Journal of Hydrology: Regional Studies 50 (2023) 101545

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Evaluation of runoff estimation from GRACE coupled with different meteorological gridded products over the Upper Blue Nile Basin

Khaled Alghafli [a,b,*], Awad M. Ali [c], Xiaogang Shi [d], William Sloan [a], Ali A.A. Obeid [e], Mohammad Shamsudduha [f]

[a] James Watt School of Engineering, University of Glasgow, Glasgow, UK
[b] National Water and Energy Center, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates
[c] Water Research Center, University of Khartoum, P.O. Box 321, Khartoum, Sudan
[d] School of Social and Environmental Sustainability, University of Glasgow, Dumfries, United Kingdom
[e] Institute of Hydraulic Engineering and Water Resources Management, TU Wien, Vienna, Austria
[f] Institute for Risk and Disaster Reduction, University College London, London, UK

ARTICLE INFO

Keywords:
Terrestrial water storage
Precipitation
Evapotranspiration
Water balance
GRACE
Upper Blue Nile basin

ABSTRACT

Study Region: The Upper Blue Nile (UBN) basin, Ethiopia.
Study Focus: In efforts to accurately close the water balance equation for the UBN basin using remote sensing products, river runoff is calculated as a residual from the water balance equation by incorporating Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS) and remote sensing products for precipitation (P) and evapotranspiration (ET). The calculated river runoff is then compared to the gauge records located at the basin's outlet. The best performing combination among the various combinations is chosen by aggregating rankings attributed to both error and linear fit metrics. The errors associated with each satellite product were assessed by forcing the In-Situ runoff to estimate the P, ET, and TWS. This methodology helps in addressing the uncertainty linked with each hydrological component.
New Hydrological Insights for the Region: The best P, ET, and TWS combination performance products to estimate runoff are SM2RAIN-CCI, GLEAM, and GRACE Spherical Harmonic products, respectively. The statistical results for the six metrics are $R^2 = 0.7$, slope = 1.6, y-intercept = - 0.5 cm, RMSE = 3 cm, MAE = 2.8 cm, and PBIAS = 36%. The uncertainty from each hydrological component was quantified and showed that improving the accuracy of P and ET estimation is a crucial step to successfully close the water balance.

1. Introduction

The accurate quantification of water cycle components over a large-scale basin is an essential step towards improving the understanding of water balance and responses of river basin to different hydrologic extremes, such as droughts and floods (Sheffield et al., 2009). The water cycle components are primarily controlled by precipitation (P), evapotranspiration (ET), runoff (R), and terrestrial water storage (TWS) changes. A dense In-Situ network of hydro-meteorological datasets is required to simulate the hydrologic cycle

* Corresponding author at: James Watt School of Engineering, University of Glasgow, Glasgow, UK.
E-mail address: 2508863a@student.gla.ac.uk (K. Alghafli).

https://doi.org/10.1016/j.ejrh.2023.101545
Received 17 May 2023; Received in revised form 4 October 2023; Accepted 5 October 2023
Available online 10 October 2023

parameters and their heterogeneity (Archfield et al., 2015; Crochemore et al., 2020; Eagleson, 1986). However, ungauged basins set a new challenge for water resource scientists to monitor water cycle components until the launch of the Gravity Recovery and Climate Experiment (GRACE) satellites in 2002 (Rodell and Famiglietti, 2001), which provided a key parameter for the terrestrial water storage estimation. GRACE mission was to provide a monthly estimate of TWS changes on Earth from gravimetry. The TWS components are surface water storage, groundwater storage, snow water storage, and soil moisture storage (Tapley et al., 2004).

The recent advances in satellites have enabled researchers to monitor water balance components regionally. A considerable effort at the international level has helped provide global gridded datasets for the water cycle components. These gridded datasets are based on In-Situ data, satellites, and reanalysis models. Prior to incorporating these products in the water budget calculation, it is recommended to assess their performance over the studied basin (Zambrano-Bigiarini et al., 2017; Zandler et al., 2019). Various studies have attempted to close the water balance equation with remote sensing data coupled with land surface models, but it was concluded that it is challenging to be achieved (Abolafia-Rosenzweig et al., 2021; Chen et al., 2020; Ferreira et al., 2013; Gao et al., 2010; Moreira et al., 2019; Pan et al., 2012; Penatti et al., 2015; Sahoo et al., 2011; Sheffield et al., 2009; Sneeuw et al., 2014; Syed et al., 2009; Syed et al., 2005; Tan et al., 2022; Wang et al., 2014a; Wang et al., 2014b; Yao et al., 2014) due to the uncertainties associated with satellite data and the limitation of land surface models to account for anthropogenic activities, coupled with the coarse resolution of GRACE. Additionally, this imbalance, referring to the disparity between water budget inflow and outflow, is inherited from signal processing uncertainties and different algorithm assumptions (Aires, 2014; Long et al., 2014; Pan et al., 2012; Sheffield et al., 2009).

Numerous studies have investigated the source of imbalance and uncertainties associated with each water budget component by applying different methods. Sheffield et al. (2009) estimated the uncertainty for each component of the water balance and removed the systematic bias for the precipitation, as it showed a high level of overestimation. Syed et al. (2009) combined the water balance equation with the atmospheric moisture budget to estimate the river runoff. The estimated river discharge was then utilized as an inflow for the global ocean mass balance to address the residual. Munier et al. (2014) employed the Simple Weight integration and the Post Processing Filter to enforce the water budget closure. Additionally, they applied a linear regression model (closure correction model) to reduce the residual from the water budget components.

Several studies have assigned a weight proportionally to the error variance of each product, and then applied a Constrained Kalman Filter data assimilation approach to close the water budget (Aires, 2014; Pan et al., 2012; Pan and Wood, 2006; Sahoo et al., 2011; Wong et al., 2021; Zhang et al., 2018). Abolafia-Rosenzweig et al. (2021) utilized three closure constraint techniques: the Constrained Kalman Filter, the Proportional Redistribution and the Multiple Collocation at the water budget level, rather than correcting each component of the water budget. Luo et al. (2021) provided an approach to assess the accuracy of the water budget closure for sixty different combinations from satellites by applying the omission error and employing the first order reliability method. Fok et al. (2023) applied the geographically weighted averaging approach to each gridded runoff value to account for the geographical heterogeneity of

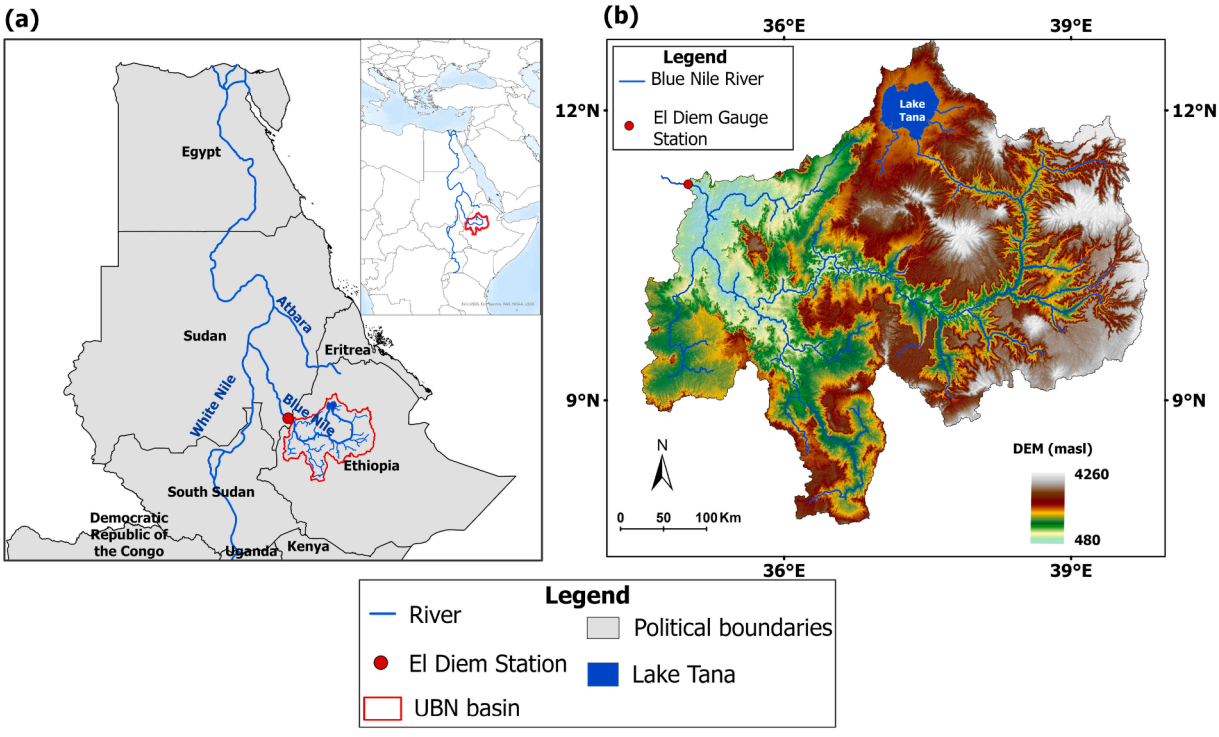

Fig. 1. (a) shows the geographic location of the UBN basin in the Nile basin (b) DEM obtained from the Shuttle Radar Topography Mission (SRTM) in meter above sea level (Farr et al., 2007). The blue line represents the Blue Nile River and the outlet of the basin is located at the El Diem as seen in red line.

each grid. Other research compared the water budget components estimated from satellite products with in-situ gauges for validation and applied various statistical approaches to evaluate the trends of the water balance components (Azarderakhsh et al., 2011; Chen et al., 2020; Ferreira et al., 2013; Lehmann et al., 2022; Liu et al., 2018; Mohanasundaram et al., 2021; Moreira et al., 2019; Oliveira et al., 2014; Syed et al., 2005).

The water budget analysis requires a detailed analysis to quantify the water fluxes entering, leaving, and stored in a basin. Several studies concluded that the largest errors in the imbalance are mainly attributed to evapotranspiration due to the heterogeneity of land cover (Abera et al., 2017; Soltani et al., 2020), coupled with the numerous parameters influencing its variations. These parameters include vegetation types, soil moisture, air temperature, relative humidity, wind speed, solar radiation and surface albedo (Martens et al., 2017; Mu et al., 2007; Zhang et al., 2019). Other researchers claimed that the deviation in water balance closure is caused by precipitation (Sheffield et al., 2009; Wang et al., 2014b; Wong et al., 2021). This discrepancy arises from the intense spatial variation of precipitation influenced by a complex topography, the uneven distribution of gauges, and the climatic condition (wind, air temperature, radiation, humidity), which affect the ability of precipitation devices to take accurate measurements (Schreiner-McGraw and Ajami, 2020; Sivapalan et al., 2003; Wang et al., 2014a; Wrzesien et al., 2019). Others argue that the imbalance is inherited from the terrestrial water storage signals captured by GRACE (Fok et al., 2023; Syed et al., 2005; Wang et al., 2015) due to the coarse resolution of GRACE and the signal leakage during postprocessing phase (Watkins et al., 2015; Wiese et al., 2016).

Quantifying the hydrological variables of the water budget on a regional scale will enhance our understanding of the water cycle. The UBN basin, also known as the Abbay basin, contributes more than 60% of the streamflow to the main River Nile annually (Conway, 2005; Senay et al., 2014). It is located in the Ethiopian highlands and has a drainage area of 175,000 km^2. The basin's outlet is near El Diem at the border with Sudan (Fig. 1). Hundreds of millions of people living downstream in Sudan and Egypt rely on the water coming from the UBN. In 2011, the construction of the hydroelectric dam known as the Grand Ethiopian Renaissance Dam in Ethiopia began upstream on the Blue Nile River and has been a concern for downstream countries (Sudan and Egypt), where it could reduce water availability during filling and will change the flow regime. Once the dam filling is completed, the dam will be the largest dam in Africa with a storage capacity of 74 billion cubic meters to generate a power capacity of 5150 Megawatts (Ezega News, 2019). Thus, accurate quantification of the hydrological cycle fluxes will help manage the water resources in an effective and sustainable manner.

In the UBN basin, the estimation of individual components of the hydrological cycle, such as P (Abebe et al., 2020; Abera et al., 2016; Ayehu et al., 2018; Bayissa et al., 2017; Dinku et al., 2011; Lakew et al., 2017; Lakew et al., 2020; Worqlul et al., 2018), ET (Allam et al., 2016; Weerasinghe et al., 2020), and TWS (Seyoum, 2018; Shamsudduha et al., 2017), using remote sensing are examined, whereas other research papers have studied the UBN basin using the water balance model (Abebe et al., 2022; Conway, 1997; Johnson Peggy and Curtis, 1994; Kim and Kaluarachchi, 2009; Mishra and Hata, 2006; Steenhuis et al., 2009; Tekleab et al., 2011).

To date, there are no satellite products capable of estimating runoff except by using altimetry and multispectral (optical and radar) sensors, which provide key parameters as an input for the hydraulic equation or to apply empirical approach. The hydraulic parameters obtained from remote sensing for the discharge estimations are water height (Koblinsky et al., 1993; Kouraev et al., 2004; Zakharova et al., 2020; Zakharova et al., 2006) and river width (Smith et al., 1996), but could have uncertainties and limitations. The recent launch of surface water ocean topography (Durand et al., 2010; Fu et al., 2012) will provide a promising product for river water surface elevation, width, and slope for rivers $\geq$ 100 m width which could be converted to river discharge. Other than that, researchers have estimated the runoff from satellites using the water budget closure approach.

Over the entire UBN basin, a few studies have focused on closing the water budget using satellites and reanalysis datasets to quantify runoff (Jung et al., 2017; Koukoula et al., 2020; Siam et al., 2013). Abera et al. (2017) applied the JGrass—NewAGE system, incorporating remote sensing solutions to estimate P, discharge, ET, and TWS, and then compared the estimated ET with the satellite-based ET. Lazin et al. (2020) simulated the water budget components using coupled routing and excess storage soil-vegetation-atmosphere-snow over the UBN and compared the simulated components with In-Situ data and global models. Jung et al. (2017) employed land surface models (LSMs) coupled with Hydrological Modeling and Analysis Platform (HyMAP) river routing schemes to generate an estimate for the hydrological components over the UBN basin. In situ gauges for river discharge were incorporated in the calculation and revealed that the Catchment LSM version Fortuna 2.5 (CLSMF2.5) outperformed Noah version 3.3 in the estimation of the water budget components. The above studies coupled modeling techniques with remote sensing products for the estimation of the water budget components. There is no research that purely assesses the closure of the water budget using remote sensing nor estimates the runoff at the UBN using satellite products. The availability of various remote sensing products for the estimation of the water budget components poses a challenge for researchers and users to decide what product one can rely on. Therefore, an assessment for these remote sensing products is required.

Although the evaluation of water budget closure has been widely applied to validate its credibility (Mohanasundaram et al., 2021; Moreira et al., 2019), there is no single correct approach for the evaluation of remote sensing products. Incorporating In-Situ gauges in the water budget closure will help verify the analysis and address the uncertainty associated with the calculation. Thus, inferring river runoff as a residual from the water budget closure can provide better information about the accuracy of satellites and their uncertainties. Compared to previous research, this research will assess river runoff derived from multiple satellites products not only by stating the statistical results, but with summing up the performance rankings for the chosen linear and error metrics, without relying solely on one metric. The Unified metrics (UM) will indicate which runoff combination among all the estimated runoffs from remote sensing products has less discrepancy. The Overall Unified metrics (OUM) are proposed in this study to assess the performance of an individual satellite product depending on the runoff score derived from the UM approach. This research assesses the majority of remote sensing products available within a similar temporal resolution of the available in situ discharge gauge to conclude which remote sensing products are better to be used for water management or research purposes. Since runoff data are available at the outlet of the

UBN basin, the target of this paper is to estimate runoff from the water budget equation and diagnose the estimated runoff with the El Diem gauge records. Moreover, our study showcases a comprehensive approach for uncertainty quantification of water budget components.

In summary, a collection of satellite measurements was acquired to quantify the runoff, which was compared with the runoff at the El Diem gauge station over 2003–2014. The questions this research aims to address are as follows: 1) What is the most suitable P, TWS, and ET combination of satellites and reanalysis products that can accurately estimate the runoff at the outlet of the UBN basin?; 2) How reliable is the applied approach for assessing the accuracy of remote sensing products?; 3) What satellite-derived component of the water budget causes the most discrepancies?. These questions are answered by employing the unified metrics (UM) and the overall unified metrics (OUM) approach for the assessment. Additionally, the El Diem river discharge gauge was forced in the water balance equation to estimate the TWS and subsequently compared with the TWS derived from GRACE. Similarly, the river discharge was forced again to estimate ET and P. These steps will help intuitively in evaluating the errors in the water budget closure and the uncertainties associated with the hydrological data products.

2. Data

2.1. Study Area

The River Nile is a transboundary river flows through 11 African countries- Burundi, Democratic Republic of Congo, Egypt, Ethiopia, Eritrea, Kenya, Rwanda, Sudan, South Sudan, Tanzania and Uganda, including three major rivers contributing to its supply namely: the Blue Nile, the White Nile, and the Atbara River (Fig. 1). The White Nile originates at a height of 1600 m above mean sea level in Northern Burundi, whereas the Blue Nile source comes from the Ethiopian Highlands near Lake Tana, 1800 m above sea level. The Atbara River headwater is also sourced from the Ethiopian highlands and close to the Blue Nile headwater which is north of Lake Tana (Williams, 2019). The UBN basin has a drainage area of 175,000 km^2 with a complex topographic distribution, which also results in variations in the climate. The elevation ranges from 4260 m in the highlands to 480 m in the lowlands near the Ethiopia-Sudan border (Fig. 1b). The topography of the northeastern parts from the basin is characterized by hills and highlands, while the southern and western parts are described as valleys. The climate of the UBN is semiarid to humid and considered relatively wet from June to September, while the dry season spans from October to May. It is estimated that 74% of the annual rainfall occurs from June to September during the wet season (Abtew et al., 2009). The mean annual rainfall of the UBN basin is estimated to be in the range of 1200–1600 mm (Conway, 1997), whereas the annual potential ET is in the range of 1000–1800 mm (Conway, 2000). The spatial distribution of annual rainfall shows high rainfall events on the southern tip and low over the north-eastern area.

2.2. Datasets

2.2.1. Terrestrial water storage

GRACE has twin satellites following each other at a specific distance of ~ 220 km, and a change in the distance is due to a change in Earth's gravity field (Rodell and Famiglietti, 2001). The gravity field maps are derived from the spherical harmonic coefficient up to a maximum degree-1 (generally 60), which provides a spectral representation of the Earth's gravity field (Sun et al., 2016; Swenson and Wahr, 2006). The monthly gravity field was transformed to interpret the changes as a change in the terrestrial water storage after applying multiple corrections which are glacial isostatic adjustment (Peltier et al., 2018), atmospheric and oceanic corrections using the Atmosphere and Ocean De-liasing Level 1-B model (Flechtner et al., 2015). In the latest version of GRACE (RL06), the degree-2 zonal term ΔC_{20} was substituted with the satellite-laser ranging, offering more accurate measurements with less noise (Cheng and Ries, 2019). These steps are applied for the two forms of GRACE solutions, which are the spherical harmonics form (SH) and mass concentration solutions (mascon).

Multiple processing centers provide GRACE data in SH form (e.g., University of Texas Center for Space Research (CSR), German Research Center for Geosciences (GFZ) and Jet Propulsion Laboratory (JPL) (Chambers, 2006; Tapley et al., 2004) and mascon (JPL and CSR). The steps discussed earlier were applied for both SH and mascon solutions.

The decorrelation filtering and 300-km Gaussian smoothing was employed for the SH-based solution to calculate the surface mass change and remove the correlated errors in gravity solutions, whereas JPL mascon applies constraints obtained from geophysical models (Scanlon et al., 2016). Thus, the major difference between the two products is in the postprocessing technique. GRACE TWS is subject to leakage during postprocessing and the application of filters for the SH solutions, and additional leakage could occur where coastal mass variation signals leak into the land and result in large uncertainties (Watkins et al., 2015). As for the mascon solution, the Coastline Resolution Improvement filter is applied to minimize the leakage between the land and ocean boundaries (Wiese et al., 2016). The leaked signal is returned and adjusted for the SH-based products by incorporating TWS derived from models. The Global Land Data Assimilation System-NOAH is incorporated to quantitively estimate errors (Rodell et al., 2004b). The scale factor is derived based on the root mean square difference between the unfiltered and filtered TWS. This factor is multiplied with the GRACE-filtered TWS. As for the mascon solutions, the scale factor obtained from the National Centre for Atmospheric Research's Community Land Model 4.0 (NCAR CLM 4.0) (Lawrence et al., 2011) is incorporated with applying the least squares fit to enhance the resolution from 3° to 0.5° and reduce the leakage (Wiese et al., 2016).

The latest dataset release of GRACE RL06 from CSR, GFZ, JPL, and JPL mascon are used. The following four solutions have been used: (1) GRCTellus Land RL06 release of GRACE (Landerer, 2021; Landerer and Swenson, 2012) and (2) GRACE mascon solution obtained from JPL release 06 Version 02 (Watkins et al., 2015; Wiese et al., 2016). It is suggested to use GRACE for areas larger than

100,000 km^2, and the UBN basin complies with GRACE native resolution. In this study, the three SH GRACE products were averaged to reduce the uncertainties associated with data processing. Thus, GRACE mascon and averaged SH GRACE were utilized for the water balance components estimation. The first-order derivative of TWS was computed from GRACE monthly TWS to represent the storage flux (Eq. 1).

$$\frac{ds}{dt} = TWS_{t+1} - TWS_t \quad (1)$$

where ds/dt represents the TWS flux between two consecutive months. Due to battery issues, GRACE has missing data for certain months, and it was linearly interpolated as suggested by the GRACE handbook (Landerer and Cooley, 2021).

2.2.2. Precipitation

Previous research has concluded that the water budget closure faced an imbalance error as a result of the overestimation of remotely sensed precipitation products (Oliveira et al., 2014; Sheffield et al., 2009; Sneeuw et al., 2014; Xie et al., 2019). Seven precipitation products are chosen for assessing the water budget closure. These products are Climate Hazards InfraRed Precipitation with Stations data version 2.0 (CHIRPSv2), which is developed by the US Geological Survey and the Climate Hazards Group of the University of California (Funk et al., 2015); Climatic Research Unit Time-Series version 4.06 (CRU TS4.06) (Harris et al., 2020); the European Centre for Medium-Range Weather Forecasts ERA5 reanalysis dataset (ERA5) (Hersbach et al., 2020); Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis TRMM 3B43 V7 (Huffman et al., 2019; Kummerow et al., 1998); the Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM Final Run (IMERGv6) (Hou et al., 2014); the Climate Forecast System Reanalysis and Reforecast (CFSR) (Saha et al., 2010); and the Soil Moisture to Rain derived from the European Space Agency Climate Change Initiative (SM2RAIN-CCI) (Ciabatta et al., 2018).

These precipitation datasets are either derived from satellites, In-Situ gauges, or combined reanalysis. CHRIPS synthesizes infrared satellite imagery with In-Situ rainfall gauges, and the utilized infrared satellites are derived from global gridded satellite data of the National Oceanic and Atmospheric Administration (NOAA) spanning 1981–2008 (Knapp et al., 2011) and the 2000-present datasets sourced from the Climate Prediction Center of NOAA (Janowiak et al., 2001). The In-Situ rainfall gauges used in generating CHIRPS data are collected from multiple public sources consisting of around 14,000 gauges with monthly records and 200,000 gauges with daily records. The CRU is a global gridded gauge analysis product that relies on 10,000 gauges and is derived by using the angular-distance weighting interpolating method for a dense network of weather station observations (Harris et al., 2020). The ERA5 reanalysis dataset is derived from combining Earth system model estimates with observations through data assimilation. ERA5 is the fifth-generation reanalysis product replacing the previous version ERA-Interim. ERA5 dataset includes significant atmospheric improvements (e.g., convection, radiation, orographic drag, wind, pressure gradient), with several schemes and parametrizations applied during assimilation. It has a temporal resolution of one hour and ranges from 1950 to the present with a spatial resolution of ~ 25 km (Hersbach et al., 2020).

The TRMM is the first satellite with a mission to measure P in tropical and subtropical areas through infrared sensors and microwave, including a P radar. Several instruments were on the spacecraft, including the precipitation radar, the TRMM Microwave Imager, an optical-infrared radiometer, a lightening detection sensor, and a cloud radiation radiometer (Haddad et al., 1997). It is a joint project between the Japan Aerospace Exploration Agency and the US National Aeronautics and Space Administration and was launched in 1997 (Kummerow et al., 1998). The 3B43 algorithm is applied on a monthly scale to generate the best-estimated precipitation rate by coupling the precipitation data from the Global Precipitation Climatology Centre (GPCC) gauges with the 3-hourly merged infrared estimates derived from multiple satellites. Then, the corrected monthly best-estimated precipitation is combined with the rain gauges data by applying the inverse error variance weighting (Huffman et al., 2007). The TRMM 3B43 V7 product from TRMM was used for the calculation in this study. GPM is a satellite to measure precipitation and snowfall globally every three hours. It is a successor to the TRMM with improvements in the precipitation radar (dual-frequency precipitation radar) to detect the rain drop distribution. TRMM's precipitation radar operates at a single frequency of Ku-band (13.8 GHz) whereas the GPM operates at Ku-band (13.6 GHz) and Ka-band (35.5 GHz) (Hou et al., 2014). The high frequency channels on the precipitation radar and the microwave imager aboard the GPM will improve the detection of light intensity precipitation, falling snow and frozen precipitation (Field and Heymsfield, 2015; Mugnai et al., 2007). It provides rainfall data at a high spaciotemporal resolution using the progressive precipitation retrieval algorithms obtained from several satellite sensors.

The CFSR is a reanalysis dataset with a resolution of ~38 km from 1979 to present. It is a global product coupling the atmosphere-

Table 1
Summary of the precipitation products used in this study.

Product	Spatial resolution	Temporal resolution	Time period
CHRIPSv2	0.05°	daily	1983-present
CRU TS4.06	0.5°	monthly	1901-present
CFSR	0.5°	hourly	1979–2017
ERA5	0.25°	monthly	1981-present
GPM-IMERGv6	0.1°	daily	2000- present
SM2RAIN-CCI	0.25°	daily	1998–2015
TRMM 3B43 V7	0.25°	monthly	1998–2019

ocean-land surface-sea ice system to simulate those components in an hourly time resolution and a horizontal resolution of 0.5° latitude × 0.5° longitude. The atmospheric model takes into account the carbon dioxide concentration and solar variation, with applying bias correction for radiance data obtained from NOAA (Saha et al., 2010). Finally, the SM2RAIN rainfall dataset developed by Brocca et al. (2013) is an approach to estimate rainfall from In-Situ and satellite-based soil moisture. The data are provided in a global scale with a daily time span. The applied method in the SM2RAIN-CCI is based on the "bottom-up" approach that incorporates surface soil moisture readings from multiple sensors for the estimation of precipitation. Active and passive microwave satellite soil moisture products are combined to estimate precipitation (Dorigo et al., 2017) using the SM2RAIN algorithm (Brocca et al., 2014). Additional details related to these seven rainfall products are summarized in Table 1.

2.2.3. Actual evapotranspiration products

Three gridded actual ET (ETa) products were chosen, and each has a different technique to estimate ETa. The selected products are the Global Land Evaporation Amsterdam Model GLEAMv3.6a (Martens et al., 2017; Miralles et al., 2011); MOD16 obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) (Mu et al., 2007; Mu et al., 2011); and the Penman-Monteith-Leuning (PMLv2) model (Gan et al., 2018; Zhang et al., 2019; Zhang et al., 2016). Table 2.

The two versions of GLEAM are different in their forcing data and have different temporal coverage. GLEAM products are based on the Priestley and Taylor equation (Priestley and Taylor, 1972) and driven by the daily satellite dataset (GLEAMv3.6a) and reanalysis (GLEAMv3.6b) dataset (air temperature, humidity, net radiation, and air pressure). Additional to the Priestly and Taylor model, GLEAM incorporates the Gash analytical model to simulate the forest interception loss by considering the vegetation and the precipitation characteristics. GLEAM also employs an evaporative stress formula to depict the response of soil and vegetation throughout the evaporation process (Miralles et al., 2011). Multi-source precipitation products are forced for the calculation of GLEAM ET namely the TRMM 3B42v7 (Huffman et al., 2007) and the Multi-Source Weighted Ensemble Precipitation (MSWEP) (Beck et al., 2017). Furthermore, the other forced in variables in the GLEAM model, such as radiation, air temperature, soil moisture, snow-water equivalent and vegetation optical depth are also retrieved from remote sensing products (Miralles et al., 2011). It also employs the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis- ERA-Interim data for the ETa estimation (Dee et al., 2011).

The calculation of ETa from MOD16 is constructed based on the energy-balance schemes using the Penman-Monteith equation. This ETa calculation is driven by the Global Modeling and Assimilation Office (GMAO) meteorological reanalysis data (Global Modeling and Assimilation Office, 2004) and MODIS data for albedo, leaf area index, land cover, and vegetation index. The improved version of ETa derived from MOD16 considers additional inputs in the applied algorithm which are soil heat flux, vegetation cover fraction, evapotranspiration during both daytime and nighttime. Further advancements have been made in the estimation of aerodynamic resistance, stomatal conductance and boundary layer resistance (Mu et al., 2011).

Both GLEAM and MOD16 rely on the energy-balance method to estimate ET. The PMLv2 model couples the carbon and water fluxes through canopy conductance and integrates the leaf area index obtained from MODIS. It was developed by Leuning et al. (2008) and further improvements were applied by Zhang et al. (2010). The input meteorological reanalysis data in PMLv2 is retrieved from the Global Land Data Assimilation System Version 2.1 (GLDAS 2.1) (Zhang et al., 2019). PMLv2 requires fewer parameters compared to MOD16 and GLEAM products (Gan et al., 2018).

All three ET products (GLEAM, MOD16, and PMLv2) were validated against the eddy flux towers. For example, GLEAM product underwent a comparison with soil moisture records obtained from 2325 devices, and the evaporation fluxes involved a comparison with 91 eddy-covariance towers. The correlation between GLEAM soil moisture and the In-Situ soil moisture records resulted in a coefficient of 0.64. Similarly, a correlation of 0.81 occurred between the eddy covariance-tower and GLEAM ETa. The daily ETa derived from MOD16 was validated against 46 eddy flux towers distributed globally. The mean absolute bias ranged from 0.31 mm day^{-1} to 0.33 mm day^{-1}. In a parallel manner, the assessment of PMLv2 involved a comparison with 95 flux towers, resulting in a PBIAS and RMSE values of − 1.8% and 0.69 mm day^{-1}, respectively (Zhang et al., 2019). Additional details related to the validation of ETa datasets can be found in Gan et al. (2018); Miralles et al. (2011); Mu et al. (2011). To keep in line with the spatial resolutions of the different remote sensing products, an average of the entire basin was taken into the water budget closure, and the temporal resolutions of all the products were converted to a monthly resolution.

2.2.4. Observed discharge

The daily streamflow records of the El Diem station were obtained from the Ministry of Irrigation and Water Resources of Sudan (MoIWR) spanning from 1990 to 2014. Since the MoIWR conducts continuous water level measurements, the discharge at El Diem is calculated using the rating curves method. Conversely, due to the Roseires dam's filling practice occurring downstream between June and October, which could adversely impact the water level measurements at the El Diem gauge station, the water balance approach is applied for estimating the discharge solely during the filling phase. Notably, the reason of choosing the period for the analysis from

Table 2
Summary of ETa products.

Product	Estimation Method	Spatial resolution	Temporal resolution	Time period
GLEAMv3.6a	Priestley-Taylor	0.25°(~28 km)	daily	2003-present
MOD16	Penman-Monteith	0.0043° (~500 m)	8 days	2000-present
PMLv2	Penman-Monteith-Leuning	0.05° (~5.6 km)	8 days	2002-present

2003 to 2014 is to ensure a match in the temporal resolution between the observed discharge and satellite products, particularly with the terrestrial water storage component derived from GRACE.

3. Method

3.1. Evaluation of the water budget closure

To assess the performance of different satellite products, the monthly runoff data were estimated as a residual from the water budget closure using remote sensing products and compared with the runoff gauge at the outlet of the basin (El-Diem) (Eq. 2). Given that the data accessible from the El Diem gauge is from 2003 to 2014, the analysis is carried out within a similar timespan of the gauge records.

$$R = P - ET - \frac{ds}{dt} \tag{2}$$

We examined the best performance combination to estimate the runoff at the outlet of the basin by considering multiple criteria. One statistical measure may not be sufficient to validate the calculated runoff from satellites. Thus, the discrepancy between the calculated and observed monthly runoff required a series of statistical measures to judge which remote sensing combination perform the best. The statistical metrics used in the assessment were three linear fit metrics (the coefficient of determination (R^2; Eq. 3), the y-intercept (Eq. 4) and the slope (Eq. 5)), and three error metrics (the Root Mean Squared Error (RMSE; Eq. 6), the Mean Absolute Error (MAE; Eq. 7) and the Percent Bias (PBIAS; Eq. 8) (Gupta et al., 1999)).

$$R^2 = \frac{\left(n\left(\sum_{i=1}^{n} G_i S_i\right) - \left(\sum_{i=1}^{n} G_i\right)\left(\sum_{i=1}^{n} S_i\right)\right)^2}{\left(n\left(\sum_{i=1}^{n} G_i^2\right) - \left(\left(\sum_{i=1}^{n} G_i\right)^2\right)\right)\left(n\left(\sum_{i=1}^{n} S_i^2\right) - \left(\left(\sum_{i=1}^{n} S_i\right)^2\right)\right)} \tag{3}$$

$$Intercept \quad (mm) = \overline{S} - Slope \times \overline{G} \tag{4}$$

$$Slope = \frac{\left(\sum_{i=1}^{n} G_i S_i\right) - \frac{\left(\sum_{i=1}^{n} G_i\right)\left(\sum_{i=1}^{n} S_i\right)}{n}\Big)}{\left(\sum_{i=1}^{n} G_i^2\right) - \frac{\left(\sum_{i=1}^{n} G_i\right)^2}{n}} \tag{5}$$

$$RMSE \quad (mm) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)^2} \tag{6}$$

$$MAE(mm) = \frac{\sum_{i=1}^{n}|S_i - G_i|}{n} \tag{7}$$

$$PBIAS(\%) = \frac{\sum_{i=1}^{n}(S_i - G_i) \times 100}{\sum_{i=1}^{n}(G_i)} \tag{8}$$

where G_i is the observed monthly runoff from the El Diem gauge station, S_i is the estimated runoff using the water balance equation, n is the number of data records, $\overline{G}$ is the mean of the observed runoff, and $\overline{S}$ is the mean of the calculated runoff.

The coefficient of determination R^2 explains the relationship between the dependant and independent variance, and the value ranges from 0 (no correlation) to 1 (perfect match). A high R^2 value indicates low error variance. However, a high uncertainty has been observed in relying on regression statistics due to the assumption of linear relationship between the observed and the simulated values which are rare to occur (N. Moriasi et al., 2007). The slope indicates whether the estimated runoff overestimates (>1) or underestimates (<1) the observed runoff. Moreover, the y-intercept indicates whether a lag exists (y-intercept‡0) or not (y-intercept = 0). Thus, the error metrics (RMSE, MAE, PBIAS) are applied in the analysis. The RMSE and MAE indicate the discrepancy magnitude between the estimated and real runoff records data. It is suggested that acceptable RMSE and MAE occur when their values are less than half of the standard deviation of the measured data. The PBIAS measures the average tendency of the runoff derived from remote sensing and whether it is smaller or larger than the In-Situ discharge (Gupta et al., 1999).

Once all the statistical metrics are calculated, a method based on summing up the performance rankings was applied herein to decide what combination is the best on estimating the runoff. The Unified Metric (UM) method was applied previously and showed a significance contribution on choosing the best satellites products when compared with In-Situ gauges (Ali et al., 2023; Basheer and

Elagib, 2019; Elagib and Mansell, 2000). The UM will give an indication of the best-performing product of runoff combination derived from the TWS, ETa, and P (Eq. 9).

$$UM_c = \sum_{i=1}^{M} R_{c,m_i} \tag{9}$$

Where UM_c is the unified metric score, R_{c,m_i} is the rank of the combination c based on the performance metric m_i, and M is the number of the performance metrics which are six metrics used in this analysis. Since we used 7 precipitation, 3 ETa, and 2 TWS flux, we evaluated 42 runoff combinations as shown in Fig. 2. To know which datasets are more reliable and accurate to estimate the runoff, a score was given for each combination. The runoff combinations are classified based on the score given from the UM which ranges between 6 and 252 indicating the best and worst performance in all metrics, respectively. The UM is a simple and informative way to evaluate the 42 estimated runoff and to give insight on the most reliable runoff combination among the tested ones. Furthermore, to assess the performance of an individual product (e.g., CHIRPS, GLEAM) compared to all other similar products for P, ETa, and TWS, the Overall Unified Metric was applied:

$$OUM_p = \sum UM_{cep} \tag{10}$$

where OUM is the overall unified metrics for a product p. UM_{cep} is the UM values of all combinations belong to the product. Therefore, the range of OUM depends on the product type (P: 36 – 1512; ETa: 84 – 3528; TWS: 126 – 5292). For example, each product for the precipitation has six possible combinations which each can have a UM value ranging from 6 and 252 resulting in OUM range between 36 and 1512. Additional to the UM and OUM, the boxplot was estimated for the 42 monthly and annual discharge combination. This method will contribute on providing further understanding on the water budget closure.

3.2. Uncertainty analysis

The uncertainty associated with runoff calculation is addressed in this research by computing the 95% confidence bounds of the monthly and the annual records of precipitation, evapotranspiration, runoff, and terrestrial water storage. In other words, the uncertainty of each product will be assessed individually then as a combination. To quantify the level of uncertainty related to each of the three variables (P, ET, and TWS flux), we analyse the variation in model performance caused by changes in product selection for each variable. In simpler terms, the greater the change in performance due to product selection changes, the higher the level of uncertainty associated with the variable. When examining the combinations, we elaborate on the uncertainty arising in the estimation of runoff as a response to alterations in the 42 selected combination of products. Further evaluation was applied by calculating the standard deviation and the standard error. These analyses are informative on giving an indication of how the data are spread and its deviation from the mean.

To conduct a more comprehensive examination of the uncertainty associated with ET, TWS flux and P, we undertake a reverse

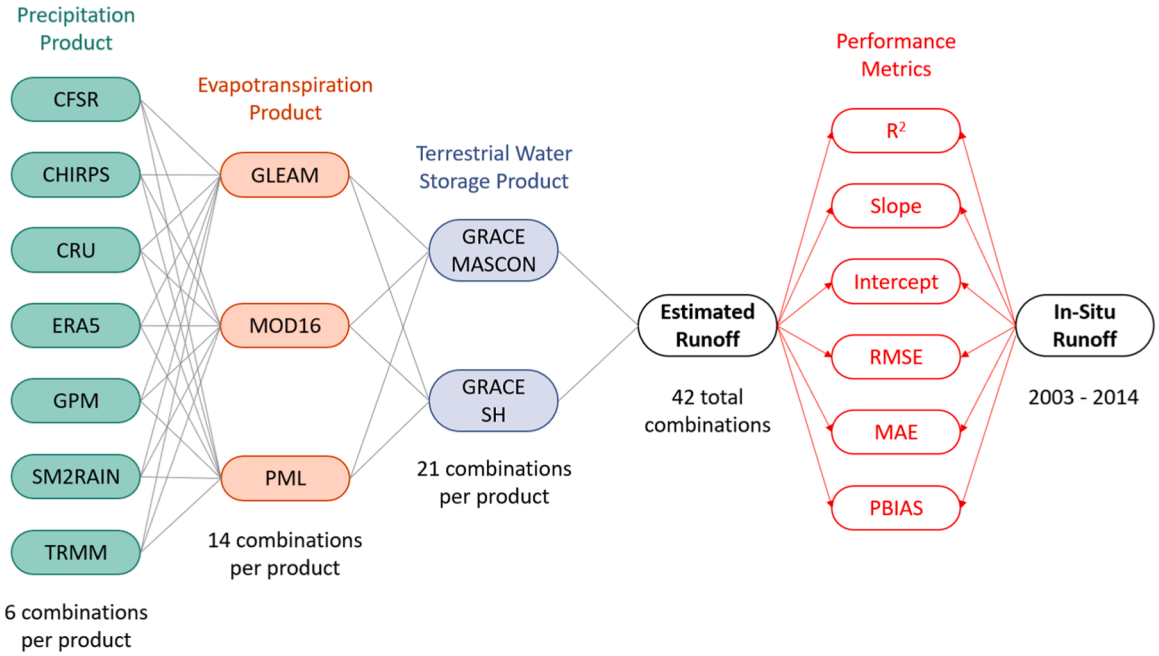

Fig. 2. Selected remote sensing products, possible combinations, and chosen performance metrics.

calculation of ET, TWS flux and P in correspondence with the observed runoff at the El Diem gauge station (R_{obs}). This calculation is executed employing the revised water balance Eqs. (11–13) provided below. By undertaking such estimations, we gain valuable insights into the temporal occurrence of heightened uncertainty and the associated constraints inherent to these variables. The calculated TWS flux was then compared with TWS flux from GRACE SH and GRACE mascon, while the estimated ETa was compared with GLEAM, MOD16 and PML. This comparative analysis was also applied to the P. Subsequently, the 95% confidence bound was estimated to evaluate the ET, ds/dt and P obtained from the reverse water balance calculations using In-Situ runoff with the satellite products. This comparison will help in the selection of the most optimum satellite products which will then be compared with the UM and OUM output.

$$ET = P - R_{obs} - \frac{ds}{dt} \tag{11}$$

$$\frac{ds}{dt} = P - R_{obs} - ET \tag{12}$$

$$P = R_{obs} + ET + \frac{ds}{dt} \tag{13}$$

This rearrangement will give a better insight on the water fluxes of the hydrological cycle to address the uncertainty associated with each product. Once the ETa, ds/dt and P are calculated by forcing the In-Situ river runoff, the uncertainty range for the hydrological components are calculated in a combination and individually. The most dominant source of errors is identified by finding the largest percentage of total non-closure errors among the three hydrological components (ETa, ds/dt and P). This approach is compared with the previous approaches for the purpose of addressing the source of uncertainty and if all the approaches are consistent.

Last, previous research applied in different catchments concluded an existence of imbalance (Sahoo et al., 2011; Sheffield et al., 2009) while closing the water budget, therefore; it is also expected herein this study. The wate budget imbalance is usually defined as a residual (ΔRe) when closing the water budget equation ($\Delta Re = P - ET - R - \frac{ds}{dt}$). Thus, we conducted an extensive analysis to furtherly assess the imbalance and the uncertainties associated with the remote sensing products used in this research. A conclusion will be drawn on stating the reason of the imbalance. Furthermore, this will help on the validation of TWS flux from GRACE and determining which GRACE products performed the best, since TWS flux is difficult to estimate using observed data.

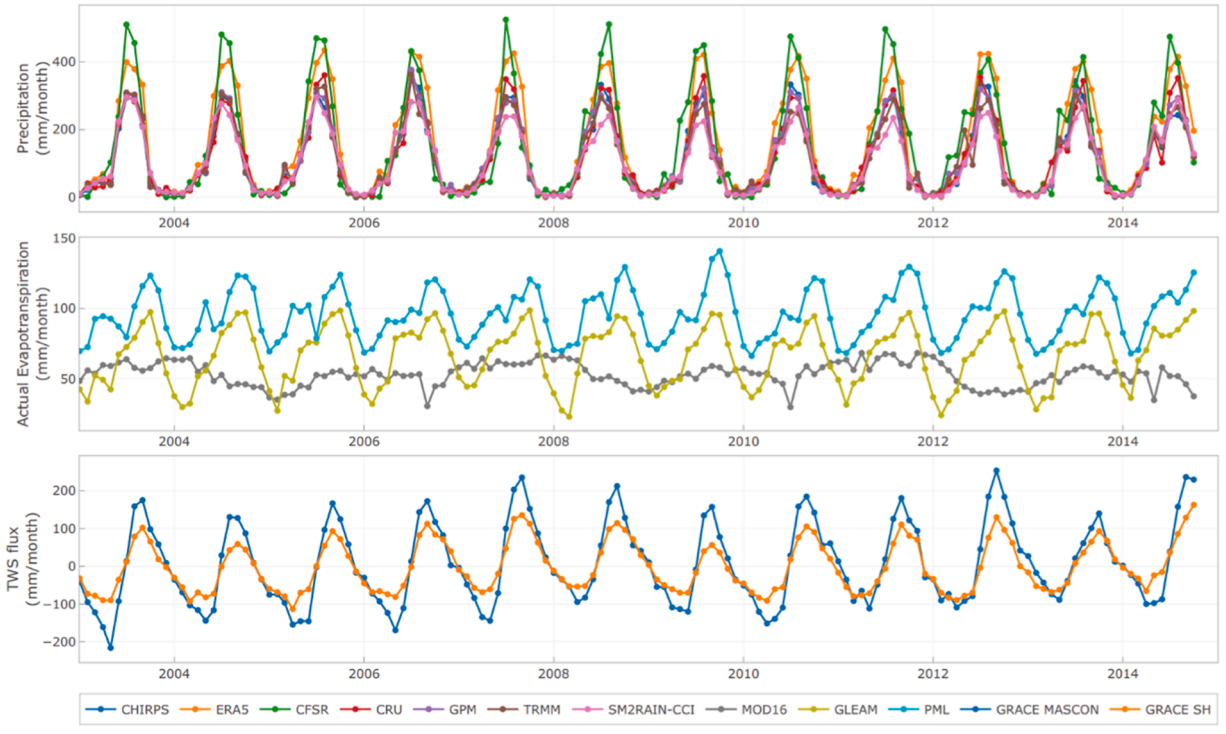

Fig. 3. Monthly precipitation, actual evapotranspiration, and terrestrial water storage data from multiple remote sensing products.

4. Results and Discussion

4.1. Monthly values of remote sensing products

Fig. 3 illustrates the monthly values of the different products. The seven precipitation products (see Table 1) have similar trend variations with different magnitudes as shown in Fig. 3. All precipitation data showed consistency in variation with peak events occurring in July or August (R^2 ranges from 0.6 to 0.9), indicating the rainy season, while January has the lowest monthly mean rainfall. The differences in the precipitation magnitudes between the precipitation products can be attributed to the differences in (1) the revisiting time of the incorporated satellites, (2) the methods employed for calibration (3) the utilized algorithm to derive the precipitation rate, and (4) the specific types of satellites employed to capture precipitation in either liquid or frozen forms (Donat et al., 2014; Huffman et al., 2007). Such differences affect the reliability of the precipitation for discharge simulations and would possibly constrain their use for some hydrological applications (e.g., flood forecasting).

The actual evapotranspiration (ETa) showed consistency between the PML and GLEAM (R^2 =0.8) whereas MOD16 algorithm showed a different trend variation due to its limitation to account for the evaporation from water bodies. Lake Tana and the Blue Nile river contribute significantly to the water evaporation process over the UBN basin; thus this is a drawback for MOD16 to accurately estimate the ETa. This could be the cause of the difference in the trend variation when compared with GLEAM and PML. Furthermore, the incorporated parameter for each ETa dataset has different forcing products, so it is expected to have inconsistent trends and magnitudes among the ETa products. It is important to mention that the three ETa products GLEAM, MOD16 and PMLv2 utilize different meteorological reanalysis data as input in the ETa estimation namely: ECMWF, GMAO, and GLDAS 2.1, respectively (Martens et al., 2017; Mu et al., 2011; Zhang et al., 2019). Based on literature, ETa is difficult to estimate due to the spatial heterogeneity of land cover (Giorgi and Avissar, 1997) and large source of uncertainty inherited from the input parameters (e.g., vegetation characteristics, surface temperature, and radiation) where it is difficult to assess its accuracy (Ferguson et al., 2010). TWS flux obtained from GRACE mascon and the averaged GRACE SH showed strong agreement (R^2 =0.9). Both TWS products showed annual seasonality, however, a wider range was found in case of GRACE mascon (i.e., lower minimum and higher maximum values).

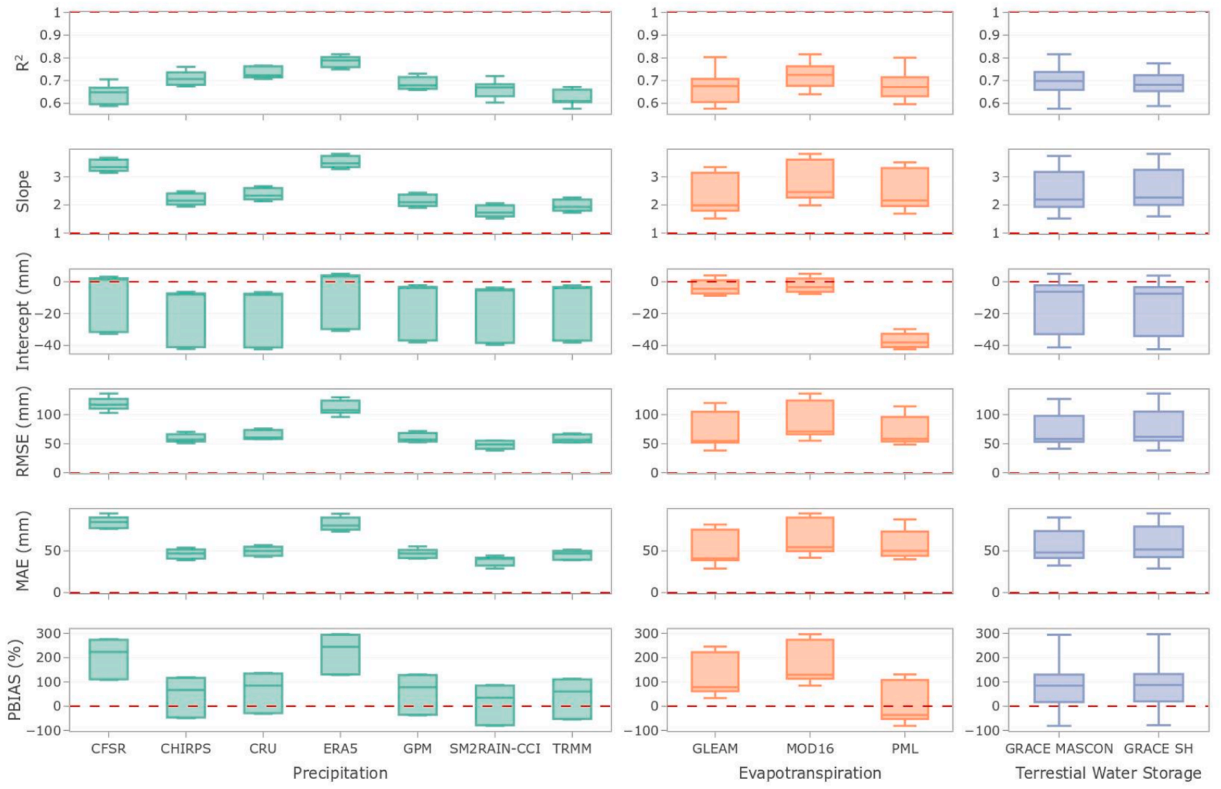

Fig. 4. The performance metrics for all the satellites products based on the runoff estimation when compared with the El-diem gauge station. The dashed red line represents the optimum result. Each boxplot includes the combinations belong to each product (6 for each P, 14 for each ETa, and 21 for each TWS product).

4.2. Product selection uncertainty in runoff estimation

The performance metrics of the estimated runoff from all possible combinations are detailed in Table S.1. It should be noted that various metrics were taken into account to ensure a comprehensive evaluation. Moreover, Fig. 4 displays the performance metrics of the combinations that each product belongs to. This analysis allows the investigation of the uncertainty of runoff estimation caused by the selection of the different input types (P, ETa, or TWS flux).

Many key findings could be extracted from Fig. 4. The runoff was found to be most sensitive to the selection of the precipitation product. Compared to the evapotranspiration and GRACE statistical metrics (Fig. 4), one can see that the precipitation products derived from remote sensing have different magnitudes which signified the variation in the statistical metrics. The runoff was found to be less sensitive to the evapotranspiration and almost not sensitive to the terrestrial water storage.

This significant difference in rainfall is caused due to multiple reasons which are the adopted satellite algorithm, the mismatch in the temporal and spatial resolution between the used remote sensing products, the model parameterization, and the interpolation and calibration process (Donat et al., 2014). In other words, each rainfall product has a different technique on estimating the rainfall rate. Based on the OUM and Fig. 4, the SM2RAIN was the best performance product while the CFSR was the least. The SM2RAIN is based on the novel "bottom up" approach using the soil moisture satellites to accurately quantify the rainfall amount (Brocca et al., 2013) whereas the CFSR relies upon model-derived analysis and surface observation stations (Saha et al., 2010). ERA5 is a reanalysis data, and it has been observed that it underestimated convective precipitation and overestimated precipitation in a complex topographic area (Amjad et al., 2020; Beck et al., 2019). This complies with our research and the runoff combinations that incorporated ERA5 had the largest PBIAS and slope indicating an overestimation (see Fig. 4). This high variation between the satellite products is expected since the region is mountainous where complex terrain could result in high uncertainty in estimating the precipitation (Mei et al., 2014) and this was also observed in East Africa (Cattani et al., 2016).

4.3. Evaluation of estimated runoff combinations

All the possible combinations of the selected 12 remote sensing products were used to estimate the runoff using water budget closer which plotted as a 95% confidence interval as show in Fig. 5. Among the 42 combinations and based on the UM calculations (Table S.2), we found that the best combination is SM2RAIN-CCI + GLEAM + GRACE SH. The statistical results for the six chosen metrics are $R^2 = 0.7$, slope = 1.6, y-intercept = - 0.5 cm, RMSE = 3 cm, MAE = 2.8 cm, and PBIAS = 36%. Table S.2 presents the UM values for all combinations in order of performance, with the best-performing combination at the top and the worst-performing at the bottom.

Results show that all the calculated runoff overestimated the amount of runoff during the rainy seasons and the monthly runoff were either overestimated or underestimated with a PBIAS ranges from − 80–296% (Fig. 5.b and Table S.1), while during the dry season, the runoff was slightly outperforming. Previous research have shown an overestimation of runoff when satellites products are incorporated in the UBN basin (Koukoula et al., 2020). Furthermore, Chen et al. (2020) observed that in the wet season, river runoff obtained from In-Situ gauge underestimated the runoff derived from remote sensing products in the Amazon. Such discrepancies during the wet season could reflect the underestimation of the In-Situ gauge when flood water overflows riverbanks (Chen et al., 2020; Eom et al., 2017). This will require field investigation to investigate if any uncertainties are associated with the In-Situ measurements.

In general, the statistical analysis results of monthly runoff during the dry season (October-May) were more accurate than the results during the wet season (June-September). Not surprisingly, previous research has also shown that closing the water budget in mountainous region (Wang et al., 2014b) and over the UBN (Ali et al., 2023) exhibited the largest water imbalances. Considering if the imbalance in the water budget closure caused by the groundwater outflow from the basin; therefore, it should be examined. There is no major aquifer over the UBN basin and even if so, the slow movement of groundwater relative to surface water is minimum to consider in causing a dramatic change and impacting the GRACE signal (Rodell et al., 2004a; Rodell et al., 2011). Another important finding is that the runoff estimation derived from remote sensing was able to capture the runoff events and was within the 95% confidence bound on the years of 2006, 2008, 2011 and 2012 (Fig. 5.a).

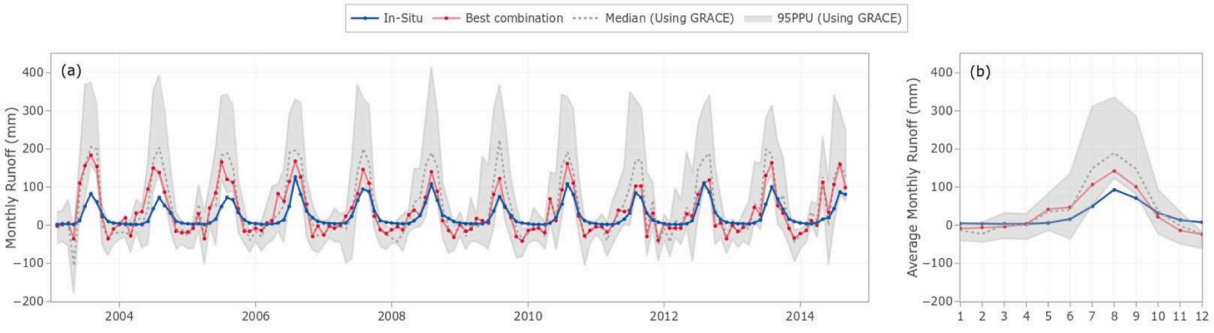

Fig. 5. The shaded colour represents the 95% confidence bound for the R, TWS and ETa (a-b) the 42 runoff combinations overestimated the runoff in the wet season. The best combination is based on SM2RAIN-CCI+GLEAM+GRACE SH.

Furthermore, to assess which individual product of the P, ETa and TWS performed best in the estimation of runoff, the overall unified metrics (OUM) was calculated for each satellite product (Table 3) and revealed that SM2RAIN-CCI, GLEAM, and GRACE mascon are the best individual products compared to the considered ones in this study. The OUM indicates how the individual remote sensing product performs when coupled with other remote sensing products only. This approach revealed that GRACE mascon outperformed GRACE SH. Thus, further analysis is required since both GRACE mascon and GRACE SH have different preprocessing methods as discussed earlier. Forcing the In-Situ runoff into the water balance equation will help furtherly investigating the most optimum solutions from GRACE.

4.4. *In-Situ runoff forced in the water balance equation*

To furtherly examine the imbalance in the water budget closure, the In-Situ runoff data was forced in the water balance equation in three different instances. First, it was employed to estimate the terrestrial water storage and compared it with the two TWS products. Second, it was forced again to estimate the evapotranspiration for the comparison with the ETa products obtained from GLEAM, MOD16, and PML. Last, it was also employed to calculate the precipitation for the comparison with satellite-based P (Fig. 6). All precipitation records retrieved from remote sensing products overestimated the precipitation calculated from the forced river runoff in the water budget equation, as indicated in Fig. 6 (panels a and b). Similar to the estimated runoff from satellites (Fig. 5), it is shown that all the precipitation products exhibited an overestimation during the wet season and was able to be within the 95% confidence bound during the dry season. Based on the 95% confidence bound, the precipitation values obtained from SM2RAIN-CCI is the most optimum products. This finding corresponds with the results from UM and the OUM. Consequently, the three different methods showed similar outcomes. The SM2RAIN-CCI performed very good when compared with the observed rainfall data over Africa (Ciabatta et al., 2018).

It is clearly shown that the estimated TWS in good agreement with GRACE mascon and GRACE SH. This is an indication that GRACE products are reliable and have less uncertainty. In terms of the 95% confidence bound, the GRACE mascon outperformed the SH as demonstrated in Fig. 6 (panels e and f). The TWS derived from GRACE mascon outperformed SH both in the monthly analysis and seasonal analysis. This result contradicts with the Unified Metric and matches with the OUM. Thus, the 95% confidence bound output should be prioritized in this analysis if a user is interested in monitoring the TWS flux during the rainy season. Previous research has demonstrated that GRACE mascon outperformed GRACE SH (Alghafli et al., 2023; Bhanja et al., 2016; Fok et al., 2023; Neves et al., 2020; Scanlon et al., 2016; Wong et al., 2021). This aligns with the OUM method and when rearranging the water balance equation for the TWS flux estimation as displayed in Fig. 6.

The ETa is difficult to measure directly and has large uncertainties at the basin scale. Thus, the second step is to force the In-Situ runoff gauge in the water budget closure to estimate the ETa. The analysis revealed high uncertainties in the ETa derived from satellites products when compared with the ETa derived from the forced In-Situ gauge Fig. 6 (panels c and d). The satellite products showed a lag of 1 month compared to the estimated ETa with underestimated magnitudes. This discrepancy indicates that ETa requires further investigation and analysis. Our findings align with previous studies that assessed the ETa over the Blue Nile basin and concluded that GLEAM and MOD16 tend to underestimate the ETa (McNamara et al., 2021; Trambauer et al., 2014; Weerasinghe et al., 2020). ETa obtained from MOD16 is driven by meteorological reanalysis data from GMAO and MODIS data. To address the ETa uncertainty obtained from MOD16, Heinsch et al. (2006) compared GMAO meteorological data and MODIS datasets for land cover and leaf area index with in In-Situ measurements and revealed an overestimation. Due to the various parameters incorporated in MOD16 and even if all the MOD16 parameters are accurately estimated, the leaf area index inaccuracy could lead to biases and a misrepresentative values for ETa (Mu et al., 2011). In-Situ measurements for the ETa plays a key role in the calibration process. It is important to highlight that in Africa, only six flux towers are available for the ETa estimation, indicating a limited coverage. None of these flux towers are located in the UBN basin (Baldocchi, 2008).

The performance of the different products against the calculations forced by In-Situ runoff is summarized in Table 4. The median R^2 and RMSE values were used to evaluate the different products. The main difference between GRACE mascon and SH is the higher error resulted from the latter product. The median R^2 score for both GRACE products was similar, but there was a difference of 6.82 mm in the median RMSE. This finding aligns with the OUM results and contradicts with the UM findings. The UM objective is to find the most optimum runoff combination rather than investigating the performance of individual products. Thus, the mentioned details are for the purpose of analysing different satellite products. For the ETa products, it aligned with the rank of the products showed in Table 3. GLEAM resulted in the highest R^2 and lowest RMSE medians compared to PML and MOD16. The median R^2 was not high for the three products. This indicates the complexity of estimating ETa, given the many factors impacting its variations. In a similar manner, the

Table 3
The Overall Unified Metric (OUM) score for each remote sensing product based on the runoff estimation when compared with the In-Situ gauge (El Diem).

Rank	P (Product Name)	OUM	Eta (Product Name)	OUM	TWS (Product Name)	OUM
1	SM2RAIN-CCI	510	GLEAM	1502	GRACE mascon	2566
2	GPM	651	PML	1895	GRACE SH	2852
3	TRMM	663	MOD16	2021		
4	CHIRPS	677				
5	CRU	776				
6	ERA5	1013				
7	CFSR	1128				

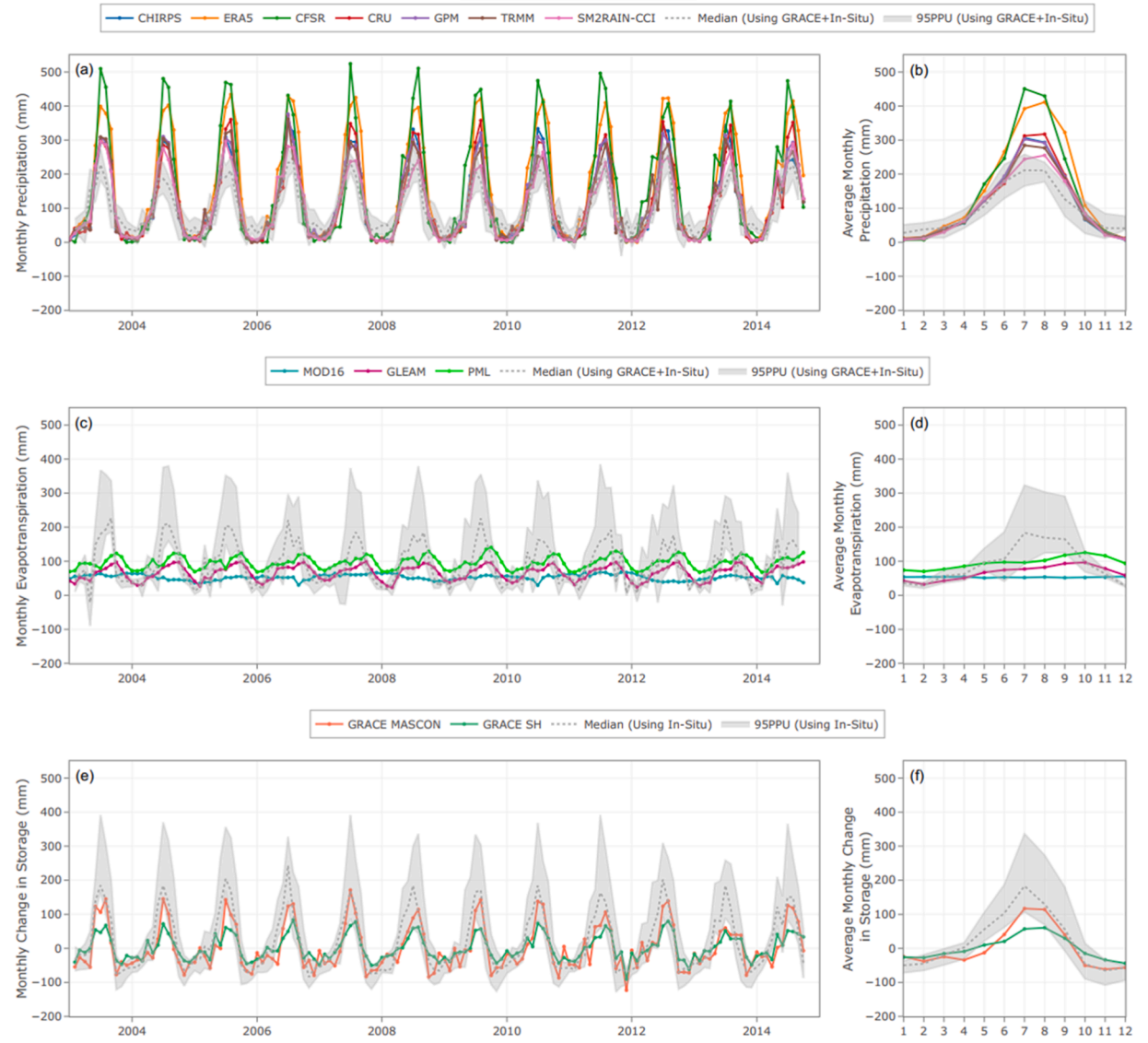

Fig. 6. Panels (a,c,e) demonstrate the estimated P, ETa, and TWS based on reverse water balance calculations using In-Situ runoff in comparison to the different considered products. Panels (b,d,f) illustrate the average monthly values for each variable. The shaded areas represent the 95% confidence bound estimated and the dashed line is the median of the different possible combinations at each case.

median R^2 for the six precipitation products were within a close range, but the major difference occurs when analysing the RMSE. The SM2RAIN-CCI showed the most optimum results in terms of the Median RMSE (48.78 mm). The best median R^2 is scored by CHIRPS followed by SM2RAIN-CCI. Considering the coefficient of determination (R^2) as the sole metric to assess the performance of one product can result in high uncertainty due to the assumption of linear relationship. The median RMSE revealed that GPM, TRMM and CHIRPS have a similar median RMSE. However, the UM showed that CHIRPS and GPM have an equal score (Table S.2). CHIRPS incorporate TRMM Multi-satellite Precipitation Analysis version 7 (TMPA 3B42 v7) data for the calibration process of the global Cold Cloud Duration precipitation estimates (Funk et al., 2015). Since this product is derived from TRMM, this indicate how these remote sensing products are not independent and consider other datasets for the calibration.

This research focused on estimating the runoff using multiple satellites products and investigated the uncertainty associated with these products. The reason of forcing the In-Situ runoff gauge and estimate the TWS, ETa and P is to assess where the imbalance is inherited from. Based on this rearrangement, one can conclude that the P and ETa are the main sources of the imbalance regionally over the UBN basin. Various research papers demonstrated that the ETa is the largest source of uncertainties in the water budget closure (Abera et al., 2017; Soltani et al., 2020). Others claimed that the largest error attributions in closing the water budget is inherited from precipitation which in an agreement with our findings (Sheffield et al., 2009; Wang et al., 2014a; Wong et al., 2021).

In this study, GRACE showed the least error and uncertainty. It is important to mention that GRACE uncertainty arises over glacier

Table 4
The performance of TWS, ETa and P selected products against calculations forced by In-Situ runoff.

Component	Product	Performance	
		Median R^2	Median RMSE (mm)
TWS	mascon	0.72	51.12
	SH	0.72	57.94
ETa	GLEAM	0.38	54.91
	PML	0.17	58.34
	MOD16	0.01	71.15
P	CHIRPS	0.82	56.92
	ERA5	0.79	107.77
	CFSR	0.77	117.22
	CRU	0.79	60.69
	GPM	0.8	56.91
	TRMM	0.77	56.49
	SM2RAIN-CCI	0.81	48.78

regions and basins near the coast due to the limitation of GRACE to model the ice mass balance (Wiese et al., 2016). The UBN is an inner basin which should have low errors and uncertainty. Restoring the leaked signals from the post processing step could results in an uncertainty too, but both GRACE mascon and SH adjust the signal leakage during the postprocessing steps by incorporating the National Centre for Atmospheric Research's Community Land Model 4.0 (NCAR CLM 4.0) (Lawrence et al., 2011) and GLDAS Noah, respectively (Rodell et al., 2004b; Wiese et al., 2016). The postprocessing for both GRACE forms are discussed in detail in Section 2.2.1.

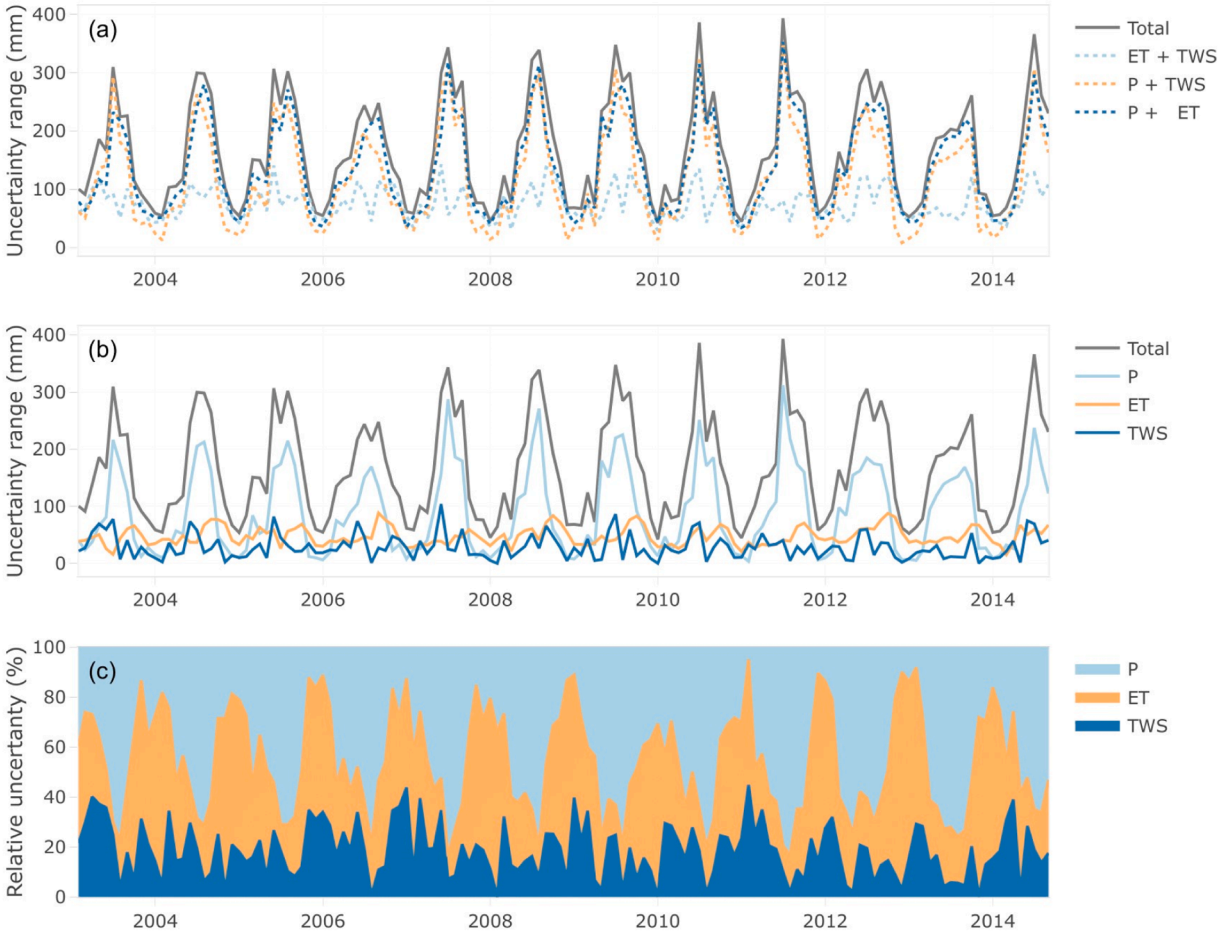

Fig. 7. Uncertainty analysis of water budget components. (a) Total and combined uncertainty ranges, (b) Uncertainty range of individual components, and (c) Relative uncertainty contribution of each individual component (Precipitation, Evapotranspiration, and Terrestrial Water Storage).

4.5. Uncertainty range and relative contribution

The estimated water budget components, namely runoff, precipitation (P), evapotranspiration (ETa), and total water storage (TWS), provide the means to assess both uncertainty and relative contributions. By considering the uncertainty range of runoff across the 42 possible combinations (Fig. 5), one can determine the overall uncertainty, as depicted in Fig. 7.a. Additionally, leveraging the estimated precipitation from Fig. 6.a, we calculated the collective uncertainty range encompassing ETa + TWS, as illustrated in Fig. 7.a. Analogously, we evaluated the combined uncertainty of P + TWS and ETa + TWS using the estimated evapotranspiration and TWS shown in Fig. 7, respectively. Evidently, an annual pattern of these uncertainties emerges, with high total uncertainty during the wet season (reaching 400 mm) and low uncertainty during the dry season (up to 40 mm).

Subsequently, we proceed to quantify the uncertainty associated with each individual component by subtracting the total uncertainty and the combined uncertainties, as depicted in Fig. 7.b. For example, the precipitation uncertainty is determined based on both the total uncertainty and ETa + TWS. Finally, we calculate the relative contribution of uncertainty for each component as a percentage of the total error (Fig. 7.c).

Quantifying the annual average uncertainty contribution of each component reveals that the primary source of uncertainty is precipitation (45%), followed by evapotranspiration (36.6%), while the remaining 18.4% is attributed to TWS. Analysing the average monthly relative contributions in Fig. 8, it becomes evident that from November to February, uncertainty due to ET is dominant, peaking at 59% in January and December. For the rest of the year, precipitation takes the lead, with the highest value (75.5%) occurring in August. A large discrepancy in satellites to estimate precipitation is observed in arid and semi-arid regions (Cattani et al., 2016). This complies to our finding that P is the main source of uncertainties in the UBN basin.

4.6. Further discussion on monthly and annual runoff estimation

Additional to the (1) unified metric techniques (2) the 95% confidence bound, and the (3) water balance rearrangement, the monthly and annual boxplot was conducted for the 42-runoff combination (Fig. 9). The monthly and annual boxplot showed a large variation within the rainfall products. This agrees with the previous analysis discussed earlier that precipitation is very sensitive as compared to the other hydrological parameters (Oliveira et al., 2014; Sheffield et al., 2009; Sneeuw et al., 2014; Wang et al., 2015; Xie et al., 2019). The average monthly observed runoff and the estimated runoff from the best combination (i.e., SM2RAIN-CCI + GLEAM + GRACE SH) are 26 ± 2.6 mm/month and 34 ± 4.9 mm/month, respectively, whereas the annual runoff are 306 ± 44 mm/year and 383 ± 37 mm/year. It is clearly shown that a large bias occurs for the runoff combinations that used CFSR and ERA5 (Fig. 9). This is also shown in the analysis derived from Fig. 4 where the highest slope is for the CFSR and ERA5 products.

Multiple research papers have investigated what an acceptable statistical outcomes are when evaluating the models performance (N. Moriasi et al., 2007; W. Van Liew et al., 2005). The proper selection of evaluation metric is related to the problem the researcher is trying to solve. Since the unified metrics and the overall unified metrics gave a very comprehensive conclusion by relating all the error indexes and the standard regression statistical metrics, it is important to address the best results based on each statistical metric. By examining each statistical metric individually, one can conclude that the statistical weight or priority should be awarded for the statistical metrics with the high variation in the results or for the purpose of the research. For example, the R^2 in this study varies from $R^2 = 0.57$ to $R^2 = 0.85$ (Table S.1). Both the least and the highest score are acceptable in this study therefore, it worths noting that R^2 is not a strong criteria to rely on for this study. The RMSE, MAE and PBIAS are the most reliable and important metrics to focus on when evaluating the error indices. The most optimum runoff combination result with the least RMSE and MAE are scored when using the following combinations: SM2RAIN-CCI (P), GLEAM (ETa), and GRACE SH (TWS). This complies with the results obtained from the Unified Metrics. This step is crucial to ensure that the score is not based solely on the weight from the R^2, slope and y-intercept whereas RMSE, MAE, and PBIAS are important metrics to consider in hydrology filed. The PBIAS is important in judging whether to consider this runoff combination or not. For example, an acceptable PBIAS for the monthly runoff estimation is classified as satisfactory when PBIAS ≤ 25% (N. Moriasi et al., 2007) or ≤ 40% (W. Van Liew et al., 2005). In our research, the most optimum PBIAS was scored for the runoff combination when SM2RAIN-CCI + GLEAM + GRACE mascon is used with a PBIAS of 30%. This outcome aligns with the

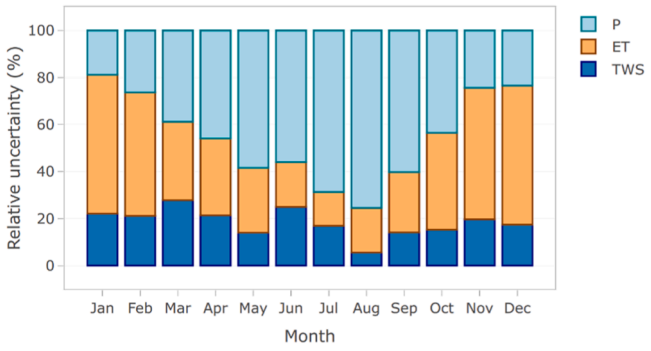

Fig. 8. Average monthly contribution of water budget components to the total uncertainty.

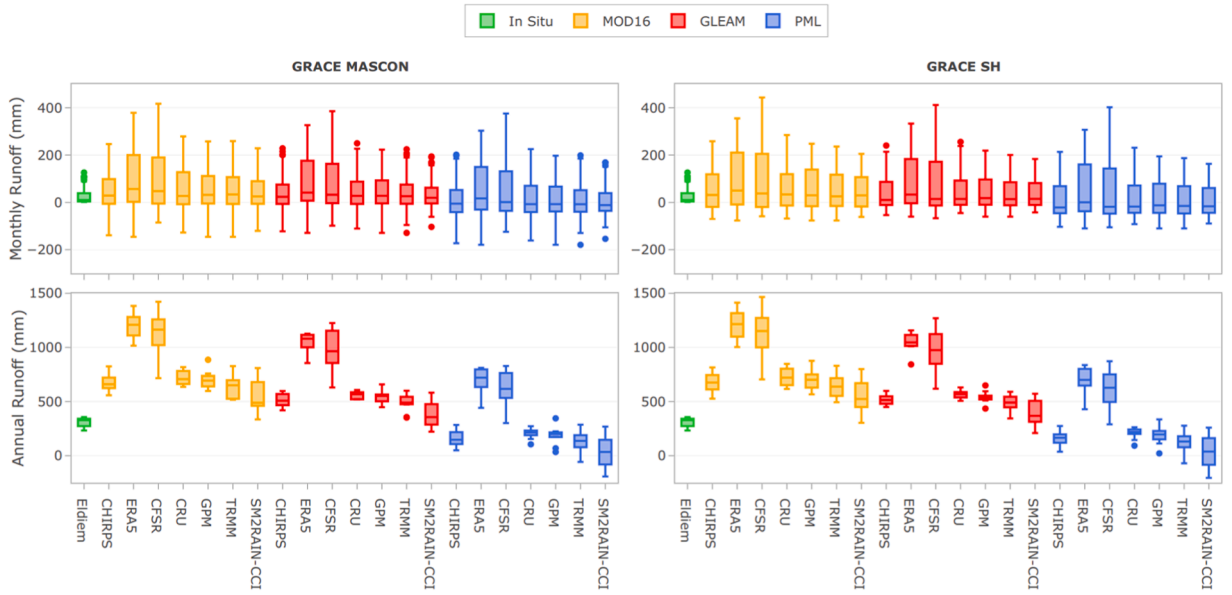

Fig. 9. The monthly and annual runoff magnitudes for the El Diem gauge station and the estimated runoff from 42 remote sensing combinations showing imbalance of the water budget closure. The columns indicate the used TWS product (left: GRACE mascon, right: GRACE SH). The temporal scale is monthly in the first row and annually in the second row.

findings obtained from the overall unified metrics approach and the 95% confidence bound analysis.

Previous research over the UBN basin demonstrated that the most optimum satellites products to estimate the precipitation is CHIRPS (Ali et al., 2023; Ayehu et al., 2018; Bayissa et al., 2017; Jung et al., 2017), and in terms of capturing the total precipitation volume, SM2RAIN-CCI (Abera et al., 2016; Abera et al., 2017), which matches with our analysis if (not) considering the precipitation product derived from SM2RAIN-CCI. These studies did not incorporate the SM2RAIN-CCI in their analysis. The studies that incorporated the SM2RAIN-CCI in the analysis concluded that it outperformed the other products. Various research papers showed that SM2RAIN-CCI datasets outperforms other satellite-based precipitation products due to its unique approach on estimating the precipitation (Brocca et al., 2019; Paredes-Trejo et al., 2019; Pradhan and Indu, 2020).

In terms of the ETa products, certain studies showed that GLEAM and MOD16 caused large discrepancies in estimating the runoff (Abera et al., 2017) whereases one research validated that PMLv2 outperformed MOD16 and GLEAM3.5b over the UBN (McNamara et al., 2021). In our research, the result revealed that the recent version from GLEAM 3.6b outperformed MOD16 and PMLv2. GLEAM and MOD16 were evaluated over 837 catchments over the globe and revealed similar results to our finding that GLEAM outperformed MOD16 (Miralles et al., 2016). GLEAM outperformed PMLv2 and showed lower uncertainty as compared to PMLv2 when compared with the flux towers (Li et al., 2022). The discrepancies between the different ET products are caused by the bias in the forcing meteorological data, lack of validation data, and the mismatch in the spatial and temporal resolution.

As for the terrestrial water storage, most studies have either averaged the SH from different centres or used GRACE mascon for validation purpose. However, neither of the two GRACE products has been assessed nor validated over the UBN basin (Abera et al., 2017; Koukoula et al., 2020). Conversely, in our study, we found that GRACE mascon slightly outperformed GRACE SH when tested using the OUM (Table 3). In terms of the individual assessment of satellite products, GRACE mascon is slightly closer than GRACE SH to the optimum values of the performance metrics (Fig. 4). However, when the goal is to estimate runoff, the combination having SM2RAIN-CCI, GLEAM, and GRACE SH is the most optimum solution. The difference in choosing the best products is subjective to the unique objective or question the author is trying to answer. The spatial distribution of the water balance components is beyond the scope of this research; therefore, the assessment of the satellites products in this research was within the frame of quantifying the runoff at the outlet of the UBN basin. The use of remote sensing, leveraging GRACE, to estimate the runoff showed to be a promising potential to provide runoff information for ungauged basins or basins with no ground measurements.

5. Conclusion

The water budget closure using remote sensing products provided an opportunity to estimate the runoff and gave a useful information on the hydrological fluxes (runoff, evapotranspiration, and terrestrial water storage) over the UBN basin, Ethiopia. The assessments of satellite products and reanalysis datasets were achieved by estimating the runoff at the basin's outlet. The Unified Metric and Overall Unified Metric assisted in the evaluation of the reliability of runoff estimations over the basin. The assessment was implemented by incorporating multiple criteria without the reliance on one statistical approach. The different investigated criteria concluded that the best-performing combination using terrestrial water storage, actual evapotranspiration, and precipitation products

are GRACE SH, GLEAM, SM2RAIN-CCI respectively. Moreover, they were also best-performing individual products expect for TWS where GRACE mascon slightly outperformed GRACE SH in terms of the 95% confidence bound. The difference between both GRACE forms is considered minimum. Forcing the In-Situ runoff to estimate the P, TWS flux, and the ETa attributed to reveal source of uncertainties. The study also conclude that precipitation and evapotranspiration are the largest source of uncertainties. The water balance offset was demonstrated in this study using multiple statistical approaches and all the approaches showed similar output except for the TWS flux. Even though the satellites products provided a promising result, access to meteorological data over the UBN is expected to improve the calculation. The intense hydrological network observation is required to have a better management of the recent built dam over the Upper Blue Nile basin. Overall, the analysis and evaluation demonstrated the potential of remote sensing for estimating runoff, which could help address the ongoing challenges in ungauged basins and basins lacking ground information.

CRediT authorship contribution statement

Khaled Alghafli: Formal analysis, Data curation, Visualization, Validation, Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Awad M. Ali:** Visualization, Formal analysis, Validation, Resources, Conceptualization, Writing - review & editing. **Xiaogang Shi:** Supervision, Conceptualization. **William Sloan:** Supervision. **Ali A.A. Obeid:** Conceptualization. **Mohammad Shamsudduha:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that has been used is confidential.

Acknowledgements

We would like to acknowledge the Ministry of Irrigation and Water Resources of Sudan for providing us the discharge data at El Diem.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2023.101545.

References

Abebe, S.A., et al., 2020. Spatial and temporal evaluation of the latest high-resolution precipitation products over the Upper Blue Nile River Basin, Ethiopia. Water. https://doi.org/10.3390/w12113072.

Abebe, S.A., Qin, T., Zhang, X., Li, C., Yan, D., 2022. Estimating the water budget of the Upper Blue Nile River Basin with water and energy processes (WEP) model. Front. Earth Sci. 10.

Abera, W., Brocca, L., Rigon, R., 2016. Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. Atmos. Res. 178–179, 471–483. https://doi.org/10.1016/j.atmosres.2016.04.017.

Abera, W., Formetta, G., Brocca, L., Rigon, R., 2017. Modeling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data. Hydrol. Earth Syst. Sci. 21 (6), 3145–3165. https://doi.org/10.5194/hess-21-3145-2017.

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 Sens. Environ. 252, 112191 https://doi.org/10.1016/j.rse.2020.112191.

Abtew, W., Melesse, A.M., Dessalegne, T., 2009. Spatial, inter and intra-annual variability of the Upper Blue Nile Basin rainfall. Hydrol. Process. 23 (21), 3075–3082. https://doi.org/10.1002/hyp.7419.

Aires, F., 2014. Combining datasets of satellite-retrieved products. Part I: methodology and water budget closure. J. Hydrometeorol. 15 (4), 1677–1691. https://doi.org/10.1175/JHM-D-13-0148.1.

Alghafli, K., et al., 2023. Groundwater recharge estimation using in-situ and GRACE observations in the eastern region of the United Arab Emirates. Sci. Total Environ. 867, 161489 https://doi.org/10.1016/j.scitotenv.2023.161489.

Ali, A.M., Melsen, L.A., Teuling, A.J., 2023. Inferring reservoir filling strategies under limited data availability using hydrological modelling and Earth observation: the case of the Grand Ethiopian Renaissance Dam (GERD). Hydrol. Earth Syst. Sci. Discuss. 2023, 1–39. https://doi.org/10.5194/hess-2023-19.

Allam, M.M., Jain Figueroa, A., McLaughlin, D.B., Eltahir, E.A.B., 2016. Estimation of evaporation over the upper Blue Nile basin by combining observations from satellites and river flow gauges. Water Resour. Res. 52 (2), 644–659. https://doi.org/10.1002/2015WR017251.

Amjad, M., Yilmaz, M.T., Yucel, I., Yilmaz, K.K., 2020. Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. J. Hydrol. 584, 124707 https://doi.org/10.1016/j.jhydrol.2020.124707.

Archfield, S.A., et al., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51 (12), 10078–10091. https://doi.org/10.1002/2015WR017498.

Ayehu, G.T., Tadesse, T., Gessesse, B., Dinku, T., 2018. Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. Atmos. Meas. Tech. 11 (4), 1921–1936. https://doi.org/10.5194/amt-11-1921-2018.

Azarderakhsh, M., Rossow, W.B., Papa, F., Norouzi, H., Khanbilvardi, R., 2011. Diagnosing water variations within the Amazon basin using satellite data. J. Geophys. Res.: Atmospheres 116 (D24). https://doi.org/10.1029/2011JD015997.

Baldocchi, D.D., 2008. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Aust. J. Bot. 56, 1–26.
Basheer, M., Elagib, N.A., 2019. Performance of satellite-based and GPCC 7.0 rainfall products in an extremely data-scarce country in the Nile Basin. Atmos. Res. 215, 128–140. https://doi.org/10.1016/j.atmosres.2018.08.028.
Bayissa, Y., Tadesse, T., Demisse, G., Shiferaw, A., 2017. Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia. Remote Sens. https://doi.org/10.3390/rs9070669.
Beck, H.E., et al., 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615. https://doi.org/10.5194/hess-21-589-2017.
Beck, H.E., et al., 2019. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23 (1), 207–224. https://doi.org/10.5194/hess-23-207-2019.
Bhanja, S.N., Mukherjee, A., Saha, D., Velicogna, I., Famiglietti, J.S., 2016. Validation of GRACE based groundwater storage anomaly using in-situ groundwater level measurements in India. J. Hydrol. 543, 729–738. https://doi.org/10.1016/j.jhydrol.2016.10.042.
Brocca, L., et al., 2014. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. J. Geophys. Res.: Atmospheres 119 (9), 5128–5141. https://doi.org/10.1002/2014JD021489.
Brocca, L., et al., 2019. SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations. Earth Syst. Sci. Data 11 (4), 1583–1601. https://doi.org/10.5194/essd-11-1583-2019.
Brocca, L., Moramarco, T., Melone, F., Wagner, W., 2013. A new method for rainfall estimation through soil moisture observations. Geophys. Res. Lett. 40 (5), 853–858. https://doi.org/10.1002/grl.50173.
Cattani, E., Merino, A., Levizzani, V., 2016. Evaluation of monthly satellite-derived precipitation products over East Africa. J. Hydrometeorol. 17 (10), 2555–2573. https://doi.org/10.1175/JHM-D-15-0042.1.
Chambers, D., 2006. Evaluation of new GRACE time-variable gravity data over the ocean. Geophys. Res. Lett. 331 https://doi.org/10.1029/2006GL027296.
Chen, J., et al., 2020. Basin-scale river runoff estimation from GRACE gravity satellites, climate models, and in situ observations: a case study in the Amazon Basin. Water Resour. Res. 56 (10), e2020WR028032 https://doi.org/10.1029/2020WR028032.
Cheng, M., Ries, J., 2019. GRACE Technical Note 11.
Ciabatta, L., et al., 2018. SM2RAIN-CCI: a new global long-term rainfall data set derived from ESA CCI soil moisture. Earth Syst. Sci. data 10 (1), 267–280. https://doi.org/10.5194/essd-10-267-2018.
Conway, D., 1997. A water balance model of the Upper Blue Nile in Ethiopia. Hydrol. Sci. J. 42 (2), 265–286. https://doi.org/10.1080/02626669709492024.
Conway, D., 2000. The climate and hydrology of the Upper Blue Nile River. Geogr. J. 166 (1), 49–62. https://doi.org/10.1111/j.1475-4959.2000.tb00006.x.
Conway, D., 2005. From headwater tributaries to international river: observing and adapting to climate variability and change in the Nile basin. Glob. Environ. Change 15 (2), 99–114. https://doi.org/10.1016/j.gloenvcha.2005.01.003.
Crochemore, L., et al., 2020. Lessons learnt from checking the quality of openly accessible river flow data worldwide. Hydrol. Sci. J. 65 (5), 699–711. https://doi.org/10.1080/02626667.2019.1659509.
Dee, D.P., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. https://doi.org/10.1002/qj.828.
Dinku, T., Connor, S., Ceccato, P., 2011. Evaluation of satellite rainfall estimates and gridded gauge products over the Upper Blue Nile Region. In: Melesse, A.M. (Ed.), Nile River Basin: Hydrology, Climate and Water Use. Springer, Netherlands, Dordrecht, pp. 109–127. https://doi.org/10.1007/978-94-007-0689-7_5.
Donat, M.G., et al., 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. J. Clim. 27 (13), 5019–5035. https://doi.org/10.1175/JCLI-D-13-00405.1.
Dorigo, W., et al., 2017. ESA CCI soil moisture for improved earth system understanding: state-of-the art and future directions. Remote Sens. Environ. 203, 185–215. https://doi.org/10.1016/j.rse.2017.07.001.
Durand, M., et al., 2010. The surface water and ocean topography mission: observing terrestrial surface water and oceanic submesoscale eddies. Proc. IEEE 98 (5), 766–779.
Eagleson, P.S., 1986. The emergence of global-scale hydrology. Water Resour. Res. 22 (9S), 6S–14S. https://doi.org/10.1029/WR022i09Sp0006S.
Elagib, N.A., Mansell, M.G., 2000. New approaches for estimating global solar radiation across Sudan. Energy Convers. Manag. 41 (5), 419–434. https://doi.org/10.1016/S0196-8904(99)00123-5.
Eom, J., Seo, K.-W., Ryu, D., 2017. Estimation of Amazon River discharge based on EOF analysis of GRACE gravity data. Remote Sens. Environ. 191, 55–66.
Ezega News, 2019. Power Generation Capacity of GERD Slashed to 5150MW – Ethiopian Minister.
Farr, T.G., et al., 2007. The shuttle radar topography mission. Rev. Geophys. 45 (2) https://doi.org/10.1029/2005RG000183.
Ferguson, C.R., Sheffield, J., Wood, E.F., Gao, H., 2010. Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA. Int. J. Remote Sens. 31 (14), 3821–3865. https://doi.org/10.1080/01431161.2010.483490.
Ferreira, V.G., Gong, Z., He, X., Zhang, Y., Andam-Akorful, S.A., 2013. Estimating Total Discharge in the Yangtze River Basin Using Satellite-Based Observations. Remote Sens. 3415–3430. https://doi.org/10.3390/rs5073415.
Field, P.R., Heymsfield, A.J., 2015. Importance of snow to global precipitation. Geophys. Res. Lett. 42 (21), 9512–9520. https://doi.org/10.1002/2015GL065497.
Flechtner, F., Dobslaw, H., Fagiolini, E., 2015. AOD1B Product Description Document for Product Release 05, ⟨http://www-app2.gfz-potsdam.de/pb1/op/grace/results/grav/AOD1B_20151214.pdf⟩.
Fok, H.S., Chen, Y., Ma, Z., Ferreira, V.G., Tenzer, R., 2023. Geographically-weighted water balance approach for satellite-hydrologic runoff estimation in Mekong Basin under ENSO. Int. J. Appl. Earth Obs. Geoinf. 118, 103234 https://doi.org/10.1016/j.jag.2023.103234.
Fu, L.-L., Alsdorf, D., Morrow, R., Rodriguez, E., Mognard, N., 2012. SWOT: the Surface Water and Ocean Topography Mission: Wide-swath Altimetric Elevation on Earth. Jet Propulsion Laboratory, National Aeronautics and Space, Pasadena, CA.
Funk, C., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2 (1), 150066. https://doi.org/10.1038/sdata.2015.66.
Gan, R., et al., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. Ecohydrology 11 (5), e1974. https://doi.org/10.1002/eco.1974.
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. Int. J. Remote Sens. 31 (14), 3955–3978. https://doi.org/10.1080/01431161.2010.483488.
Giorgi, F., Avissar, R., 1997. Representation of heterogeneity effects in Earth system modeling: Experience from land surface modeling. Rev. Geophys. 35 (4), 413–437. https://doi.org/10.1029/97RG01754.
Global Modeling and Assimilation Office, 2004. File Specification for GEOS-DAS Gridded Output Version 5.3. DOI:⟨https://gmao.gsfc.nasa.gov/operations/GMAO-1001v5.3.pdf⟩.
Gupta, H.V., Sorooshian, S., Yapo Patrice, O., 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. J. Hydrol. Eng. 4 (2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135).
Haddad, Z.S., et al., 1997. The TRMM 'day-1' radar/radiometer combined rain-profiling algorithm. J. Meteorol. Soc. Jpn. Ser. II 75 (4), 799–809.
Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci. Data 7 (1), 109. https://doi.org/10.1038/s41597-020-0453-3.
Heinsch, F.A., et al., 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. IEEE Trans. Geosci. Remote Sens. 44 (7), 1908–1925. https://doi.org/10.1109/TGRS.2005.853936.
Hersbach, H., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049. https://doi.org/10.1002/qj.3803.
Hou, A.Y., et al., 2014. The Global Precipitation Measurement Mission. Bull. Am. Meteorol. Soc. 95 (5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., Tan, J., 2019. In: Greenbelt MD, Andrey Savtchenko (Ed.), GPM IMERG Final Precipitation L3 1 day 0.1° x 0.1° V06. Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/GPM/IMERGDF/DAY/06.
Huffman, G.J., et al., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. J. Hydrometeorol. 8 (1), 38–55. https://doi.org/10.1175/JHM560.1.
Janowiak, J.E., Joyce, R.J., Yarosh, Y., 2001. A real-time global half-hourly pixel-resolution infrared dataset and its applications. Bull. Am. Meteorol. Soc. 82 (2), 205–218. https://doi.org/10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2.
Johnson Peggy, A., Curtis, P.D., 1994. Water balance of Blue Nile River Basin in Ethiopia. J. Irrig. Drain. Eng. 120 (3), 573–590. https://doi.org/10.1061/(ASCE)0733-9437(1994)120:3(573).
Jung, H.C., et al., 2017. Upper Blue Nile basin water budget from a multi-model perspective. J. Hydrol. 555, 535–546. https://doi.org/10.1016/j.jhydrol.2017.10.040.
Kim, U., Kaluarachchi, J.J., 2009. Hydrologic model calibration using discontinuous data: an example from the upper Blue Nile River Basin of Ethiopia. Hydrol. Process. 23 (26), 3705–3717. https://doi.org/10.1002/hyp.7465.
Knapp, K.R., et al., 2011. Globally Gridded satellite observations for climate studies. Bull. Am. Meteorol. Soc. 92 (7), 893–907. https://doi.org/10.1175/2011BAMS3039.1.
Koblinsky, C.J., Clarke, R.T., Brenner, A.C., Frey, H., 1993. Measurement of river level variations with satellite altimetry. Water Resour. Res. 29 (6), 1839–1848. https://doi.org/10.1029/93WR00542.
Koukoula, M., Nikolopoulos, E.I., Dokou, Z., Anagnostou, E.N., 2020. Evaluation of global water resources reanalysis products in the Upper Blue Nile River Basin. J. Hydrometeorol. 21 (5), 935–952. https://doi.org/10.1175/JHM-D-19-0233.1.
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 Sens. Environ. 93 (1), 238–245. https://doi.org/10.1016/j.rse.2004.07.007.
Kummerow, C., Barnes, W., Kozu, T., Shiue, J., Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. J. Atmos. Ocean. Technol. 15 (3), 809–817.
Lakew, H.B., Moges, S.A., Asfaw, D.H., 2017. Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile Basin: a case study of Gilgel Abbay. Hydrology. https://doi.org/10.3390/hydrology4030039.
Lakew, H.B., Moges, S.A., Asfaw, D.H., 2020. Hydrological performance evaluation of multiple satellite precipitation products in the upper Blue Nile basin, Ethiopia. J. Hydrol.: Reg. Stud. 27, 100664 https://doi.org/10.1016/j.ejrh.2020.100664.
Landerer, F.W., 2021. CSR TELLUS GRACE Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly Release 6.0 version 04 in netCDF/ASCII/GeoTIFF Formats. NASA Physical Oceanography DAAC. https://doi.org/10.5067/TELND-3AC64.
Landerer, F.W., Cooley, S.S., 2021. Gravity Recovery and Climate Experiment Follow-On (GRACE-FO): level-3 data product user handbook. NASA Jet Propulsion Laboratory, California Institute of Technology.
Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48 (4) https://doi.org/10.1029/2011WR011453.
Lawrence, D.M., et al., 2011. Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. J. Adv. Model. Earth Syst. 3 (1) https://doi.org/10.1029/2011MS00045.
Lazin, R., Shen, X., Koukoula, M., Anagnostou, E., 2020. Evaluation of the Hyper-Resolution Model-Derived Water Cycle Components Over the Upper Blue Nile Basin. J. Hydrol. 590, 125231 https://doi.org/10.1016/j.jhydrol.2020.125231.
Lehmann, F., Vishwakarma, B.D., Bamber, J., 2022. How well are we able to close the water budget at the global scale? Hydrol. Earth Syst. Sci. 26 (1), 35–54. https://doi.org/10.5194/hess-26-35-2022.
Leuning, R., Zhang, Y., 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 Resour. Res. 44 (10).
Li, C., et al., 2022. Error characterization of global land evapotranspiration products: Collocation-based approach. J. Hydrol. 612, 128102 https://doi.org/10.1016/j.jhydrol.2022.128102.
Liu, W., et al., 2018. Investigating water budget dynamics in 18 river basins across the Tibetan Plateau through multiple datasets. Hydrol. Earth Syst. Sci. 22 (1), 351–371. https://doi.org/10.5194/hess-22-351-2018.
Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50 (2), 1131–1151. https://doi.org/10.1002/2013WR014581.
Luo, Z., et al., 2021. A new method for assessing satellite-based hydrological data products using water budget closure. J. Hydrol. (Amst.) 594, 125927. https://doi.org/10.1016/j.jhydrol.2020.125927.
Martens, B., et al., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10 (5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.
McNamara, I., et al., 2021. How well do gridded precipitation and actual evapotranspiration products represent the key water balance components in the Nile Basin? J. Hydrol.: Reg. Stud. 37, 100884 https://doi.org/10.1016/j.ejrh.2021.100884.
Mei, Y., Anagnostou, E.N., Nikolopoulos, E.I., Borga, M., 2014. Error Analysis of Satellite Precipitation Products in Mountainous Basins. J. Hydrometeorol. 15 (5), 1778–1793. https://doi.org/10.1175/JHM-D-13-0194.1.
Miralles, D.G., et al., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15 (2), 453–469. https://doi.org/10.5194/hess-15-453-2011.
Miralles, D.G., et al., 2016. The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. Hydrol. Earth Syst. Sci. 20 (2), 823–842. https://doi.org/10.5194/hess-20-823-2016.
Mishra, A., Hata, T., 2006. A grid-based runoff generation and flow routing model for the Upper Blue Nile basin. Hydrol. Sci. J. 51 (2), 191–206. https://doi.org/10.1623/hysj.51.2.191.
Mohanasundaram, S., et al., 2021. An application of GRACE mission datasets for streamflow and baseflow estimation in the Conterminous United States basins. J. Hydrol. 601, 126622 https://doi.org/10.1016/j.jhydrol.2021.126622.
Moreira, A.A., et al., 2019. Assessment of terrestrial water balance using remote sensing data in South America. J. Hydrol. 575, 131–147. https://doi.org/10.1016/j.jhydrol.2019.05.021.
Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111 (4), 519–536. https://doi.org/10.1016/j.rse.2007.04.015.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.
Mugnai, A., Levizzani, V., Bauer, P., Turk, J., 2007. Snowfall measurements by proposed European GPM Mission. Springer-Verlag,, pp. 655–674.
Munier, S., et al., 2014. Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model. J. Geophys. Res.: Atmospheres, 119(21): 12,100-12,116. https://doi.org/10.1002/2014JD021953.
N. Moriasi, D., et al., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50 (3), 885–900. https://doi.org/10.13031/2013.23153.
Neves, M.C., Nunes, L.M., Monteiro, J.P., 2020. Evaluation of GRACE data for water resource management in Iberia: a case study of groundwater storage monitoring in the Algarve region. J. Hydrol. Reg. Stud. 32 (C), 100734 https://doi.org/10.1016/j.ejrh.2020.100734.
Oliveira, P.T.S., et al., 2014. Trends in water balance components across the Brazilian Cerrado. Water Resour. Res. 50 (9), 7100–7114. https://doi.org/10.1002/2013WR015202.
Pan, M., et al., 2012. Multisource Estimation of Long-Term Terrestrial Water Budget for Major Global River Basins. J. Clim. 25 (9), 3191–3206. https://doi.org/10.1175/JCLI-D-11-00300.1.

Pan, M., Wood, E.F., 2006. Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. J. Hydrometeorol. 7 (3), 534–547. https://doi.org/10.1175/JHM495.1.
Paredes-Trejo, F., Barbosa, H., dos Santos, C.A.C., 2019. Evaluation of the Performance of SM2RAIN-Derived Rainfall Products over Brazil. Remote Sens. https://doi.org/10.3390/rs11091113.
Peltier, R., Donald F, W.A., Drummond, R., 2018. Comment on "An Assessment of the ICE-6G_C (VM5a) Glacial Isostatic Adjustment Model" by Purcell et al. J. Geophys. Res.: Solid Earth 123 (2), 2019–2028. https://doi.org/10.1002/2016JB013844.
Penatti, N.C., Almeida, T.I.Rd, Ferreira, L.G., Arantes, A.E., Coe, M.T., 2015. Satellite-based hydrological dynamics of the world's largest continuous wetland. Remote Sens. Environ. 170, 1–13. https://doi.org/10.1016/j.rse.2015.08.031.
Pradhan, A., Indu, J., 2020. Performance Assessment of GPM IMERG, SM2RAIN-CCI, SM2RAIN-ASCAT rainfall estimates over the Indian Subcontinent, 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), pp. 164–167. DOI:⟨10.1109/ICSIDEMPC49020.2020.9299605⟩.
Priestley, C.H.B., Taylor, R.J., 1972. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Mon. Weather Rev. 100 (2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.
Rodell, M., et al., 2004a. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31 (20) https://doi.org/10.1029/2004GL020873.
Rodell, M., et al., 2004b. The Global Land Data Assimilation System. Bull. Am. Meteorol. Soc. 85 (3), 381–394. https://doi.org/10.1175/BAMS-85-3-381.
Rodell, M., Famiglietti, J.S., 2001. An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). Water Resour. Res. 37 (5), 1327–1339. https://doi.org/10.1029/2000wr900306.
Rodell, M., McWilliams, E.B., Famiglietti, J.S., Beaudoing, H.K., Nigro, J., 2011. Estimating evapotranspiration using an observation based terrestrial water budget. Hydrol. Process. 25 (26), 4082–4092. https://doi.org/10.1002/hyp.8369.
Saha, S., et al., 2010. The NCEP Climate Forecast System Reanalysis. Bull. Am. Meteorol. Soc. 91 (8), 1015–1058. https://doi.org/10.1175/2010BAMS3001.1.
Sahoo, A.K., et al., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sens. Environ. 115 (8), 1850–1865. https://doi.org/10.1016/j.rse.2011.03.009.
Scanlon, B.R., et al., 2016. Global evaluation of new GRACE mascon products for hydrologic applications. Water Resour. Res. 52 (12), 9412–9429. https://doi.org/10.1002/2016WR019494.
Schreiner-McGraw, A.P., Ajami, H., 2020. Impact of Uncertainty in Precipitation Forcing Data Sets on the Hydrologic Budget of an Integrated Hydrologic Model in Mountainous Terrain. Water Resour. Res. 56 (12), e2020WR027639 https://doi.org/10.1029/2020WR027639.
Senay, G.B., Velpuri, N.M., Bohms, S., Demissie, Y., Gebremichael, M., 2014. Understanding the hydrologic sources and sinks in the N ile B asin using multisource climate and remote sensing data sets. Water Resour. Res. 50 (11), 8625–8650.
Seyoum, W.M., 2018. Characterizing water storage trends and regional climate influence using GRACE observation and satellite altimetry data in the Upper Blue Nile River Basin. J. Hydrol. 566, 274–284. https://doi.org/10.1016/j.jhydrol.2018.09.025.
Shamsudduha, M., et al., 2017. Recent changes in terrestrial water storage in the Upper Nile Basin: an evaluation of commonly used gridded GRACE products. Hydrol. Earth Syst. Sci. 21 (9), 4533–4549. https://doi.org/10.5194/hess-21-4533-2017.
Sheffield, J., Ferguson, C.R., Troy, T.J., Wood, E.F., McCabe, M.F., 2009. Closing the terrestrial water budget from satellite remote sensing. Geophys. Res. Lett. 36 (7) https://doi.org/10.1029/2009GL037338.
Siam, M.S., Demory, M.-E., Eltahir, E.A.B., 2013. Hydrological Cycles over the Congo and Upper Blue Nile Basins: Evaluation of General Circulation Model Simulations and Reanalysis Products. J. Clim. 26 (22), 8881–8894. https://doi.org/10.1175/JCLI-D-12-00404.1.
Sivapalan, M., et al., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48 (6), 857–880. https://doi.org/10.1623/hysj.48.6.857.51421.
Smith, L.C., Isacks, B.L., Bloom, A.L., Murray, A.B., 1996. Estimation of Discharge From Three Braided Rivers Using Synthetic Aperture Radar Satellite Imagery: Potential Application to Ungaged Basins. Water Resour. Res. 32 (7), 2021–2034. https://doi.org/10.1029/96WR00752.
Sneeuw, N., et al., 2014. Estimating Runoff Using Hydro-Geodetic Approaches. Surv. Geophys. 35 (6), 1333–1359. https://doi.org/10.1007/s10712-014-9300-4.
Soltani, S.S., Ataie-Ashtiani, B., Danesh-Yazdi, M., Simmons, C.T., 2020. A probabilistic framework for water budget estimation in low runoff regions: A case study of the central Basin of Iran. J. Hydrol. 586, 124898 https://doi.org/10.1016/j.jhydrol.2020.124898.
Steenhuis, T.S., et al., 2009. Predicting discharge and sediment for the Abay (Blue Nile) with a simple model. Hydrol. Process. 23 (26), 3728–3737. https://doi.org/10.1002/hyp.7513.
Sun, Y., Riva, R., Ditmar, P., 2016. Optimizing estimates of annual variations and trends in geocenter motion and J2 from a combination of GRACE data and geophysical models. J. Geophys. Res.: Solid Earth 121 (11), 8352–8370. https://doi.org/10.1002/2016JB013073.
Swenson, S., Wahr, J., 2006. Post-processing removal of correlated errors in GRACE data. Geophys. Res. Lett. 33 (8) https://doi.org/10.1029/2005GL025285.
Syed, T.H., et al., 2005. Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance. Geophys. Res. Lett. 32 (24) https://doi.org/10.1029/2005GL024851.
Syed, T.H., Famiglietti, J.S., Chambers, D.P., 2009. GRACE-Based Estimates of Terrestrial Freshwater Discharge from Basin to Continental Scales. J. Hydrometeorol. 10 (1), 22–40. https://doi.org/10.1175/2008JHM993.1.
Tan, X., Liu, B., Tan, X., Chen, X., 2022. Long-Term Water Imbalances of Watersheds Resulting From Biases in Hydroclimatic Data Sets for Water Budget Analyses. Water Resour. Res. 58 (3), e2021WR031209 https://doi.org/10.1029/2021WR031209.
Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE Measurements of Mass Variability in the Earth System. Science 305 (5683), 503–505. https://doi.org/10.1126/science.1099192.
Tekleab, S., et al., 2011. Water balance modeling of Upper Blue Nile catchments using a top-down approach. Hydrol. Earth Syst. Sci. 15 (7), 2179–2193. https://doi.org/10.5194/hess-15-2179-2011.
Trambauer, P., et al., 2014. Comparison of different evaporation estimates over the African continent. Hydrol. Earth Syst. Sci. 18 (1), 193–212. https://doi.org/10.5194/hess-18-193-2014.
W. Van Liew, M., G. Arnold, J., D. Bosch, D., 2005. Problems and potential of autocalibrating a hydrologic model. Trans. ASAE 48 (3), 1025–1040. https://doi.org/10.13031/2013.18514.
Wang, S., et al., 2014a. Assessment of water budget for sixteen large drainage basins in Canada. J. Hydrol. 512, 1–15. https://doi.org/10.1016/j.jhydrol.2014.02.058.
Wang, S., McKenney, D.W., Shang, J., Li, J., 2014b. A national-scale assessment of long-term water budget closures for Canada's watersheds. J. Geophys. Res.: Atmospheres 119 (14), 8712–8725. https://doi.org/10.1002/2014JD021951.
Wang, S., Huang, J., Yang, D., Pavlic, G., Li, J., 2015. Long-term water budget imbalances and error sources for cold region drainage basins. Hydrol. Process. 29 (9), 2125–2136. https://doi.org/10.1002/hyp.10343.
Watkins, M.M., Wiese, D.N., Yuan, D.-N., Boening, C., Landerer, F.W., 2015. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. J. Geophys. Res.: Solid Earth 120 (4), 2648–2671. https://doi.org/10.1002/2014JB011547.
Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., van Griensven, A., 2020. Can we trust remote sensing evapotranspiration products over Africa? Hydrol. Earth Syst. Sci. 24 (3), 1565–1586. https://doi.org/10.5194/hess-24-1565-2020.
Wiese, D.N., Landerer, F.W., Watkins, M.M., 2016. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. Water Resour. Res. 52 (9), 7490–7502. https://doi.org/10.1002/2016WR019344.
Williams, M., 2019. The Atbara. In: Williams, M. (Ed.), The Nile Basin: Quaternary Geology, Geomorphology and Prehistoric Environments. Cambridge University Press, Cambridge, pp. 164–175. https://doi.org/10.1017/9781316831885.013.
Wong, J.S., et al., 2021. Assessing water balance closure using multiple data assimilation– and remote sensing–based datasets for Canada. J. Hydrometeorol. 22 (6), 1569–1589. https://doi.org/10.1175/JHM-D-20-0131.1.

Worqlul, A.W., et al., 2018. Performance of bias corrected MPEG rainfall estimate for rainfall-runoff simulation in the upper Blue Nile Basin, Ethiopia. J. Hydrol. 556, 1182–1191. https://doi.org/10.1016/j.jhydrol.2017.01.058.

Wrzesien, M.L., Pavelsky, T.M., Durand, M.T., Dozier, J., Lundquist, J.D., 2019. Characterizing Biases in Mountain Snow Accumulation From Global Data Sets. Water Resour. Res. 55 (11), 9873–9891. https://doi.org/10.1029/2019WR025350.

Xie, J., Xu, Y.-P., Gao, C., Xuan, W., Bai, Z., 2019. Total Basin Discharge From GRACE and Water Balance Method for the Yarlung Tsangpo River Basin, Southwestern China. J. Geophys. Res.: Atmospheres 124 (14), 7617–7632. https://doi.org/10.1029/2018JD030025.

Yao, Y., et al., 2014. Estimation of the terrestrial water budget over northern China by merging multiple datasets. J. Hydrol. 519, 50–68. https://doi.org/10.1016/j.jhydrol.2014.06.046.

Zakharova, E., Nielsen, K., Kamenev, G., Kouraev, A., 2020. River discharge estimation from radar altimetry: Assessment of satellite performance, river scales and methods. J. Hydrol. 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 Geosci. 338 (3), 188–196. https://doi.org/10.1016/j.crte.2005.10.003.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., Ribbe, L., 2017. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. Hydrol. Earth Syst. Sci. 21 (2), 1295–1320. https://doi.org/10.5194/hess-21-1295-2017.

Zandler, H., Haag, I., Samimi, C., 2019. Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. Sci. Rep. 9 (1), 15118. https://doi.org/10.1038/s41598-019-51666-z.

Zhang, Y., et al., 2010. Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05 spatial resolution. Water Resour. Res. 46 (5).

Zhang, Y., et al., 2016. Multi-decadal trends in global terrestrial evapotranspiration and its components. Sci. Rep. 6 (1), 19124. https://doi.org/10.1038/srep19124.

Zhang, Y., et al., 2018. A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. Hydrol. Earth Syst. Sci. 22 (1), 241–263. https://doi.org/10.5194/hess-22-241-2018.

Zhang, Y., et al., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens. Environ. 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031.