi.org/10.1029/2020WR028647>
- Penman, H.L., Keen, B.A., 1948. Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 193, 120–145. <https://doi.org/10.1098/rspa.1948.0037>
- 1595 Petrie, J., Diplas, P., Gutierrez, M., Nam, S., 2013. Data evaluation for acoustic Doppler current profiler measurements obtained at fixed locations in a natural river. *Water Resources Research* 49, 1003–1016. <https://doi.org/10.1002/wrcr.20112>
- Räsänen, T.A., Someth, P., Lauri, H., Koponen, J., Sarkkula, J., Kumm, M., 2017. Observed river discharge changes due to hydropower operations in the Upper Mekong Basin. *Journal of Hydrology* 545, 28–41. <https://doi.org/10.1016/j.jhydrol.2016.12.023>
- 1600 Reichle, R.H., De Lannoy, G.J.M., Forman, B.A., Draper, C.S., Liu, Q., 2014. Connecting Satellite Observations with Water Cycle Variables Through Land Data Assimilation: Examples Using the NASA GEOS-5 LDAS. *Surv Geophys* 35, 577–606. <https://doi.org/10.1007/s10712-013-9220-8>
- Roca, R., Alexander, L.V., Potter, G., Bador, M., Jucá, R., Contractor, S., Bosilovich, M.G., Cloché, S., 2019. FROGS: a daily $1^\circ \times 1^\circ$ gridded precipitation database of rain gauge, satellite and reanalysis products (preprint). *Atmosphere – Meteorology*. <https://doi.org/10.5194/essd-2019-51>
- 1605 Rodell, M., Beaudoin, H.K., L'Ecuyer, T.S., Olson, W.S., Famiglietti, J.S., Houser, P.R., Adler, R., Bosilovich, M.G., Clayson, C.A., Chambers, D., Clark, E., Fetzer, E.J., Gao, X., Gu, G., Hilburn, K., Huffman, G.J., Lettenmaier, D.P., Liu, W.T., Robertson, F.R., Schlosser, C.A., Sheffield, J., Wood, E.F., 2015. The Observed State of the Water Cycle in the Early Twenty-First Century. *Journal of Climate* 28, 8289–8318. <https://doi.org/10.1175/JCLI-D-14-00555.1>
- 1610 Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoin, H.K., Landerer, F.W., Lo, M.-H., 2018. Emerging trends in global freshwater availability. *Nature* 557, 650–+. <https://doi.org/10.1038/s41586-018-0123-1>
- Rodell, M., Houser, P.R., Jambor, U., Gottschalk, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society* 85, 381–394. <https://doi.org/10.1175/BAMS-85-3-381>
- 1615 Rodell, M., Velicogna, I., Famiglietti, J., 2009. Satellite Based Estimates of Groundwater Depletion in India. *Nature* 460, 999–1002. <https://doi.org/10.1038/nature08238>
- Saemian, P., Elmi, O., Stroud, M., Riggs, R., Kitambo, B.M., Papa, F., Allen, G.H., Tourian, M.J., 2025. Satellite Altimetry-based Extension of global-scale in situ river discharge Measurements (SAEM). *Earth Syst. Sci. Data* 17, 2063–2085. <https://doi.org/10.5194/essd-17-2063-2025>
- 1620 Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, Jie, Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, Jun, Hou, Y.-T., Chuang, H.-Y., Juang, H.-M., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R., Rutledge, G., Goldberg, M., 2010. NCEP Climate Forecast System Reanalysis (CFSR) 6-hourly Products, January 1979 to December 2010. <https://doi.org/10.5065/D69K487J>
- 1625 Sahoo, A.K., Pan, M., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sensing of Environment* 115, 1850–1865. <https://doi.org/10.1016/j.rse.2011.03.009>
- 1630 Salwey, S., Coxon, G., Pianosi, F., Singer, M.B., Hutton, C., 2023. National-Scale Detection of Reservoir Impacts Through Hydrological Signatures. *Water Resources Research* 59, e2022WR033893. <https://doi.org/10.1029/2022WR033893>
- Save, H., Bettadpur, S., Tapley, B.D., 2016. High-resolution CSR GRACE RL05 mascons. *JGR Solid Earth* 121, 7547–7569. <https://doi.org/10.1002/2016JB013007>
- Schneider, D.P., Deser, C., Fasullo, J., Trenberth, K.E., 2013. Climate Data Guide Spurs Discovery and Understanding. *EoS Transactions* 94, 121–122. <https://doi.org/10.1002/2013EO130001>
- 1635 Schneider, U., Hänsel, S., Finger, P., Rustemeier, E., Ziese, M., 2022. GPCC Full Data Monthly Version 2022 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data: Globally Gridded Monthly Totals. https://doi.org/10.5676/DWD_GPCC/FD_M_V2022_025



- Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational
1640 Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB
Approach. *J American Water Resour Assoc* 49, 577–591. <https://doi.org/10.1111/jawr.12057>
- Shiklomanov, A.I., Lammers, R.B., 2009. Record Russian river discharge in 2007 and the limits of analysis. *Environ. Res.
Lett.* 4, 045015. <https://doi.org/10.1088/1748-9326/4/4/045015>
- Shiklomanov, A.I., Yakovleva, T.I., Lammers, R.B., Karasev, I.Ph., Vörösmarty, Charles.J., Linder, E., 2006. Cold region
1645 river discharge uncertainty—estimates from large Russian rivers. *Journal of Hydrology* 326, 231–256.
<https://doi.org/10.1016/j.jhydrol.2005.10.037>
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K., 2018. A Review of Global Precipitation Data Sets: Data
Sources, Estimation, and Intercomparisons. *Reviews of Geophysics* 56, 79–107.
<https://doi.org/10.1002/2017RG000574>
- 1650 Sun, W., Ishidaira, H., Bastola, S., 2012. Calibration of hydrological models in ungauged basins based on satellite radar
altimetry observations of river water level. *Hydrological Processes* 26, 3524–3537. <https://doi.org/10.1002/hyp.8429>
- Syed, T.H., Famiglietti, J.S., Chen, J., Rodell, M., Seneviratne, S.I., Viterbo, P., Wilson, C.R., 2005. Total basin discharge for
the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance. *Geophysical Research
Letters* 32, 2005GL024851. <https://doi.org/10.1029/2005GL024851>
- 1655 Talsma, C.J., Good, S.P., Jimenez, C., Martens, B., Fisher, J.B., Miralles, D.G., McCabe, M.F., Purdy, A.J., 2018. Partitioning
of evapotranspiration in remote sensing-based models. *Agricultural and Forest Meteorology* 260–261, 131–143.
<https://doi.org/10.1016/j.agrformet.2018.05.010>
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: Mission overview
and early results. *Geophysical Research Letters* 31, 2004GL019920. <https://doi.org/10.1029/2004GL019920>
- 1660 Tarpanelli, A., Amarnath, G., Brocca, L., Massari, C., Moramarco, T., 2017. Discharge estimation and forecasting by MODIS
and altimetry data in Niger-Benue River. *Remote Sensing of Environment* 195, 96–106.
<https://doi.org/10.1016/j.rse.2017.04.015>
- Tarpanelli, A., Brocca, L., Barbetta, S., Faruolo, M., Lacava, T., Moramarco, T., 2015. Coupling MODIS and Radar Altimetry
Data for Discharge Estimation in Poorly Gauged River Basins. *IEEE J. Sel. Top. Appl. Earth Observations Remote
Sensing* 8, 141–148. <https://doi.org/10.1109/JSTARS.2014.2320582>
- 1665 Tourian, M.J., Sneeuw, N., Bardossy, A., 2013. A quantile function approach to discharge estimation from satellite altimetry
(ENVISAT). *Water Resour. Res.* 49, 4174–4186. <https://doi.org/10.1002/wrcr.20348>
- Vishwakarma, B.D., Devaraju, B., Sneeuw, N., 2018. What Is the Spatial Resolution of grace Satellite Products for Hydrology?
Remote Sensing 10, 852. <https://doi.org/10.3390/rs10060852>
- 1670 Voroessmarty et al, 2001. Global Water Data: A Newly Endangered Species. *EOS Trans., American Geophysical Union*, 82,
54–58.
- Wang, S., Fu, B., Piao, S., Lü, Y., Ciais, P., Feng, X., Wang, Y., 2016. Reduced sediment transport in the Yellow River due
to anthropogenic changes. *Nature Geosci* 9, 38–41. <https://doi.org/10.1038/ngeo2602>
- 1675 Wang, S., Song, S., Zhang, H., Yu, L., Jiao, C., Li, C., Wu, X., Zhao, W., Best, J., Roberts, P., Fu, B., 2025. Anthropogenic
impacts on the Yellow River Basin. *Nat Rev Earth Environ* 6, 656–671. <https://doi.org/10.1038/s43017-025-00718-2>
- Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., Liu, C., 2007. A Gauge-Based Analysis of Daily
Precipitation over East Asia. *Journal of Hydrometeorology* 8, 607–626. <https://doi.org/10.1175/JHM583.1>
- 1680 Yoon, Y., Durand, M., Merry, C.J., Rodríguez, E., 2013. Improving Temporal Coverage of the SWOT Mission Using
Spatiotemporal Kriging. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6,
1719–1729. <https://doi.org/10.1109/JSTARS.2013.2257697>
- Yu, Y., Wang, H., Shi, X., Ran, X., Cui, T., Qiao, S., Liu, Y., 2013. New discharge regime of the Huanghe (Yellow River):
Causes and implications. *Continental Shelf Research* 69, 62–72. <https://doi.org/10.1016/j.csr.2013.09.013>
- 1685 Zakharova, E., Nielsen, K., Kamenev, G., Kouraev, A., 2020. River discharge estimation from radar altimetry: Assessment of
satellite performance, river scales and methods. *Journal of Hydrology* 583, 124561.
<https://doi.org/10.1016/j.jhydrol.2020.124561>
- Zakharova, E.A., Kouraev, A.V., Cazenave, A., Seyler, F., 2006. Amazon River discharge estimated from TOPEX/Poseidon
altimetry. *Comptes Rendus. Géoscience* 338, 188–196. <https://doi.org/10.1016/j.crte.2005.10.003>



- 1690 Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment* 222, 165–182. <https://doi.org/10.1016/j.rse.2018.12.031>
- Zhang, Y., Leuning, R., Hutley, L.B., Beringer, J., McHugh, I., Walker, J.P., 2010. Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution. *Water Resources Research* 46, 2009WR008716. <https://doi.org/10.1029/2009WR008716>
- 1695 Zhang, Y., Pan, M., Sheffield, J., Siemann, A.L., Fisher, C.K., Liang, M., Beck, H.E., Wanders, N., MacCracken, R.F., Houser, P.R., Zhou, T., Lettenmaier, D.P., Pinker, R.T., Bytheway, J., Kummerow, C.D., Wood, E.F., 2018. A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. *Hydrol. Earth Syst. Sci.* 22, 241–263. <https://doi.org/10.5194/hess-22-241-2018>
- 1700 Zhu, Y., Liu, S., Yi, Y., Xie, F., Grünwald, R., Miao, W., Wu, K., Qi, M., Gao, Y., Singh, D., 2021. Overview of terrestrial water storage changes over the Indus River Basin based on GRACE/GRACE-FO solutions. *Science of The Total Environment* 799, 149366. <https://doi.org/10.1016/j.scitotenv.2021.149366>