Journal of Hydrology: Regional Studies 37 (2021) 100884

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

How well do gridded precipitation and actual evapotranspiration products represent the key water balance components in the Nile Basin?

Ian McNamara [a,*], Oscar M. Baez-Villanueva [a,b], Ali Zomorodian [a], Saher Ayyad [a,c], Mauricio Zambrano-Bigiarini [d,e], Modathir Zaroug [f], Azeb Mersha [g], Alexandra Nauditt [a], Milly Mbuliro [f], Sowed Wamala [f], Lars Ribbe [a]

[a] Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Cologne University of Applied Sciences, Cologne, Germany
[b] Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany
[c] Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany
[d] Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile
[e] Center for Climate and Resilience Research, Universidad de Chile, Santiago, Chile
[f] Nile Basin Initiative Secretariat, Entebbe, Uganda
[g] Eastern Nile Technical Regional Office, Addis Ababa, Ethiopia

ARTICLE INFO

Keywords:
Precipitation
Evapotranspiration
Water balance
Remote sensing
Random Forest
GRACE

ABSTRACT

Study region: Nile Basin, Africa.
Study focus: The accurate representation of precipitation (*P*) and actual evapotranspiration (ETa) patterns is crucial for water resources management, yet there remains a high spatial and temporal variability among gridded products, particularly over data-scarce regions. We evaluated the performance of eleven state-of-the-art *P* products and seven ETa products over the Nile Basin using a four-step procedure: (i) *P* products were evaluated at the monthly scale through a point-to-pixel approach; (ii) streamflow was modelled using the Random Forest machine learning technique, and simulated for well-performing catchments for 2009–2018 (to correspond with ETa product availability); (iii) ETa products were evaluated at the multiannual scale using the water balance method; and (iv) the ability of the best-performing *P* and ETa products to represent monthly variations in terrestrial water storage (ΔTWS) was assessed through a comparison with GRACE Level-3 data.
New hydrological insights for the region: CHIRPSv2 was the best-performing *P* product (median monthly KGE' of 0.80) and PMLv2 and WaPORv2.1 the best-performing ETa products over the majority of the evaluated catchments. The application of the water balance using these best-performing products captures the seasonality of ΔTWS well over the White Nile Basin, but overestimates seasonality over the Blue Nile Basin. Our study demonstrates how gridded *P* and ETa products can be evaluated over extremely data-scarce conditions using an easily transferable methodology.

* Corresponding author at: Betzdorfer Str. 2, 50679 Cologne, Germany
E-mail address: ian.mcnamara01@gmail.com (I. McNamara).

https://doi.org/10.1016/j.ejrh.2021.100884
Received 6 May 2021; Received in revised form 23 July 2021; Accepted 26 July 2021
Available online 14 August 2021

Table 1
Summary of most relevant studies of gridded *P* and ETa products in the Nile Basin.

Study	Region	Products	Temp. scale	Relevant conclusions
P studies				
Dinku et al. (2008a)	Ethiopian highlands and Zimbabwe	CMORPH, NRLB, PERSIANN, RFEv2, TRMM-3B42, TRMM-3B42RT	Daily, 10-day	Over Ethiopia, CMORPH and TRMM-3B42 performed the best. PERSIANN was the only product with a bias greater than 1.
Dinku et al. (2008b)	Ethiopian highlands	CRU TS2.1, GPCC-mon, GPCC-ful v3, GPCC-clm v1.1, NOAA–CPC	Monthly	GPCC-ful v3 was the best-performing product, followed by NOAA-CPC.
Romilly and Gebremichael (2011)	Ethiopia	CMORPH, PERSIANN, TRMM 3B42RT	Multiannual	Generally, TRMM 3B42RT and CMORPH outperformed PERSIANN. Over the Blue Nile, the bias ratio decreased with increasing elevation.
Thiemig et al. (2012)	Baro-Akobo Basin	CMORPH, ERA-Interim, GPROF 6.0, GSMaP-MVK, PERSIANN, RFEv2, TRMM 3B42v6	Daily, monthly, seasonal, multiannual	In the point-to-pixel analysis, CMORPH was the best-performing product in terms of both pBIAS and *r*.
Diem et al. (2014)	Western Uganda	ARC2, RFEv2, TRMM 3B42v7	Daily, 10-day, monthly, seasonal	TRMM 3B42v7 was the best-performing product; however, all three products performed poorly in the southern part of the study area.
Abera et al. (2016)	Upper Blue Nile Basin	CFSR, CMORPH, SM2R-CCI, TAMSAT, TRMM 3B42v7	Daily	CMORPH, TAMSAT and SM2R-CCI were the best-performing products.
Bayissa et al. (2017)	Upper Blue Nile	ARC2, CHIRPSv2, PERSIANN-CDR, TARCATv2, TRMM 3B42v7	10-day, monthly, seasonal	TARCATv2 and CHIRPSv2 performed the best overall and CHIRPSv2 at the monthly temporal scale. All products underestimated seasonal rainfall amounts.
Ayehu et al. (2018)	Upper Blue Nile Basin	ARC2, CHIRPSv2, TAMSATv2, TAMSATv3	10-day, monthly	CHIRPSv2 was the best-performing product at the dekadal (i.e., 10-day) and monthly temporal scales.
Dinku et al. (2018)	Eastern Africa	ARC2, CHIRP, CHIRPSv2, TAMSATv3	Daily, 10-day, monthly	CHIRP and CHIRPSv2 performed best at the dekadal and monthly time scales while TAMSATv3 performed best at the daily temporal scale.
Basheer and Elagib (2019)	South Sudan	ARC2, CHIRPSv2, GPCCv7.0, MSWEPv2.0, PERSIANN-CDR, TAMSATv2	Monthly, annual	GPCCv7.0 was the best-performing product, followed by CHIRPSv2. Note that rain gauge data from the five evaluated stations are used to correct some of the evaluated products.
Camberlin et al. (2019)	Five regions over Central Africa	ARCv2, CHIRPSv2, CMORPH, PERSIANN-CDR, TAMSAT/TARCATv3, TAPEER, TRMM 3B42v7	Daily, monthly	In eastern Democratic Republic of the Congo, Burundi and western Uganda (the only studied area overlapping the Nile Basin), TAPEER performed best at the daily temporal scale (correlation coefficient), while TRMM 3B42v7 and CHIRPSv2 performed best at the monthly temporal scale.
Abebe et al. (2020)	Upper Blue Nile Basin	CHIRPSv2, ENACTS, IMERGv6, MSWEPv2.2, SM2RAIN-ASCATv1.1	Daily, monthly, seasonal	At the daily and monthly temporal scales, the best-performing products were ENACTS and CHIRPSv2, respectively. All product performances decreased for the highest evaluated elevation band (> 2,500 m a.s.l.).
Lakew et al. (2020)	Upper Blue Nile Basin	CMORPH, ERA-Interim, GPCCv1.1, MSWEPv1.1, TRMM 3B42v7	Daily	MSWEPv1.1 was the best-performing product at simulating daily flow over the three catchments. In the highest-elevation catchment, all products except ERA-Interim underestimated total *P*.
Nashwan et al. (2020)	Egypt	ARC2, CHIRPSv2, GSMaPv6, PERSIANN-CCS, TAMSATv3	Daily	ARC2, CHIRPSv2 and GSMaPv6 exhibited the best KGE's over all evaluated stations. Rainfall events were best detected by ARC2 and GSMaPv6.
ETa studies				
Alemu et al. (2014)	Nile Basin	SSEBop, MOD16	Monthly, annual	SSEBop and MOD16 showed similar patterns over rainfed croplands; however, over irrigated croplands they varied in seasonality and magnitude, while over grasslands and wetlands they showed similar seasonality but varied in magnitude.
Al Zayed et al. (2016)	Gezira Irrigation Scheme, Sudan	Product: MODI6A2; other calculation methods: METRIC, SEBAL, SSEB	Daily, seasonal	SSEB and SEBAL provided ETa estimates close to values derived from water balance calculations, while MOD16A2 underestimated ETa. SSEB gave the best performance overall.
Ayyad et al. (2019)	Egypt	EEFlux, MOD16A2, SSEBop	Monthly, annual	SSEBop was the best-performing product, followed by MOD16A2.
Blatchford et al. (2020)	Africa	WaPORv2.1	10-day, annual	WaPORv2.1 shows good spatial and temporal consistency for most climate classes. WaPORv2.1 generally overestimates ETa over irrigated areas, while for rainfed areas, it overestimates low ETa values and underestimates high ETa values.

(*continued on next page*)

Table 1 (*continued*)

Study	Region	Products	Temp. scale	Relevant conclusions
Weerasinghe et al. (2020)	Africa	CMRSET, ETMonitor, GLEAMv3.2a, LandFlux-EVAL, MOD16A3, MTE, SSEBopv4, WaPORv1.1, WECANNv1.0	Multiannual	CMRSET, SSEBopv4 and WaPORv1.1 best estimated long-term ETa quantities across Africa, and along with ETMonitor, best represented ETa spatial patterns. Over the Upper Blue Nile Basin, all products underestimated ETa, while over the Nile Basin, all products except for CMRSET and WaPORv1.1 underestimated ETa.
P & ETa studies				
Bastiaanssen et al. (2014)	Nile Basin	*P*: TRMM 3B43v7, FEWS-NET; ETa: SSEBop	Multiannual	The *Q* derived using the evaluated *P* and ETa products had a good agreement with observed *Q* for the 14 catchments evaluated.
Abera et al. (2017)	Upper Blue Nile Basin	*P*: SM2R-CCI; ETa: GLEAMv3_BETA, MOD16	Daily, monthly	Model estimated ETa showed higher seasonal variations than both the GLEAMv3_BETA and MOD16 products. The combination of SM2R-CCI as a *P* input and model calibrated ETa and *Q* in the application of the water balance at the monthly temporal scale outputted derived ΔTWS very similar to that obtained by GRACE.
Koukoula et al. (2020)	Upper Blue Nile Basin	18 water resources reanalysis datasets	Monthly	Datasets with satellite-derived *P* forcing exhibited better *Q* performance than those that used reanalysis-based *P*. Most products showed an overall underestimation of TWS compared to GRACE, caused by an overall underestimation of ETa.

1. Introduction

Precipitation (*P*) and actual evapotranspiration (ETa; also referred to as evaporation, see Miralles et al., 2020) are two major components of the water cycle, and their accurate characterisation in time and space is of vital importance in hydrology, water resources management, agricultural planning and many other fields. *P* can be measured directly by rain gauges, which provide accurate information at the point scale; however, a dense station network is required to obtain reliable estimates of its spatial patterns (Garcia et al., 2008; Villarini et al., 2008; Adhikary et al., 2015). While *P* is directly measured, methods such as the Bowen ratio and the eddy covariance are commonly used to derive estimates of ETa rates from ground-based measurements at discrete locations, but in addition to the high density of measurements required to understand spatial patterns of ETa, the installation of these instruments can be labour and cost-intensive (Wang and Dickinson, 2012; Zhang et al., 2016).

Over the previous decades, high quality gridded *P* and ETa datasets based on novel techniques have become widely available, which account for the spatial and temporal distributions of these variables. Gridded *P* products are typically based on: (i) the interpolation of ground-based measurements; (ii) satellite-based retrievals (using infrared and/or passive microwave data); and (iii) reanalysis models (Sun et al., 2018). On the other hand, remote sensing based ETa estimates are mainly derived through calculation methods such as the Surface Energy Balance (SEB; Bastiaanssen et al., 1998; Allen et al., 2007), Penman-Monteith (PM; Mu et al., 2007; Leuning et al., 2008), Priestley-Taylor (PT; Fisher et al., 2008), and T_s-VI space methods (Price, 1990), among others (e.g., Ryu et al., 2011; Wang and Bras, 2011). For comprehensive reviews on methodologies used to derive *P* and ETa products, the reader is referred to Sun et al. (2018) and Zhang et al. (2016), respectively.

Gauge-based gridded *P* products lose accuracy over areas where the gauge network density is low (Xie et al., 1996; Schneider et al., 2014). Satellite-based *P* products perform better at longer temporal scales (Maggioni and Massari, 2018; Sun et al., 2018), and their performance is influenced by climate, with lower performances over arid regions (Sun et al., 2018) as well as for intense *P* events (Herold et al., 2017; Baez-Villanueva et al., 2020). Satellite-based products tend to perform better in the tropics because the convective systems are more related to cold cloud-top temperatures (Levizzani et al., 2001; Nauss and Kokhanovsky, 2006), while reanalysis products are consistent with large-scale circulation systems (Gehne et al., 2016) because of their ability of simulate synoptic-scale weather systems (Haiden et al., 2012; Zhu et al., 2014). The performance of *P* products is commonly evaluated through point-to-pixel analyses, hydrological modelling and triple collocation procedures (e.g., Stisen and Sandholt, 2010; Beck et al., 2017; Massari et al., 2017; Zambrano-Bigiarini et al., 2017; Baez-Villanueva et al., 2018).

Gridded ETa products are typically derived through the extrapolation of methodologies that were initially developed for local scale applications (McCabe et al., 2016). Because ETa in gridded products is derived as an energy variable (i.e., energy flux), its estimation relies on accurate representations of numerous input variables, most notably net radiation, near surface air temperature, water vapour pressure and visible and near infrared (VNIR) measurements for phenology and vegetation cover (Fisher et al., 2017). Some relevant conclusions from global evaluations of ETa products are: reanalysis based datasets have the highest global mean estimates, followed by remote sensing based datasets and then land surface models (Mueller et al., 2011); not all products account for irrigated areas (Pan et al., 2020); and that there is an increased variability in ETa estimates over arid and semi-arid regions (Pan et al., 2020). Assessing the water balance closure is a common approach to evaluate the accuracy of these products when insufficient ground-based ETa measurements are available (e.g., Senay et al., 2011; Liu et al., 2016; Moreira et al., 2019; Weerasinghe et al., 2020; Abolafia-Rosenzweig

et al., 2021).

There is no best-performing *P* product or ETa product worldwide, with the relative performance of products specific to the study area (Miralles et al., 2016; Beck et al., 2017), meaning that the performance of a product over a particular area should be assessed prior to any application (Zambrano-Bigiarini et al., 2017; Khan et al., 2018; Zandler et al., 2019). In the data-scarce continent of Africa, there is a high spatial and temporal variability among estimates from gridded *P* and ETa products (Mueller et al., 2011; Trambauer et al., 2014; Awange et al., 2016; Beck et al., 2017; Sun et al., 2018; Weerasinghe et al., 2020), while in the Nile Basin, relatively few ground-based observations are used by some of these gridded products to correct their estimates, resulting in a reduced reliability (Karimi and Bastiaanssen, 2015). For instance, systematic underestimations of *P* have been reported over the northern arid regions of the Nile (Beck et al., 2020) and of ETa over the Blue Nile (Weerasinghe et al., 2020). Table 1 presents a summary of the most relevant studies that have assessed the performances of gridded *P* and ETa products over or within the Nile Basin, where it is evident that most studies assess either *P* products or ETa products rather than both.

Of the identified studies that assess both *P* and ETa product performance over the Nile Basin, Bastiaanssen et al. (2014) considers only two *P* products (TRMM 3B43v7 and FEWS-NET) and one ETa product (SSEBop) to evaluate water balance closure at the multiannual scale. Over the Upper Blue Nile Basin, Abera et al. (2017) used one *P* product (SM2R-CCI) to force a hydrological model and compared the ETa model outputs with two gridded ETa products (GLEAMv3_BETA and MOD16), while Koukoula et al. (2020) assessed

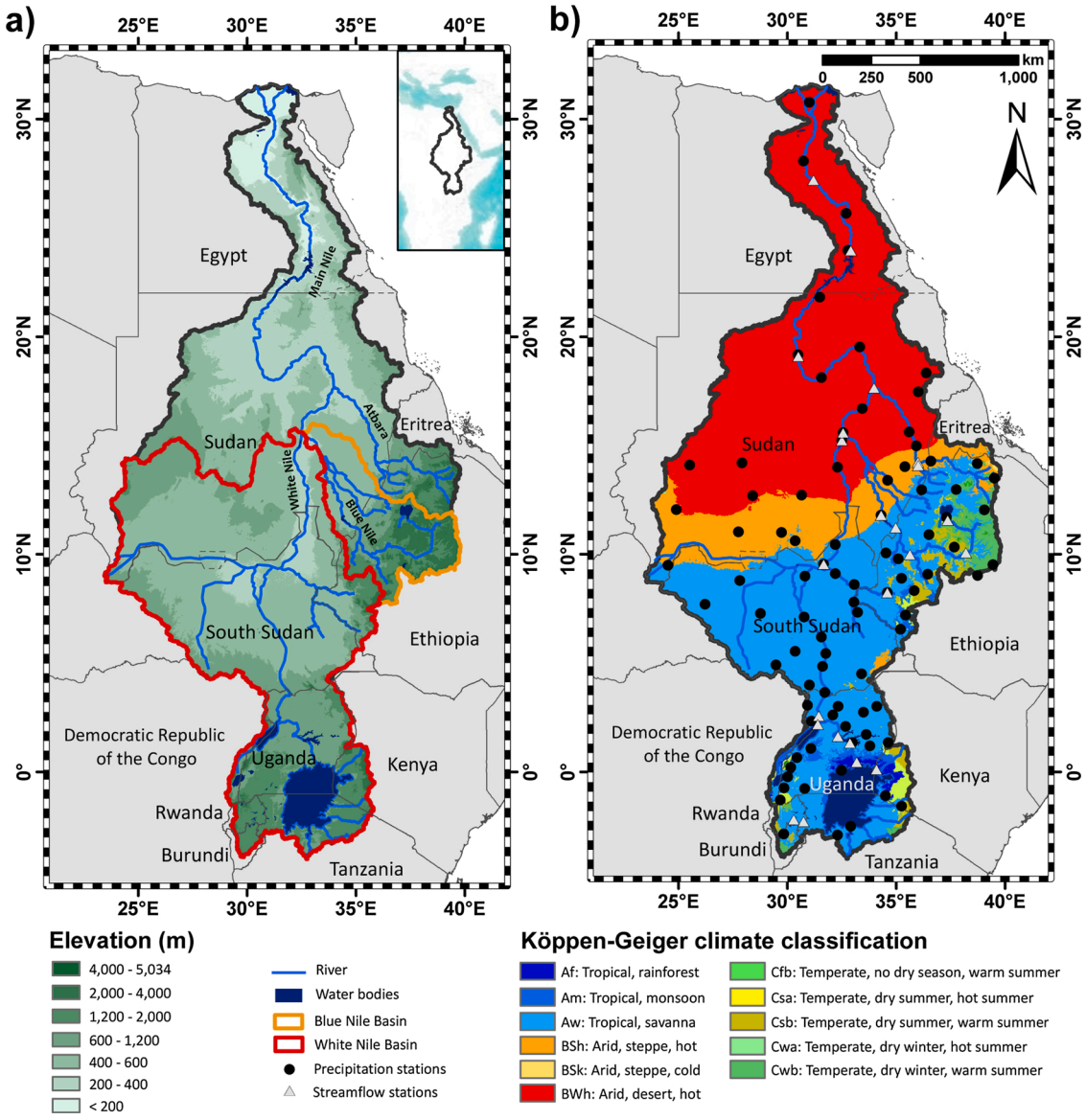

Fig. 1. Study area: (a) elevation (Jarvis et al., 2008); and (b) climate zones according to the Köppen-Geiger climate classification (Beck et al., 2018). The locations of streamflow and precipitation stations used in this study are also depicted.

18 reanalysis datasets at the monthly temporal scale against in situ *P* and *Q* measurements as well as a satellite-based ETa dataset. In addition, two studies listed in Table 1 used the water balance approach to only evaluate ETa product performance: Weerasinghe et al. (2020) used a mean ensemble of gridded *P* inputs to evaluate ETa products at the multiannual scale; and Blatchford et al. (2020) evaluated only one ETa product (WaPORv2.1) over the Nile Basin.

To the best of our knowledge, there is no study that comprehensively assesses the performance of both gridded *P* and ETa products at different temporal scales and over multiple catchments in the Nile Basin. A greater understanding of the spatial and temporal patterns of the key components of the water balance over this basin will greatly assist many decision-making processes related to water resources management, such as the optimisation of agricultural practices, streamflow (*Q*) modelling, drought characterisation, and the development of national and transboundary water plans (Mohamed et al., 2006; Bayissa et al., 2015; Abdelmoneim et al., 2020; Shokr, 2020; Ayyad and Khalifa, 2021). Through the implementation of statistical comparisons between point-based measurements and gridded product estimates, machine learning techniques to simulate *Q* and the application of the water balance, this study evaluates the spatial and temporal performance of numerous gridded *P* and ETa products over the Nile Basin. Based on these results, we assess whether the best-performing *P* and ETa products are able to represent: (i) seasonality of the terrestrial water storage (TWS) over Nile subcatchments; and (ii) the multiannual water balance of the entire Nile Basin.

2. Study area

The Nile River Basin extends over eleven African countries and has a drainage area of about 3.15 million km^2. The upper areas of the basin are located in a tropical climate, the Blue Nile and White Nile merge in Sudan, forming the Main Nile that then flows through the arid Egyptian climate before discharging into the Mediterranean Sea (see Fig. 1). The Nile River, together with its tributaries and lakes, serve domestic and industrial uses as well as irrigated agriculture, hydropower dams and vast fisheries, making the water resources of vital importance for the livelihoods of those living in the basin (Awulachew, 2012). Despite its importance, data are scarce, and disputes regarding water resources have led to minimal data sharing (Swain, 2011; Johnston and Smakhtin, 2014; Digna et al., 2017; Zeitoun et al., 2020).

3. Data

3.1. Ground-based data

To evaluate the performance of the gridded *P* and ETa products over the Nile Basin, we used monthly time series from 90 rain gauges and 23 *Q* stations provided by the Nile Basin Initiative (NBI). The locations of these stations are shown in Fig. 1. The temporal availability of ground-based data varies according to station; 2011 is the last year for which any station has *P* data available, and 2014 for *Q* data. Information related to the location, period of data availability and station administrator of the rain gauge data is included in Table S1 of the Supplementary Material.

3.2. Gridded data

3.2.1. Gridded precipitation products

Based on results of previous studies, eleven *P* products were selected for evaluation: African Rainfall Climatology, version 2 (ARC2, Novella and Thiaw, 2013); Climate Hazards InfraRed Precipitation with Stations data version 2.0 (CHIRPSv2, Funk et al., 2015); Climate Prediction Center (CPC) Morphing technique version 1.0-BLD, gauge-satellite blended precipitation product (CMORPHv1,

Table 2
Summary of the evaluated *P* products.

Product	Product type	Spatial resolution[a]	Spatial coverage	Temporal resolution[b]	Time period
ARC2	Satellite and gauge	0.10° (~11 km)	Africa	Daily[c]	1983–present
CHIRPSv2	Satellite, reanalysis and gauge	0.05° (~5.5 km)	50°S–50°N	Daily	1981–present
CMORPHv1	Satellite and gauge	0.25° (~28 km)	60°S–60°N	3-hourly	1998–present
CRU TS4.04	Gauge	0.50° (~55 km)	Global	Monthly	1901–present
ERA5	Reanalysis	0.25° (~28 km)	Global	Monthly	1981–present
GPCCv2020	Gauge	0.25° (~28 km)	Global	Monthly	1891–present
IMERGv6	Satellite and gauge	0.10° (~11 km)	Global	Daily	2000–present
MSWEPv2.2	Merged	0.10° (~11 km)	Global	Daily	1979–2021
PERSIANN-CDR	Satellite and gauge	0.25° (~28 km)	60°S–60°N	6-hourly[c]	1983–present
TAMSATv3.1	Satellite and gauge	0.25° (~28 km)[d]	Africa	Daily	1983–present
TRMM 3B42v7	Satellite and gauge	0.25° (~28 km)	50°S–50°N	3-hourly	1998–2019

[a] Equivalent grid size in kilometres is calculated at the equator.
[b] Most *P* products are available at multiple temporal resolutions; the resolution listed in the table refers to the product downloaded for this study.
[c] Data not available for all time steps.
[d] Dataset has a spatial resolution of 0.0375° but a rescaled dataset at 0.25° was acquired.

Joyce et al., 2004; Xie et al., 2017); Climatic Research Unit Time-Series version 4.04 (CRU TS4.04, Harris et al., 2020); The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis dataset (ERA5, Hersbach et al., 2020); Global Precipitation Climatology Centre version 2020 (GPCCv2020, Becker et al., 2013; Schneider et al., 2016, 2020); Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM Final Run (IMERGv6, Huffman et al., 2015, 2019); Multi-Source Weighted-Ensemble Precipitation (MSWEPv2.2, Beck et al., 2019); Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Climate Data Record (PERSIANN-CDR, Hsu et al., 1997; Ashouri et al., 2015); Tropical Applications of Meteorology using SATellite and ground based observations version 3.1 (TAMSATv3.1, Tarnavsky et al., 2014; Maidment et al., 2014, 2017); and Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) (TRMM 3B42v7, Huffman et al., 2007, 2010). These eleven products are summarised in Table 2.

3.2.2. Gridded actual evapotranspiration products

We selected seven gridded ETa products that encompass a variety of ETa calculation methodologies (Table 3). The products selected and their method for deriving ETa are as follows:

- The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis dataset (ERA5; Hersbach et al., 2020) derives ETa through a land surface model;
- The NASA Global Land Assimilation System (GLDAS; Rodell et al., 2004) uses satellite and ground-based data to generate optimal fields of land-surface states and fluxes. GLDAS version 2.1 (GLDASv2.1) is forced with a combination of model and observation based forcing datasets (Rui et al., 2018);
- The Global Land Evaporation Amsterdam Model (GLEAM; Miralles et al., 2011; Martens et al., 2016, 2017) estimates different components of land evapotranspiration based on the PT equation. Two versions of the product are available: GLEAMv3.5a, based mainly on reanalysis data; and GLEAMv3.5b, based mainly on satellite data;
- MOD16 (Mu et al., 2007, 2011) ETa is calculated based on the PM equation and uses daily meteorological reanalysis data (air pressure, air temperature, humidity and radiation) in combination with remotely sensed products obtained from MODIS such as albedo, land cover, and vegetation property dynamics;
- The Penman-Monteith-Leuning (PML) model was developed by Leuning et al. (2008). The PMLv2 product couples ET and gross primary products via canopy conductance theory (Gan et al., 2018; Zhang et al., 2019);
- The operational Simplified Surface Energy Balance (SSEBop) product (Senay et al., 2013) combines ET fractions generated from MODIS thermal imagery with reference ET using a thermal index approach. The product uses pre-defined, seasonally dynamic boundary conditions that are unique to each grid-cell for the cold/wet and hot/dry reference points; and
- The FAO Water Productivity through Open access of Remotely sensed derived data (WaPOR; FAO, 2020) actual Evapotranspiration and Interception (ETIa) product estimates of the sum of the soil evaporation, canopy transpiration and evaporation of intercepted *P*. WaPORv2.1 is based on the ETLook model (Pelgrum et al., 2010; Bastiaanssen et al., 2012), which uses the PM equation adapted for remote sensing data.

3.2.3. Potential evaporation and temperature products

In addition to a gridded *P* product (see Section 3.2.1), we used gridded potential evaporation (PET) and temperature (*T*) datasets for *Q* modelling. These products were acquired from the ECMWF ERA5 reanalysis dataset (Hersbach et al., 2020), downloaded at the monthly temporal scale and with a spatial resolution of ~0.28° (30 km at the equator). The PET product is calculated based on the surface energy balance, while *T* represents the mean temperature 2 m above the land surface, calculated through interpolation between the lowest ERA5 model level and the land surface.

3.2.4. GRACE data

The Gravity Recovery And Climate Experiment (GRACE) uses measurements of the relative distance between centre of mass of two

Table 3
Summary of the evaluated ETa products.

Product	Calculation method	Spatial resolution[a]	Spatial coverage	Temporal resolution[b]	Time period
ERA5	Land surface model	0.25° (~28 km)	Global	Monthly	1981–present
GLDASv2.1	Land surface model	0.25° (~28 km)	Global	3 hourly	2000–present
GLEAMv3.5b[c]	Priestley-Taylor	0.25° (~28 km)	Global	Daily	2003–2020
MOD16	Penman-Monteith	0.0043° (~500 m)	Global	Annual[d]	2000–present
PMLv2	Penman-Monteith-Leuning	0.05° (~5.6 km)	Global	8 day	2002–present
SSEBop	Surface energy balance	0.0092° (~1 km)	Global	10 day	2003–present
WaPORv2.1	ETLook	0.0022° (~250 m)	Africa and the Near East	10 day	2009–present

[a] Some products are available at multiple spatial resolutions; the resolution listed in the table refers to the product downloaded for this study.
[b] Most products are available at multiple temporal resolutions; the resolution listed in the table refers to the product downloaded for this study.
[c] The analysis of GLEAMv3.5a was not included in this manuscript because it was consistently outperformed over our study area by GLEAMv3.5b.
[d] The annual product (MOD16A3) was selected for this study because it employs a gap-filling technique for cells where estimates were sometimes not available at the 8-day temporal scale because of cloud cover and aerosols (Running et al., 2019).

identical co-orbiting satellites to derive mean and time variable components of Earth's gravity field, which can be used to monitor changes in water storage, among other processes (Wahr et al., 1998; Rodell and Famiglietti, 1999; Tapley et al., 2004). The measurements are consolidated into a series of gravity field estimates in the form of spherical harmonics, and background models are included so that the contributions of changes in TWS can be isolated (Dahle et al., 2019). GRACE data have been used in numerous studies of water resources both in Africa, and specifically focusing on the Nile Basin (e.g., Shamsudduha et al., 2017; Seyoum, 2018; Abd-Elbaky and Jin, 2019; Hasan et al., 2019, 2021; Koukoula et al., 2020).

Level-3 GRACE gridded data products (RL06v3) from the three official GRACE data centres (JPL, GFZ and UTCSR) were downloaded for period of data availability for the initial GRACE mission (04/2002–06/2017) from the podaac portal: https://podaac.jpl.nasa.gov/ (Yuan, 2018; Dahle et al., 2019; Wiese et al., 2019; Save, 2020). As described by Landerer and Swenson (2012), GRACE Level-3 data are generated by downscaling the filtered standard (native) GRACE products with an approximate spatial resolution of 3° (~330 km at the equator; cell size of approximately 100,000 km^2) to a resolution of 1° (~110 km at the equator). Because of this downscaling process, adjacent cells are highly correlated to each other, meaning that more representative results are obtained when evaluating catchments larger than the ~100,000 km^2 footprint of the native GRACE data (Scanlon et al., 2016; Cooley and Landerer, 2020). The Level-3 products provide information of TWS deviation from a mean condition calculated over 2005–2010.

4. Methods

Fig. 2 illustrates the four-step procedure implemented to evaluate the performance of the gridded *P* and ETa products. The evaluation period for each step was selected based on data availability of ground-based observations and gridded products. First, we evaluated *P* products at the monthly temporal scale through a statistical comparison against the rain gauge data over the period where the majority of *P* products are available (Section 4.1). Because all evaluated ETa products are available for 2009–2018 and we did not have sufficient *Q* data for that period, we simulated *Q* using a Random Forest (RF) model over the 23 catchments from 1981 (when all gridded RF model inputs were available) to the end of the *Q* observations corresponding to each catchment. For the catchments where modelling performed well, we then simulated *Q* for 2009–2018 (Section 4.2). Third, adopting the identified best-performing *P* product and the simulated *Q* time series, we evaluated ETa product performance for each catchment by determining the relative error in closing the water balance at the multiannual scale under the assumption that the multiannual change in storage can be neglected (Section 4.3). Finally, we compared the derived monthly ΔTWS values obtained using the best-performing *P* and ETa products (ΔTWS_{WB}) with those obtained from GRACE Level-3 data (ΔTWS_{GRACE}), starting in 2009 due to ETa product availability and finishing in 2017 due to GRACE data availability. We also validated the assumption that ΔTWS is negligible in the application of the water balance at the multiannual scale (Section 4.4). For Steps 3 and 4, we applied the water balance method at the catchment scale (see Senay et al., 2011; Long et al., 2014; Liu et al., 2016; Moreira et al., 2019; Weerasinghe et al., 2020), according to Eq. (1):

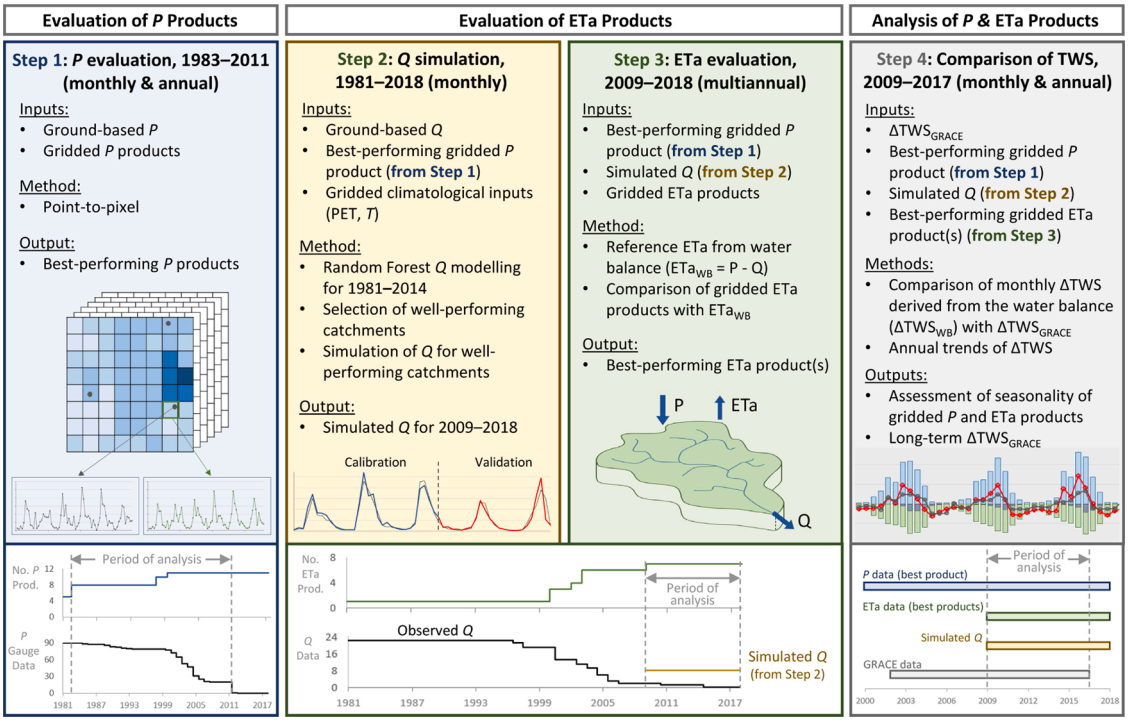

Fig. 2. Overview of the four-step procedure to evaluate the performance of *P* and ETa products. The top panels illustrate the inputs, methods and outputs while the bottom panels illustrate the temporal availability of ground-based data and evaluated products.

$$P - Q - \mathrm{ETa} = \Delta\mathrm{TWS} \tag{1}$$

where P is the precipitation; Q is the catchment runoff; ETa is the actual evapotranspiration; and ΔTWS is the change in terrestrial water storage. All values are expressed in mm, averaged over the area of the catchment.

4.1. Evaluation of precipitation products

The P product performances were evaluated using a point-to-pixel approach at the monthly temporal scale over 1983–2011. The start of the evaluation period was selected because the majority (nine) of the eleven P products have data from 1983 onwards, while the end of the evaluation period was selected because ground-based P data were only available until 2011. In this method, the time series of observed data (here, rain gauge data) are compared against simulated data (here, the corresponding grid-cells from the P products) using a performance index. The point-to-pixel approach assumes that the rain gauge measurements are representative values over the corresponding grid-cell (e.g., Thiemig et al., 2012; Zambrano-Bigiarini et al., 2017; Baez-Villanueva et al., 2018). The P products that were obtained at a sub-monthly scale were aggregated to the monthly scale, except for when temporal data gaps existed, in which case, no monthly values were calculated. To ensure a consistent point-to-pixel comparison, all products were resampled to the same spatial resolution (0.25°) using bilinear interpolation.

Some P products use data from rain gauge stations to correct their estimates; hence the inclusion of these stations in the product evaluation may bias the results. From the 90 stations described in Section 3.1, the 31 that form part of the FAO or GHCN networks, which are used by CHIRPSv2 and MSWEPv2.2 to correct their estimates, were discarded for the evaluation. For performance assessment, to evaluate the P products against the remaining 59 stations, we selected the modified Kling-Gupta efficiency (KGE'; Kling et al., 2012) because: (i) it can be decomposed into three components that allow a better understanding of the origin of mismatches; (ii) in contrast to typical measures of performance based on squared residuals, it avoids assigning a disproportional weight to high values; and (iii) it allows a fair intercomparison of performance across different climatic conditions (Schaefli and Gupta, 2007; Criss and Winston, 2008; Gupta et al., 2009; Thiemig et al., 2013; Baez-Villanueva et al., 2018). The KGE' is described in Eq. (2).

$$\mathrm{KGE}' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{2}$$

The KGE' compares observed data with simulated data, and consists of three components: the Pearson's linear correlation (r; Eq. (3)), which measures the temporal dynamics; the bias ratio (β; Eq. (4)), which measures the total volume (a value greater than 1 represents an overestimation of the simulations while a value less than 1 represents an underestimation); and the variability ratio (γ; Eq. (5)), which measures the relative dispersion between simulations and observations (Gupta et al., 2009; Kling et al., 2012).

$$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}} \tag{3}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{4}$$

$$\gamma = \frac{\mathrm{CV}_s}{\mathrm{CV}_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{5}$$

where n is the number of observations (only when both observed and simulated data exist for that time step); O_i and S_i are the observed and simulated values, respectively, at time i; $\overline{O}$ and $\overline{S}$ are the arithmetic means of the observations and the simulations, respectively; σ is the standard deviation; and μ is the mean. The KGE' and its three components present their optimal value at 1.

4.2. Streamflow modelling and evaluation

To assess the performance of ETa products through the water balance, reliable Q data were required for the period where all ETa estimates are available (2009–2018). Because only one Q station had multiple years of observed data after 2009, we implemented an RF model to simulate monthly Q for 2009–2018 over more catchments.

Machine learning techniques have been widely applied in Q modelling studies (for reviews, see Tyralis et al., 2019; Sit et al., 2020). In this study, we selected an RF model to simulate Q based on gridded P, T, and PET data. RF methods have been widely used in Q prediction for many reasons, including: (i) their demonstrated high predictive performance (Shortridge et al., 2016; Tyralis et al., 2019; Schoppa et al., 2020); (ii) their ability to represent non-linear relationships between climatological variables and Q (Shortridge et al., 2016; Yang et al., 2017; Konapala and Mishra, 2020); and (iii) their ability to not produce biased estimates or lead to overfitting (Díaz-Uriarte and Alvarez de Andrés, 2006; Biau and Scornet, 2016; Hengl et al., 2018; Tyralis et al., 2019). RF is an ensemble machine learning method that works by constructing numerous decision trees using the relationship between independent and dependent variables and can be used for supervised classification and regression tasks (Breiman, 2001; Prasad et al., 2006; Biau and Scornet, 2016). A lumped model was implemented using catchment-averaged monthly time series of the best-performing P product as well as ERA5 PET and T as inputs. The year 1981 was selected as the start date for modelling because it is the first year where both the best-performing P product (see Section 5.1) and ERA5 datasets are available. The end date for modelling each catchment corresponded

to the availability of Q data.

The general formulation used to model the monthly Q was adapted from the approach outlined in Shortridge et al. (2016):

$$Q_t = f(P_t, P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}, T_t, T_{t-1}, T_{t-2}, T_{t-3}, T_{t-4}, \mathrm{PET}_t, \mathrm{PET}_{t-1}, \mathrm{PET}_{t-2}, \mathrm{PET}_{t-3}, \mathrm{PET}_{t-4}) + \varepsilon_t \quad (6)$$

where P_t, T_t, and PET_t are the catchment-averaged monthly total precipitation, mean temperature and mean potential evaporation at month t, respectively; and ε_t is the model error. The subscript $t-n$ represents n months before the month t.

The modelled period for each catchment was divided into a calibration period using the first 70% of the available Q time series, and a validation period using the remaining 30%. Due to the differing length of available time series of Q observations in each catchment, the lengths of these periods varied accordingly. The RF model was trained over the calibration period using the formulation described in Eq. (6), with the following RF configuration: number of trees set to 1000; number of variables randomly sampled as candidates at each split set as the number of predictors divided by 3; and minimum size of terminal nodes set to 5. The performance of Q simulations was evaluated using the KGE' and its components during the validation period. Based on Thiemig et al. (2013), a KGE' greater than 0.75 was used to select a well-performing model. Furthermore, the β component (Eq. (4)) of the KGE' was used to evaluate the ability of the model to represent the total volume of estimated Q because its accurate estimation is crucial in the application of the simplified water balance. Therefore, in addition to ensuring a KGE' > 0.75, only catchments with a β component within the range of $0.80 < \beta < 1.20$ were selected to avoid introducing bias into the evaluation. The implementation of these selection criteria may result in a smaller selection of catchments to evaluate the performance of the ETa products, but at the same time, guarantees a higher degree of reliability in the Q estimates.

For the well-performing catchments, monthly Q time series were generated for 2009–2018. To ensure that the widest possible range of input variables were captured by the RF model, we re-trained the model for each selected catchment using the observed data from the entire period available (post-1981) instead of only the calibration period that was used to test model performance.

4.3. Evaluation of actual evapotranspiration products

We evaluated the performances of the gridded ETa products at the multiannual scale through the application of the water balance method (Eq. (1)) over the catchments where Q modelling performed well. Assuming that ΔTWS is negligible at the multiannual scale (Liu et al., 2016), we calculated the theoretical value of ETa according to the water balance ($\mathrm{ETa_{WB}}$) as:

$$\mathrm{ETa_{WB}} = P - Q \quad (7)$$

where all values are mean annual catchment-averaged accumulations (expressed in mm) and P is calculated using the best-performing P product. We then compared the mean annual ETa for each gridded product with $\mathrm{ETa_{WB}}$, and the closer that the two results are to each other, the better performing the ETa product is inferred to be over that catchment. Note that this method only indicates whether the total ETa estimated over the catchment is accurate, regardless of the spatial distribution of ETa.

Some of the evaluated ETa products have data gaps because estimates are not calculated over certain areas. GLDASv2.1, GLEAMv3.5b, MOD16 and PMLv2 do not calculate ETa over water bodies, while in addition, the MOD16 product does not provide ETa for non-vegetated grid-cells (Running et al., 2019). We did not consider these missing values when calculating catchment-averaged ETa values for products with data gaps. Furthermore, PMLv2 values over large water bodies, which are provided as zeroes, were discarded.

4.4. Ability to represent changes in total water storage

We selected the most downstream catchments with well-performing simulated Q for both the Blue Nile and White Nile to evaluate the best-performing P and ETa products at the monthly temporal scale. We extracted monthly catchment-averaged time series of the best-performing P and ETa products, and combined with the simulated Q, derived monthly time series of expected changes in storage ($\Delta\mathrm{TWS_{WB}}$) for 01/2009–06/2017 according to Eq. (1). We compared these expected values with $\Delta\mathrm{TWS_{GRACE}}$, which we derived from the arithmetic mean of the Level-3 solutions from the three official GRACE data centres. The arithmetic mean was used to reduce noise in the gravity field solutions (Sakumura et al., 2014). Because of the low spatial resolution of the product, we accounted for the percentage of each grid-cell that lies within the respective catchment boundary.

Furthermore, we used $\Delta\mathrm{TWS_{GRACE}}$ to evaluate the assumption that ΔTWS is negligible our application of the water balance used for ETa product evaluation at the multiannual scale. GRACE Level-3 data from the first mission are available for 04/2002–06/2017 and the evaluation of ETa products was selected as 01/2009–12/2018. We calculated the trend of ΔTWS per catchment over the overlap period 01/2009–06/2017, based on fitting a linear trendline to the 12-month rolling average TWS. As recommended by Scanlon et al. (2016) and Cooley and Landerer (2020), we did not evaluate catchments with an area less than 100,000 km^2 because of their smaller size relative to the native spatial resolution of GRACE data (see Section 3.2.4).

For all analyses, we used the R environment 3.6.0 (R Core Team, 2013) and the following R packages: raster (Hijmans, 2020), for the processing and statistical analyses of gridded datasets; hydroGOF (Zambrano-Bigiarini, 2020a), for the calculation of performance indices; hydroTSM (Zambrano-Bigiarini, 2020b), for managing time series data; randomForest (Liaw and Wiener, 2002), for building the RF model to predict Q; and rgdal (Bivand, 2020), for working with vector data. For more information on these R packages, the reader is referred to their respective literature.

5. Results

5.1. Performance of precipitation products

The performances of the eleven *P* products at the monthly temporal scale at the locations of the 59 evaluated rain gauge stations are presented as boxplots in Fig. 3. The results indicate that CHIRPSv2, with a median KGE' of 0.80, was the overall best-performing product, followed by GPCCv2020 (0.78), CRU TS4.04 (0.77), and MSWEPv2.2 (0.75). CHIRPSv2 exhibited the best median bias (1.00) as well as showing the smallest interquartile range for bias (0.11). GPCCv2020 had the best median *r* (0.89); however it presented a higher dispersion in the *r* values compared to CHIRPSv2 and CRU TS4.04. CRU TS4.04 exhibited the best median variability ratio (1.01), followed by CHIRPSv2 (0.99). All products showed similar performances (except for ERA5, ARC2, and CMORPHv1) indicating that the products can represent, to some extent, the monthly *P* patterns over the Nile Basin.

Fig. 4 shows the spatial distribution of the KGE' values, demonstrating that the best-performing *P* products (i.e., CHIRPSv2, GPCCv2020 and CRU TS4.04) perform consistently well over the spatial extent of the study area where stations were available. The northernmost station evaluated (i.e., after excluding FAO and GHCN stations) is located at a longitude of 18.1°, which means that there are no rain gauge data available to evaluate product performance in the arid north of the Nile Basin.

CHIRPSv2 was considered as the best-performing *P* product at the monthly scale because: (i) it outperformed all other *P* products (see Fig. 3); (ii) it showed a consistent high performance throughout the study area at the evaluated locations (see Fig. 4); and (iii) it is almost unbiased, with a median β of 1.00 and low dispersion compared to the other products. These results demonstrate that CHIRPSv2 can be deemed suitable for application in the evaluation of ETa products using the water balance.

5.2. Streamflow modelling

Table 4 summarises the results obtained for modelling *Q* over all catchments, listing the KGE' values for the calibration and validation periods as well as the β values for the validation period. Note that due to *Q* data availability, the modelling period varied among catchments, and the calibration period was always selected as the first 70% of the available time series for each catchment, leaving the remaining 30% for validation. The catchments fulfilling the performance criteria for the validation period (i.e., KGE' $>$ 0.75 and $0.80 < \beta < 1.20$) are underlined and shown in bold text.

The results in Table 4 show that nine catchments fulfilled the designated criteria to be considered as suitable for *Q* simulation for

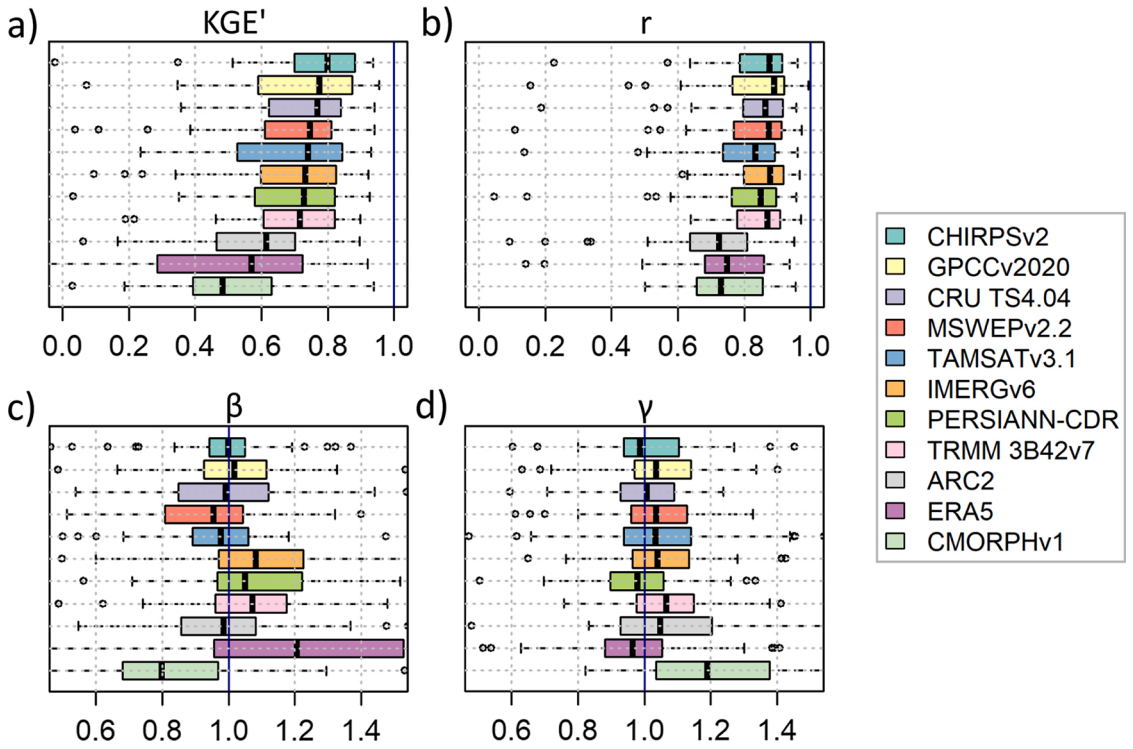

Fig. 3. *P* product performance at the monthly temporal scale: (a) KGE', (b) Pearson's linear correlation (*r*), (c) bias ratio (β); and (d) variability ratio (γ). The solid line represents the median value, the boxes bound the first and third quartiles, the whiskers extend to the most extreme data point that lies within 1.5 times the interquartile range from the box boundary and the dots represent outliers. The vertical blue line indicates the optimal value for the indicators. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

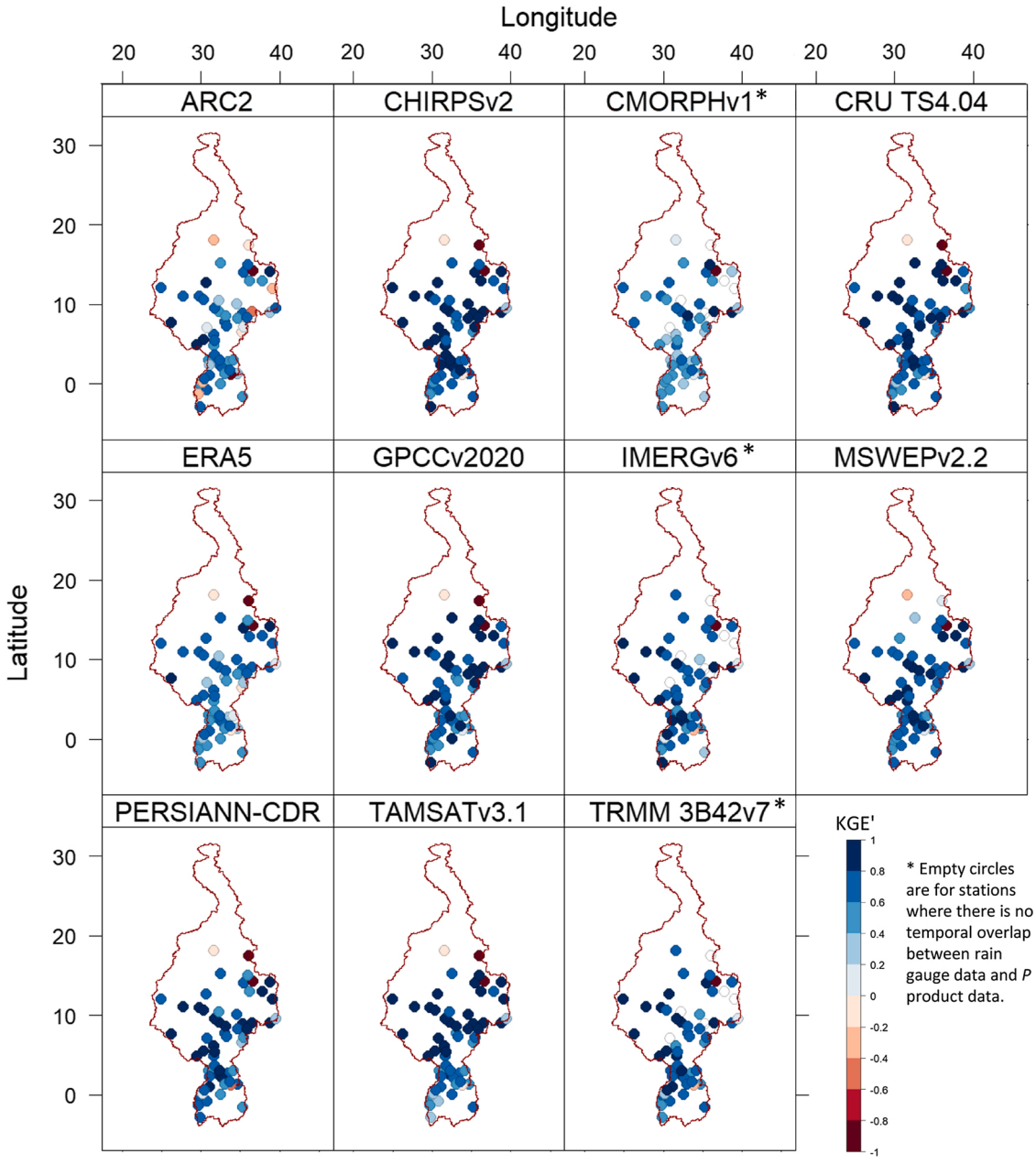

Fig. 4. Spatial performance of the *P* products according to the KGE' at the monthly temporal scale.

2009–2018. For our case study, all nine catchments with KGE' > 0.75 also yielded β values within the range $0.80 < \beta < 1.20$. Fig. 5 illustrates the hydrographs for these nine catchments, displaying both the monthly observed and simulated Q in billion cubic metres (BCM). For all nine catchments, the simulated Q matches the observations relatively well over both the calibration and the validation periods; however, some of the observed high Q values were underestimated. This is evident from the β values reported in Table 4, which are consistently slightly lower than one. For each catchment with good performance, we calculated the difference between total observed and modelled water volumes over the validation period. These differences varied from 1 to 60 mm/yr (mean of 24 mm/yr), averaged over the spatial extents of the corresponding catchments. In general, our RF model performed better in the Blue Nile (KGE' values for the validation period of ~0.80), where the majority of *P* is generated in the tropical and temperate upstream highlands, where a high seasonality in Q is present. In the upstream tropical areas of the White Nile, which is dominated by the African Great Lakes, our model had a reduced performance (KGE' values of ~0.35). In contrast, at the locations further downstream in the White Nile (as well as the Main Nile), the model performance improved (KGE' values of ~0.75). Here, the catchment outlets are located in arid climates, yet the vast majority of Q is generated in the upstream areas.

Table 4
Performance of streamflow modelling according to catchment

Station name	Lat.	Long.	Area (10^3 km^2)[a]	Period of modelling[b]	KGE' calibration	KGE' validation	β validation
Baro Gambella	8.25	34.58	23.5	1990–2006	0.96	0.69	1.25
Blue Nile Bahir Dar	**11.57**	**37.36**	**15.1**	**1981–2000**	**0.84**	**0.77**	**0.81**
Blue Nile Kessie	10.08	38.19	64.7	1981–2004	0.89	0.72	0.73
Blue Nile Mandaia	**10.00**	**35.62**	**139.0**	**1981–2003**	**0.95**	**0.85**	**0.85**
Blue Nile Abbay at Border	**11.23**	**34.97**	**174.4**	**1981–2003**	**0.96**	**0.84**	**0.84**
Blue Nile Roseries Dam	**11.80**	**34.30**	**188.4**	**1981–2005**	**0.96**	**0.87**	**0.87**
Blue Nile Khartoum	**15.60**	**32.55**	**307.1**	**1981–2000**	**0.94**	**0.84**	**0.86**
Kubur Atbara River	14.08	36.00	26.7	1981–2000	0.91	0.53	0.64
Wad El Hilew Atbara River	14.23	35.99	66.7	1981–2004	0.89	0.39	0.72
Victoria Nzoia	0.12	34.09	12.6	1981–2004	0.80	0.57	1.00
Victoria Nyabarongo	− 2.19	30.27	15.5	1981–1996	0.83	0.28	0.80
Victoria Kagera Rusumo	− 2.27	30.73	30.7	1981–1990	0.81	0.39	0.80
Jinja	0.45	33.18	263.4	1981–2014	0.66	− 0.03	1.31
Kyoga Inflow	1.33	32.87	266.8	1981–2000	0.73	0.38	0.96
Kyoga Outflow	1.64	32.31	320.8	1981–2000	0.68	0.28	0.98
Lake Albert Inflow	2.19	31.38	352.5	1981–2000	0.68	0.34	0.94
Albert Nile Panyango	2.57	31.43	409.5	1981–2010	0.61	− 0.38	1.40
White Nile Malakal	**9.55**	**31.65**	**1411.6** [c]	**1981–2005**	**0.92**	**0.88**	**0.93**
White Nile Outflow Jabel Aulia Dam	15.23	32.50	1666.1	1981–2005	0.82	0.62	0.97
Main Nile Hassanab	**17.65**	**33.97**	**2012.2**	**1981–1997**	**0.91**	**0.80**	**0.89**
Main Nile Dongola	**19.10**	**30.49**	**2643.8**	**1981–1997**	**0.93**	**0.79**	**0.84**
Aswan Dam	**23.96**	**32.90**	**2915.0**	**1981–2002**	**0.95**	**0.76**	**0.89**
Assiut Barrage	27.21	31.19	3025.5	1981–2002	0.92	0.73	0.83

[a] Delineated using the 90 m Shuttle Radar Topography Mission (SRTM) version 4.1 digital surface model (DSM) (Jarvis et al., 2008).
[b] Corresponds to all available *Q* data per station after 1981.
[c] Because the use of the SRTMv4.1 90 m could not accurately delineate the catchment boundary over the floodplains in the catchment's north-east, the delineated boundary was adjusted between 9.472N, 31.698E & 8.718N, 34.152E, based on the official catchment boundary provided by the NBI.

5.3. Performance of actual evapotranspiration products

The water balance at the multiannual scale was assessed for the nine catchments where *Q* was simulated for 2009–2018. For each catchment, Fig. 6 lists the value of mean annual ETa calculated according to the water balance (ETa_{WB}) as well as the respective mean annual ETa for all evaluated products. The closer the values of the products are to ETa_{WB}, the better performing the product is inferred to be over that catchment. The numbers in brackets represent the values of ETa according to the evaluated products minus ETa_{WB}, meaning that a positive value indicates that the product overestimates ETa and a negative value indicates underestimation.

The results show PMLv2 as the best-performing of the ETa products, under the assumption that *P* and *Q* are accurately captured. PMLv2 ETa values were the closest to ETa_{WB} in seven of the nine catchments, while WaPORv2.1 and ERA5 performed the best in one catchment each. WaPORv2.1 was the second-closest in seven catchments, PMLv2 in one and SSEBop in one. In general, ERA5, GLDASv2.1, GLEAMv3.5b, MOD16 and SSEBop all appear to systematically underestimate total ETa amounts.

5.4. Changes in terrestrial water storage

5.4.1. Monthly comparison of changes in terrestrial water storage

Figs. 7 and 8 plot the key components of the water balance at the monthly scale for two catchments: Blue Nile Khartoum and White Nile Malakal, respectively. Here, we selected the largest of the assessed catchments corresponding to the Blue Nile and White Nile because: (i) these catchments correspond to the most downstream stations that showed good *Q* modelling performance; (ii) these two basins contribute the majority of *Q* for the Main Nile, hence it is important to understand their hydrological behaviour; and (iii) errors associated with the relatively coarse footprint of GRACE are minimised when larger catchments are evaluated. For this assessment, we used CHIRPSv2 to characterise *P*, as it performed the best at the monthly scale. To characterise ETa, we selected the two best-performing ETa products because their evaluation was carried out only at the multiannual scale using aggregated ETa values over each catchment, meaning that there was higher uncertainty related to the spatio-temporal distribution of ETa in the analysis. The two figures represent the water entering the catchment as *P* (blue bars) and leaving the catchment as ETa (green and orange bars for PMLv2 and WaPORv2.1, respectively) and *Q* (blue-grey bars). Two time series representing the change in storage are included as lines: the derived monthly change in TWS from the *P*, ETa and *Q* components (ΔTWS_{WB}, in red); and the derived monthly change in TWS from the arithmetic mean of the three official Level-3 GRACE products (ΔTWS_{GRACE}, in grey).

These results illustrate how the derived monthly ΔTWS using the water balance and the best-performing *P* and ETa products corresponds well to the monthly ΔTWS estimates from GRACE over the White Nile. However, this agreement is lost, to some extent, over the Blue Nile. Furthermore, these results suggest that the seasonality of ΔTWS is better captured through the combination of CHIRPSv2 and PMLv2 than the combination of CHIRPSv2 and WaPORv2.1.

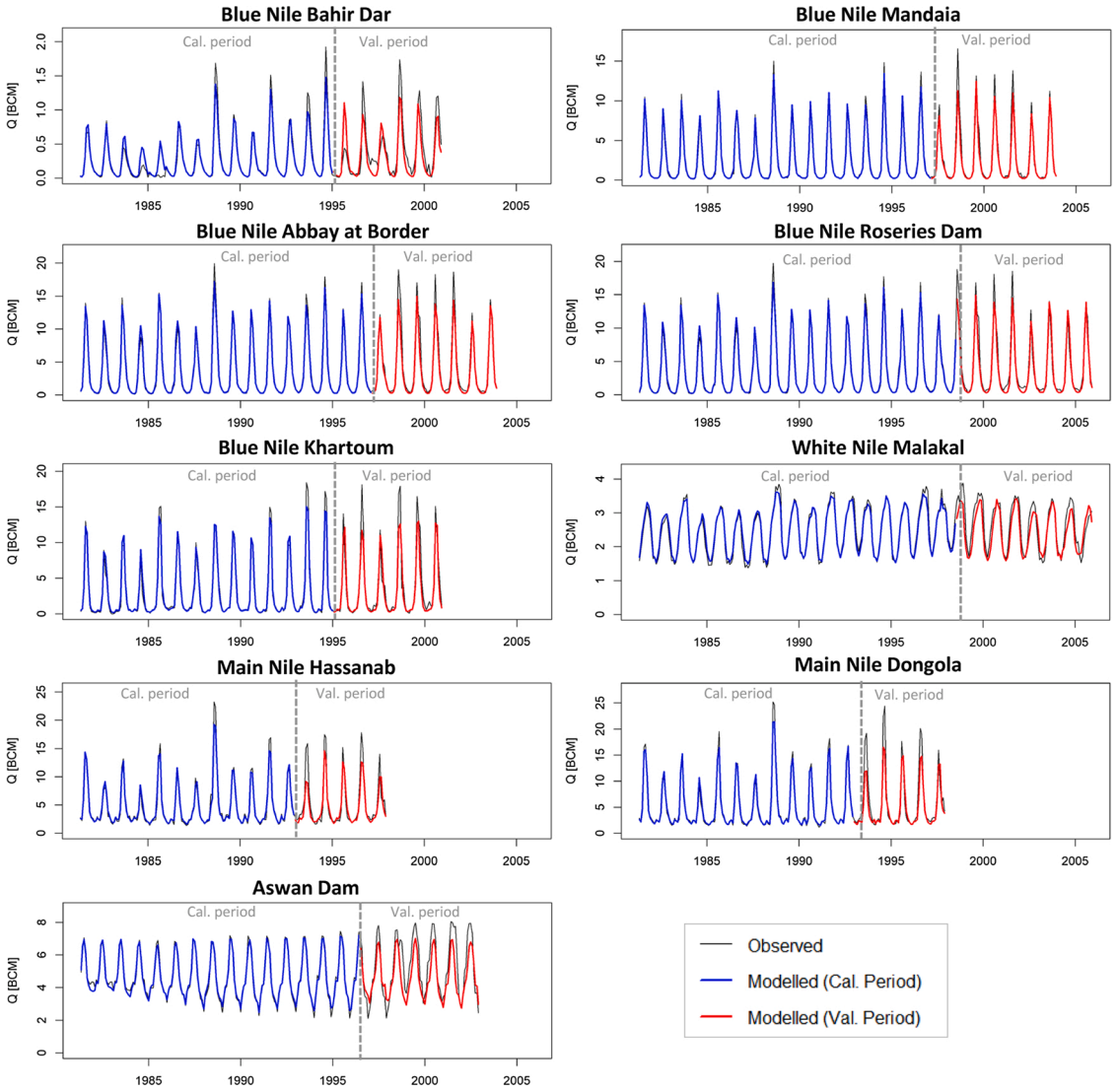

Fig. 5. Hydrographs showing the performance of the RF *Q* model for the nine selected catchments. For each catchment, the calibration period was selected as the first 70% of available data, leaving the remaining 30% for validation.

5.4.2. Annual trends in terrestrial water storage

The annual trends in ΔTWS according to the arithmetic mean of the three GRACE solutions were calculated for the eight of the nine evaluated catchments (see Fig. 5) that have an area greater than 100,000 km^2. For the four catchments in the Blue Nile Basin, mean annual ΔTWS ranged from (positive) 4.6 to 5.1 mm/yr; for White Nile Malakal, the value was 10.0 mm/yr; and for the three catchments in the Main Nile, the values ranged from 6.2 to 8.4 mm/yr. There was a consistent trend of a slight increase in TWS throughout the evaluated catchments over the period 01/2009–07/2017, but the magnitudes of these increases are small compared to the *P*, ETa and *Q* components of the water balance. Because these observed changes are small and because the GRACE data were not available over the entire period of ETa product evaluation, the storage component was not considered in the evaluation of the ETa products at the multiannual temporal scale.

6. Discussion

6.1. On the performance of P products

CHIRPSv2 had the best performance of all the evaluated *P* products, which is in agreement with numerous studies that evaluated *P* products over regions within the Nile Basin (Bayissa et al., 2017; Ayehu et al., 2018; Dinku et al., 2018; Nashwan et al., 2020). The two solely gauge-based products (GPCCv2020 and CRU TS4.04) also performed quite well (second and third best performance), although

	Catchment Name and Area (10^3 km^2)	ETa_{WB}	ERA5	GLDASv2.1	GLEAMv3.5b	MOD16	PMLv2	SSEBop	WaPORv2.1
BN	Blue Nile Bahir Dar (15.1)	1048.6	**1126 (77)**	884 (-165)	849 (-199)	441 (-608)	1311 (262)	695 (-353)	958 (-91)
BN	Blue Nile Mandaia (139.0)	1064.8	867 (-198)	855 (-210)	798 (-266)	588 (-477)	**1032 (-33)**	605 (-460)	961 (-104)
BN	Blue Nile Abbay at Border (174.4)	1043.1	873 (-171)	860 (-183)	803 (-240)	626 (-417)	**1018 (-26)**	668 (-375)	1004 (-39)
BN	Blue Nile Roseries Dam (188.4)	1037.0	866 (-171)	858 (-179)	791 (-246)	617 (-420)	**1009 (-28)**	672 (-365)	1004 (-33)
BN	Blue Nile Khartoum (307.1)	898.3	719 (-179)	752 (-147)	651 (-248)	493 (-406)	857 (-41)	618 (-280)	**882 (-17)**
WN	White Nile Malakal (1411.6)	985.6	878 (-108)	826 (-159)	768 (-217)	653 (-332)	**967 (-18)**	902 (-83)	1096 (111)
MN	Main Nile Hassanab (2012.2)	906.1	798 (-108)	760 (-146)	689 (-217)	579 (-327)	**883 (-23)**	786 (-120)	997 (91)
MN	Main Nile Dongola (2643.8)	741.2	644 (-97)	619 (-122)	556 (-185)	547 (-194)	**719 (-22)**	638 (-104)	812 (71)
MN	Aswan Dam (2915.0)	669.9	588 (-82)	559 (-111)	505 (165)	546 (-124)	**655 (-15)**	578 (-92)	741 (71)

BN = Blue Nile; WN = White Nile; MN = Main Nile

Fig. 6. ETa product performance at the multiannual temporal scale for 2009–2018: ETa_{WB} refers to the multiannual mean ETa derived from the water balance (in mm); mean annual values are reported for each ETa product; and the numbers in brackets represent the difference between ETa for each product and ETa_{WB}. The colour scale ranks the products from best-performing (dark green with bold text) to worst-performing (light green).

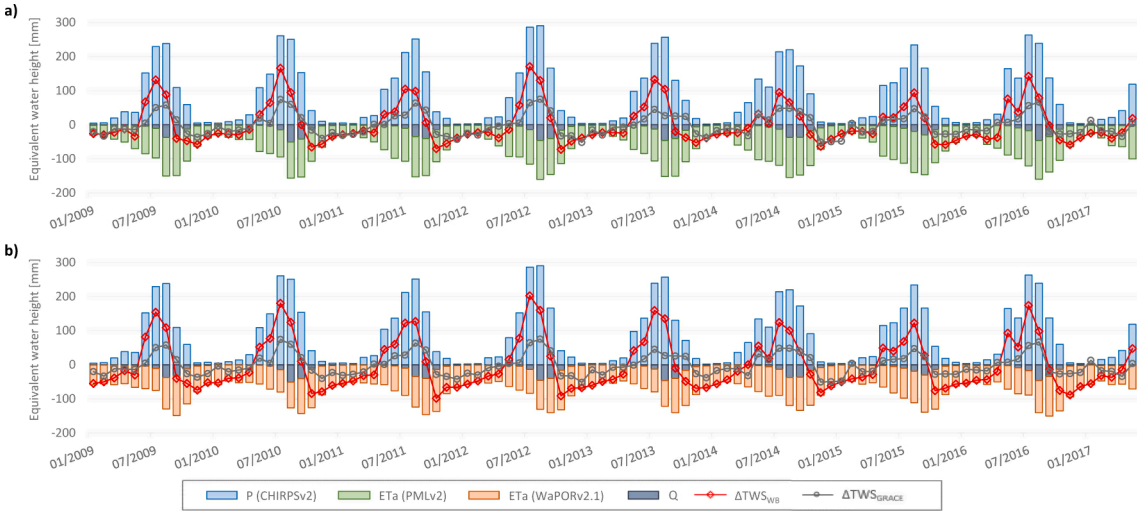

Fig. 7. Components of the water balance at the monthly scale for the Blue Nile Khartoum catchment using the: (a) PMLv2; and (b) WaPORv2.1 ETa datasets. Two estimates for monthly change in TWS are shown: that derived from applying the water balance (ΔTWS_{WB}); and that derived from the arithmetic mean of Level-3 GRACE solutions (ΔTWS_{GRACE}).

this could be attributed to the possible inclusion of regional gauges by these products. In addition, there was a relatively good performance exhibited by all satellite and gauge products, with the exception of ARC2 and CMORPHv1. ERA5, the only reanalysis product assessed, performed poorly, which is expected of reanalysis products over the tropics (Beck et al., 2017).

Although the β component of the KGE' showed that CHIRPSv2 is almost unbiased when compared to the available ground-based stations (median value of $\beta = 1.00$ and low dispersion; see Fig. 3), these stations are typically not placed at high elevations, thus making

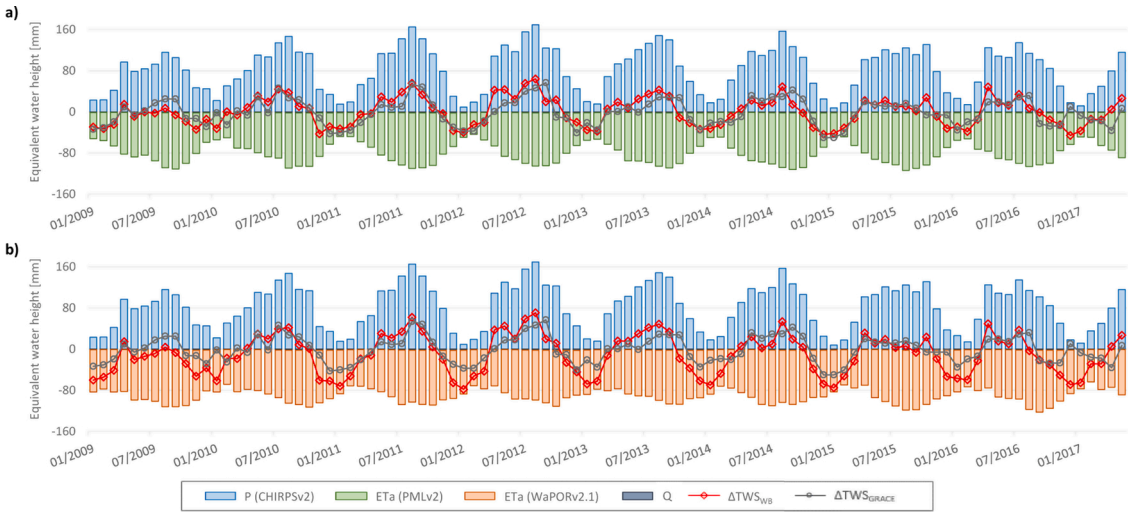

Fig. 8. Components of the water balance at the monthly scale for the White Nile Malakal catchment using the: (a) PMLv2; and (b) WaPORv2.1 ETa datasets. Two estimates for monthly change in TWS are shown: that derived from applying the water balance (ΔTWS_{WB}); and that derived from the arithmetic mean of Level-3 GRACE solutions (ΔTWS_{GRACE}).

it difficult to accurately represent the gradient of *P* related to elevation (Derin and Yilmaz, 2014). Studies have shown that satellite-based *P* products often underestimate *P* over high elevations (Beck et al., 2020), which is evident over the Ethiopian highlands (Romilly and Gebremichael, 2011; Thiemig et al., 2012; Bayissa et al., 2017). Dinku et al. (2018) evaluated the CHIRPSv2 dataset against approximately 500 rain gauge stations in Ethiopia, where a substantial portion of *Q* in the Nile Basin is generated, finding that this product underestimated the total volume of *P* (bias equal to 0.94). Similar results were obtained by Abebe et al. (2020), who reported CHIRPSv2 median bias ratios of ~0.95 for elevations >2500 m a.s.l.

Concerning product performance over the arid north of the Nile Basin, Nashwan et al. (2020) concluded that the performances of various *P* products over Egypt are relatively poor, which is in agreement with the reported lower performance of *P* products over arid areas (Zambrano-Bigiarini et al., 2017; Baez-Villanueva et al., 2020). The network of rain gauges used in this study (Fig. 4, after excluding FAO and GHCN stations) does not include stations over the northern arid region of the Nile Basin.

6.2. *On the application of the water balance to evaluate ETa products*

The RF model used to estimate *Q* performed well over the data-scarce Nile Basin, with nine of the 23 modelled catchments fulfilling the designated criteria for good performance (KGE' > 0.75 and $0.80 < \beta < 1.20$, see Table 4). However, there was a slight underestimation of the total volume of simulated *Q* due to an inability of the model to accurately capture extreme *Q* values (see Fig. 5). As RF models average all individual tree predictions, they are likely to underestimate the extreme values (Bachmair et al., 2017; Schoppa et al., 2020). This is consistent with the results reported by Shortridge et al. (2016), where RF simulated *Q* was underestimated during the wet-season and overestimated during the dry-season. Despite this, the results of our study show that the effect of this bias on the total *Q* volume modelled is small compared to the volumes of water represented by *P* and ETa.

After evaluating *P* and *Q*, we were able to evaluate ETa using the simplified water balance, and concluded that both PMLv2 and WaPORv2.1 are the best-performing products. Our results are also in agreement with those of Trambauer et al. (2014) and Weerasinghe et al. (2020), who note the large differences in total ETa estimates over Africa according to the different products. The systematic underestimation of the majority of ETa products (ERA5, GLDASv2.1, GLEAMv3.5b, MOD16 and SSEBop; see Fig. 6) is consistent with the results of Weerasinghe et al. (2020), who evaluated nine ETa products (including GLEAMv3.2a, MOD16, SSEBop and WaPORv1.1) using the water balance method and found that all products underestimated ETa over the Blue Nile Basin, while all products except CMRSET and WaPORv1.1 underestimated ETa over the entire Nile Basin. The large variation in product estimates is not surprising, given that the ETa products we assessed are derived from a range of calculation methods of varying complexity (see Section 3.2.2). Furthermore, the estimation of ETa in these products relies on accurate representations of numerous input variables (Fisher et al., 2017), with the estimation of the transpiration component identified as having the largest uncertainty (Pan et al., 2020).

The accurate representation of evaporation over water bodies is important for water balance calculations, and four of the seven evaluated ETa products (GLDASv2.1, GLEAMv3.5b, MOD16 and PMLv2) do not provide ETa estimates over large water bodies. The largest water bodies in the Nile are located in the upstream areas of the White Nile (see Fig. 1). In the calculation of mean ETa over each catchment, we excluded cells without ETa estimates. Considering the two-best performing ETa products over the Nile Basin according to our analysis, PMLv2 does not provide estimates over large water bodies and WaPORv2.1 is considered to underestimate evaporation from water bodies (FAO and IHE Delft, 2020).

In the Nile Basin, the upstream areas correspond to tropical and temperate climates, while the downstream areas are classified as

arid (see Fig. 1). Although the nine catchments assessed in our study extend over multiple climate zones, only a small contribution to total *P* and ETa comes from the arid climate areas. Therefore, our evaluation of ETa products provides low informative value about product performance over the arid climate. Over arid areas, Blatchford et al. (2020) compared WaPORv2.1 estimates to those derived from eddy covariance towers at three locations with irrigated agriculture in Egypt, finding that WaPORv2.1 overestimates ETa over these areas. Furthermore, Ayyad et al. (2019) compared estimates from ETa products with measurements from a lysimeter system over irrigated areas in Egypt, and found that SSEBop performed better than MOD16. To the best of our knowledge, no published studies have evaluated the performance of PMLv2 over the Nile Basin.

6.3. Ability of P and ETa products to represent monthly changes in TWS

The monthly ΔTWS derived from applying the water balance (Eq. (1)) using the best-performing *P* and ETa products corresponded well to the monthly ΔTWS estimates from GRACE over the White Nile (Fig. 8). In contrast, the seasonality of the ΔTWS derived from the water balance was more pronounced than that of GRACE over the Blue Nile (Fig. 7). Over the Blue Nile Basin, Abebe et al. (2020) showed that there is no substantial seasonal bias in CHIRPSv2 estimates, based on an evaluation over 38 rain gauges in the Upper Blue Nile Basin. Furthermore, the validation of our RF model showed that the seasonality of *Q* was well captured over this basin. Assuming that the seasonality of *P* and *Q* is captured well by CHIRPSv2 and the modelled *Q*, respectively, the discrepancy between the derived monthly ΔTWS_{WB} and ΔTWS_{GRACE} should be mainly due to the performance of the ETa products, or the performance of GRACE (related to measurement and leakage errors, as noted by Landerer and Swenson (2012)).

Our results suggest that PMLv2 and WaPORv2.1 ETa products tend to underestimate the seasonal variability over the Blue Nile Basin. These results suggest that although some ETa products perform relatively well at multiannual temporal scales, they are not yet able to capture the intra-annual variability over this area, which is in agreement with the reported marked seasonal differences between ETa products and calculation methods (Mueller et al., 2011; Trambauer et al., 2014; Pan et al., 2020). In the case of WaPORv2.1, a reason for this could be attributed to the fact that the Blue Nile has large areas of rainfed agriculture, for which Blatchford et al. (2020) concluded that the product overestimates and underestimates low and high ETa values, respectively. Additionally, Abera et al. (2017) obtained a very good agreement between the ΔTWS resulting from a hydrological model forced with a specific *P* product and calibrated solely with *Q* data, and ΔTWS of GRACE over the Upper Blue Nile. Their results indicate that two ETa products (GLEAMv3_BETA and MOD16) underestimate the seasonal variability of ETa over this region when compared to the resulting modelled ETa, reinforcing the idea that ETa products do not capture the intra-annual variability of ETa over the Blue Nile Basin.

6.4. Role of P and ETa product selection in establishing the water balance for the Nile Basin

Considering the major water balance components over the entire Nile Basin for 2009–2018 based on the identified best-performing products, the mean annual *P* according to CHIRPSv2 is 636 mm/yr, while the mean annual ETa according to PMLv2 and WaPORv2.1 is 616 mm/yr and 698 mm/yr, respectively. According to Molle et al. (2018), the discharge of the Nile into the Mediterranean Sea is approximately 14 BCM/yr (calculated for 2000–2013), which is equivalent to 4.4 mm/yr over the entire basin. Literature indicates that components such as groundwater discharge to the Mediterranean Sea and seawater extractions for desalination are negligible in the context of the water balance over the entire Nile Basin (Nofal et al., 2015; El-Rawy et al., 2020). Therefore, if using the *P* product and one of the ETa products recommended in this paper, the derived *Q* discharging from the Nile will either be too high (PMLv2) or negative (WaPORv2.1). Furthermore, with PMLv2, the total volumes will be affected by the missing ETa values over large water bodies.

Fig. 9 shows the spatial distribution of the estimates from the evaluated *P* and ETa products. The upper panels plot the mean annual *P* and ETa for 2009–2018 for the best-performing products, while the lower panels plot the spatial distribution of the coefficient of variation (CV) calculated for all evaluated products. As expected, the CV for the *P* products is highest over the arid regions, where Awange et al. (2016) reported a low correlation between estimates from different products, and because of the very low annual *P*, slight changes between these products result in a high CV. Values of CV for *P* are also relatively high over regions of the basin where most of the *Q* is generated, namely the Ethiopian highlands and some upstream areas in the White Nile. The CV for the ETa products is also higher over the arid regions where annual accumulated ETa values are low, which is in agreement with the results reported by Trambauer et al. (2014) and Pan et al. (2020). In addition, the CV is also high over large irrigation schemes in the arid climates, with a key reason for this being that neither GLDASv2.1 nor GLEAMv3.5b appear to detect the increased ETa associated with these irrigated croplands. The water bodies and Sudd wetlands are the other areas with a relatively high CV for ETa.

7. Conclusions

We evaluated the performance of eleven gridded *P* products through a point-to-pixel analysis, and seven gridded ETa products through the water balance over the data-scarce Nile Basin. For the water balance analysis, we generated time series of simulated *Q* through a machine learning approach and used data from Level-3 GRACE products to evaluate the changes in TWS. We evaluated the gridded *P* and ETa products using ground-based data from 90 rain gauges and 23 *Q* stations, and made the following conclusions:

(a) Based on a point-to-pixel analysis and using the KGE' and its components, CHIRPSv2 was the best-performing *P* product over the Nile Basin at the monthly scale (median KGE' of 0.80). Furthermore, CHIRPSv2 had a median bias of 1.00 and showed a much lower dispersion than all other evaluated *P* products;

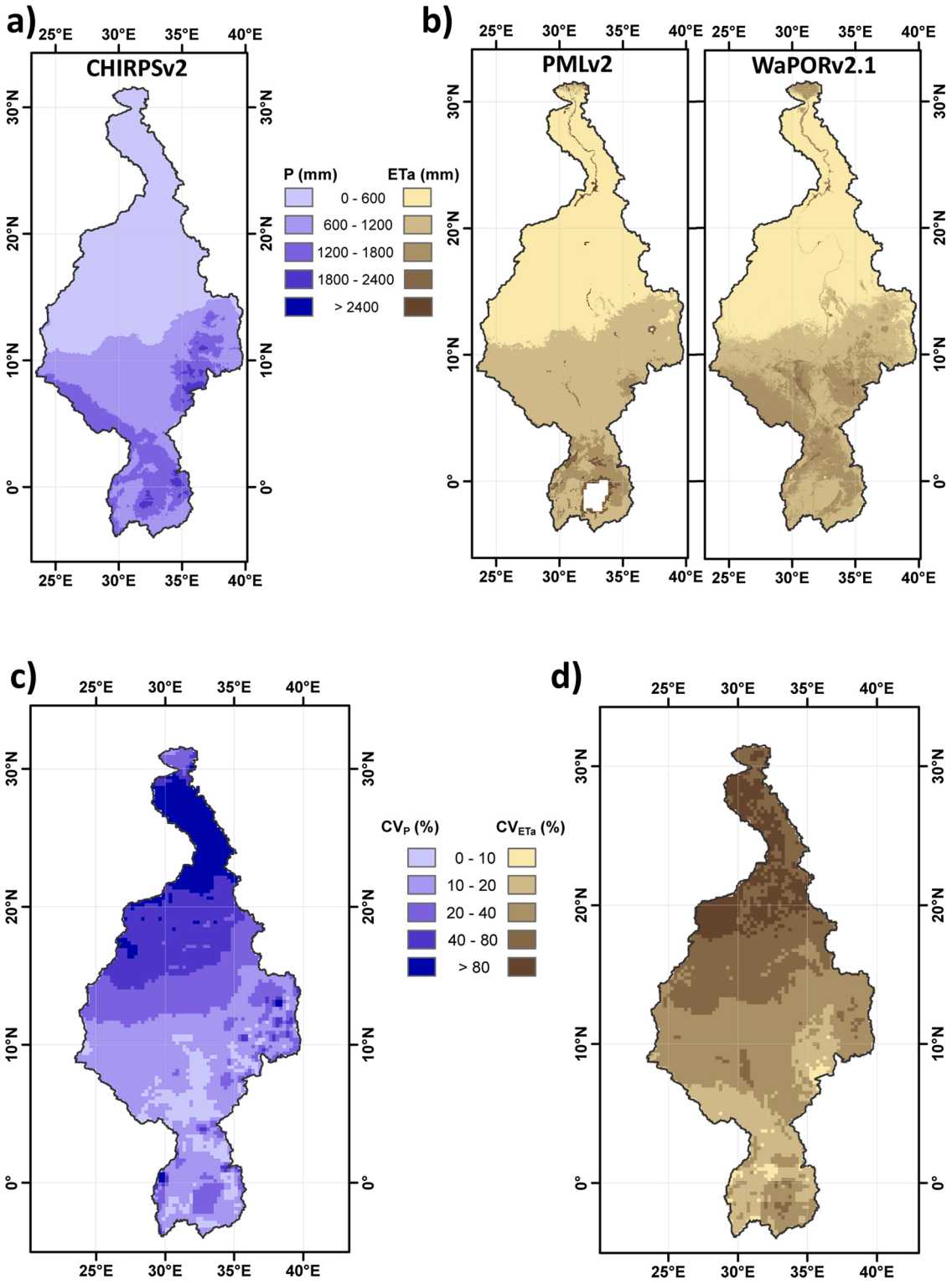

Fig. 9. Top panels: mean annual estimates (2009–2018) of *P* and ETa according to the best-performing *P* and ETa products (a and b, respectively). Bottom panels: Spatial distribution of the coefficient of variation (CV) of mean annual estimates of *P* and ETa for the evaluated products (c and d, respectively).

(b) The RF model used to simulate monthly *Q* performed well for nine of the 23 evaluated catchments and was able to capture the seasonality and magnitude of *Q*. Model performance was generally better for catchments in the Blue Nile and downstream in the Main Nile;
(c) The multiannual water balance was evaluated over nine catchments for 2009–2018 with CHIRPSv2 as the *P* input and with simulated *Q*. PMLv2 and WaPORv2.1 were the best-performing products over the majority of the catchments;
(d) Over two catchments, monthly ΔTWS calculated from the water balance using the best-performing *P* and two best-performing ETa products was compared with ΔTWS derived from GRACE Level-3 products. The seasonality of TWS derived from the best-performing *P* and ETa products is well captured over the White Nile Basin but not as well captured over the Blue Nile Basin; and
(e) Based on the results of our study, we recommend the use of CHIRPSv2 and either PMLv2 or WaPORv2.1 over the Nile Basin.

Sheffield et al. (2018) identified the pivotal role played by satellite-based products in water resources management in data-scarce regions, and our paper answers this call by providing recommendations of the best-performing *P* and ETa products over the Nile Basin. The Nile Basin has low data availability and minimal data sharing, hence this study provides an important step in the improved understanding of the spatial and temporal patterns of *P* and ETa in the region. Furthermore, the study uses novel techniques that can be transferred to other data-scarce basins around the world. Greater data availability, such as a higher density of rain gauge data, point-based ETa data and updated *Q* data would also serve a great purpose in enabling a more comprehensive evaluation of the best-performing products.

Author contributions

Ian McNamara: conceptualisation, methodology, analysis, writing, review & editing; Oscar M. Baez-Villanueva: conceptualisation, methodology, analysis, writing, review & editing; Ali Zomorodian: analysis, writing; Saher Ayyad: conceptualisation, review & editing; Mauricio Zambrano-Bigiarini: methodology, review & editing; Modathir Zaroug: methodology, review & editing; Azeb Mersha: methodology, review & editing; Alexandra Nauditt: review & editing; Milly Mbuliro: methodology, project administration; Sowed Wamala: project administration; Lars Ribbe: review & editing, funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

We would like to thank the Nile Basin Initiative (NBI) for the data provided. We would also like to thank Denis Hughes and the two anonymous reviewers for their constructive comments, which have helped to improve the quality of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ejrh.2021.100884.

References

Abd-Elbaky, M., Jin, S., 2019. Hydrological mass variations in the Nile River Basin from GRACE and hydrological models. Geodesy and Geodyn. 10, 430–438. https://doi.org/10.1016/j.geog.2019.07.004.
Abdelmoneim, H., Soliman, M.R., Moghazy, H.M., 2020. Evaluation of TRMM 3B42V7 and CHIRPS satellite precipitation products as an input for hydrological model over Eastern Nile Basin. Earth Syst. Environ. 1–14. https://doi.org/10.1007/s41748-020-00185-3.
Abebe, S.A., Qin, T., Yan, D., Gelaw, E.B., Workneh, H.T., Kun, W., Shanshan, L., Biqiong, D., 2020. Spatial and temporal evaluation of the latest high-resolution precipitation products over the Upper Blue Nile River Basin, Ethiopia. Water 12, 3072. https://doi.org/10.3390/w12113072.
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, 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, 3145–3165. https://doi.org/10.5194/hess-21-3145-2017.
Abolafia-Rosenzweig, R., Pan, M., Zeng, J., 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.
Adhikary, S.K., Yilmaz, A.G., Muttil, N., 2015. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. Hydrol. Process. 29, 2582–2599. https://doi.org/10.1002/hyp.10389.
Al Zayed, I.S., Elagib, N.A., Ribbe, L., Heinrich, J., 2016. Satellite-based evapotranspiration over Gezira irrigation scheme, Sudan: a comparative study. Agric. Water Manag. 177, 66–76. https://doi.org/10.1016/j.agwat.2016.06.027.
Alemu, H., Senay, G., Kaptue, A., Kovalskyy, V., 2014. Evapotranspiration variability and its association with vegetation dynamics in the Nile Basin, 2002–2011. Remote Sens. 6, 5885–5908. https://doi.org/10.3390/rs6075885.
Allen, R.G., Tasumi, M., Trezza, R., 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) – model. J. Irrig. Drain. Eng. 133, 380–394. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380.

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Am. Meteorol. Soc. 96, 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1.
Awange, J.L., Ferreira, V.G., Forootan, E., Andam-Akorful, S., Agutu, N.O., He, X., 2016. Uncertainties in remotely sensed precipitation data over Africa. Int. J. Climatol. 36, 303–323. https://doi.org/10.1002/joc.4346.
Awulachew, S.B., 2012. The Nile River Basin: Water, Agriculture, Governance and Livelihoods. Routledge.
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, 1921–1936. https://doi.org/10.5194/amt-11-1921-2018.
Ayyad, S., Al Zayed, I.S., Ha, V.T.T., Ribbe, L., et al., 2019. The performance of satellite-based actual evapotranspiration products and the assessment of irrigation efficiency in Egypt. Water 11, 1913. https://doi.org/10.3390/w11091913.
Ayyad, S., Khalifa, M., 2021. Will the Eastern Nile countries be able to sustain their crop production by 2050? An outlook from water and land perspectives. Sci. Total Environ. 775, 145769. https://doi.org/10.1016/j.scitotenv.2021.145769.
Bachmair, S., Svensson, C., Prosdocimi, I., Hannaford, J., Stahl, K., 2017. Developing drought impact functions for drought risk management. Nat. Hazards Earth Syst. Sci. 17, 1947–1960. https://doi.org/10.5194/nhess-17-1947-2017.
Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Thinh, N.X., 2020. RF-MEP: a novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606. https://doi.org/10.1016/j.rse.2019.111606.
Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J.D., Thinh, N.X., 2018. Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. Atmos. Res. 213, 34–50. https://doi.org/10.1016/j.atmosres.2018.05.011.
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.
Bastiaanssen, W., Cheema, M., Immerzeel, W., Miltenburg, I., Pelgrum, H., 2012. Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. Water Resour. Res. 48 https://doi.org/10.1029/2011WR010482.
Bastiaanssen, W., Karimi, P., Rebelo, L.M., Duan, Z., Senay, G., Muthuwatte, L., Smakhtin, V., 2014. Earth observation based assessment of the water production and water consumption of Nile Basin agro-ecosystems. Remote Sens. 6, 10306–10334. https://doi.org/10.3390/rs61110306.
Bastiaanssen, W.G., Menenti, M., Feddes, R., Holtslag, A., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. formulation. J. Hydrol. 212, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4.
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. 9, 669. https://doi.org/10.3390/rs9070669.
Bayissa, Y.A., Moges, S.A., Xuan, Y., Van Andel, S.J., Maskey, S., Solomatine, D.P., Griensven, A.V., Tadesse, T., 2015. Spatio-temporal assessment of meteorological drought under the influence of varying record length: the case of Upper Blue Nile Basin, Ethiopia. Hydrol. Sci. J. 60, 1927–1942. https://doi.org/10.1080/02626667.2015.1032291.
Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21, 6201–6217. https://doi.org/10.5194/hess-21-6201-2017.
Beck, H.E., Wood, E.F., McVicar, T.R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O.M., Sheffield, J., Karger, D.N., 2020. Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. J. Clim. 33, 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1.
Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100, 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1.
Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5, 180214. https://doi.org/10.1038/sdata.2018.214.
Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. Earth Syst. Sci. Data 5, 71–99. https://doi.org/10.5194/essd-5-71-2013.
Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197. https://doi.org/10.1007/s11749-016-0481-7.
Bivand, R., 2020. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R Package Version 1.5-12. https://github.com/cran/rgdal.
Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H., Zeng, Y., Pelgrum, H., Wonink, S., Karimi, P., 2020. Evaluation of WaPOR V2 evapotranspiration products across Africa. Hydrol. Process. 34, 3200–3221. https://doi.org/10.1002/hyp.13791.
Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Camberlin, P., Barraud, G., Bigot, S., Dewitte, O., Makanzu Imwangana, F., Maki Mateso, J.C., Martiny, N., Monsieurs, E., Moron, V., Pellarin, T., Philippon, N., Sahani, M., Samba, G., 2019. Evaluation of remotely sensed rainfall products over Central Africa. Q. J. R. Meteorol. Soc. 145, 2115–2138. https://doi.org/10.1002/qj.3547.
Cooley, S., Landerer, F., 2020. Gravity Recovery and Climate Experiment Follow-on (GRACE-FO) Level-3 Data Product User Handbook.
Criss, R.E., Winston, W.E., 2008. Do Nash values have value? Discussion and alternate proposals. Hydrol. Process. 22, 2723–2725. https://doi.org/10.1002/hyp.7072.
Dahle, C., Murböck, M., Flechtner, F., Dobslaw, H., Michalak, G., Neumayer, K.H., Abrykosov, O., Reinhold, A., König, R., Sulzbach, R., et al., 2019. The GFZ GRACE RL06 monthly gravity field time series: processing details and quality assessment. Remote Sens. 11, 2116. https://doi.org/10.3390/rs11182116.
Derin, Y., Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. J. Hydrometeorol. 15, 1498–1516. https://doi.org/10.1175/JHM-D-13-0191.1.
Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinform. 7, 3. https://doi.org/10.1186/1471-2105-7-3.
Diem, J.E., Hartter, J., Ryan, S.J., Palace, M.W., 2014. Validation of satellite rainfall products for western Uganda. J. Hydrometeorol. 15, 2030–2038. https://doi.org/10.1175/JHM-D-13-0193.1.
Digna, R.F., Mohamed, Y., van der Zaag, P., Uhlenbrook, S., Corzo, G., 2017. Nile River Basin modelling for water resources management – a literature review. Int. J. River Basin Manag. 15, 39–52. https://doi.org/10.1080/15715124.2016.1228656.
Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S., Ropelewski, C., 2008a. Validation of high-resolution satellite rainfall products over complex terrain. Int. J. Remote Sens. 29, 4097–4110. https://doi.org/10.1080/01431160701772526.
Dinku, T., Connor, S.J., Ceccato, P., Ropelewski, C.F., 2008b. Comparison of global gridded precipitation products over a mountainous region of Africa. Int. J. Climatol. J. R. Meteorol. Soc. 28, 1627–1638. https://doi.org/10.1002/joc.1669.
Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., Ceccato, P., 2018. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Q. J. R. Meteorol. Soc. 144, 292–312. https://doi.org/10.1002/qj.3244.
El-Rawy, M., Abdalla, F., El Alfy, M., 2020. Water resources in Egypt. In: Hamimi, Z., El-Barkooky, A., Martínez Frías, J., Fritz, H., Abd El-Rahman, Y. (Eds.), The Geology of Egypt. Springer, Cham, pp. 687–711. https://doi.org/10.1007/978-3-030-15265-9_18.
FAO, 2020. WaPOR Database Methodology. Version 2 Release, April 2020. FAO, Rome. https://doi.org/10.4060/ca9894en.
FAO, IHE Delft, 2020. Water Accounting in the Nile River Basin. FAO WaPOR Water Accounting Reports. FAO, Rome. https://doi.org/10.4060/ca9895en.
Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., et al., 2017. The future of evapotranspiration: global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resour. Res. 53, 2618–2626. https://doi.org/10.1002/2016WR020175.
Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919. https://doi.org/10.1016/j.rse.2007.06.025.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. Sci. Data 2, 150066. https://doi.org/10.1038/sdata.2015.66.
Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H., Yu, Q., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. Ecohydrology 11, e1974. https://doi.org/10.1002/eco.1974.
Garcia, M., Peters-Lidard, C.D., Goodrich, D.C., 2008. Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. Water Resour. Res. 44 https://doi.org/10.1029/2006WR005788.
Gehne, M., Hamill, T.M., Kiladis, G.N., Trenberth, K.E., 2016. Comparison of global precipitation estimates across a range of temporal and spatial scales. J. Clim. 29, 7773–7795. https://doi.org/10.1175/JCLI-D-15-0618.1.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.
Haiden, T., Rodwell, M.J., Richardson, D.S., Okagaki, A., Robinson, T., Hewson, T., 2012. Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. Mon. Weather Rev. 140, 2720–2733. https://doi.org/10.1175/MWR-D-11-00301.1.
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, 1–18. https://doi.org/10.1038/s41597-020-0453-3.
Hasan, E., Tarhule, A., Hong, Y., Moore, B., 2019. Assessment of physical water scarcity in Africa using GRACE and TRMM satellite data. Remote Sens. 11, 904. https://doi.org/10.3390/rs11080904.
Hasan, E., Tarhule, A., Kirstetter, P.E., 2021. Twentieth and twenty-first century water storage changes in the Nile River Basin from GRACE/GRACE-FO and modeling. Remote Sens. 13, 953. https://doi.org/10.3390/rs13050953.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., Gräler, B., 2018. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.
Herold, N., Behrangi, A., Alexander, L.V., 2017. Large uncertainties in observed daily precipitation extremes over land. J. Geophys. Res. Atmos. 122, 668–681. https://doi.org/10.1002/2016JD025842.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Mu noz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. https://doi.org/10.1002/qj.3803.
Hijmans, R.J., 2020. raster: Geographic Data Analysis and Modeling. R Package Version 3.3-13. https://CRAN.R-project.org/package=raster.
Hsu, K., Gao, l., Sorooshian, X., Gupta, S.H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteorol. 36, 1176–1190. https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.
Huffman, G., Adler, R., Bolvin, D., Nelkin, E., 2010. The TRMM multi-satellite precipitation analysis (TMPA). In: Gebremichael, M., Hossain, F. (Eds.), Satellite Rainfall Applications for Surface Hydrology. Springer Verlag. ISBN 978-90-481-2914-0.
Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., Tan, J., 2019. GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06, edited by Andrey Savtchenko. Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD.
Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., Yoo, S.H., 2015. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 4, p. 26.
Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38. https://doi.org/10.1175/JHM560.1.
Jarvis, A., Guevara, E., Reuter, H.I., Nelson, A., 2008. Hole-Filled SRTM for the Globe Version 4. Available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org) 15, 25–54.
Johnston, R., Smakhtin, V., 2014. Hydrological modeling of large river basins: how much is enough? Water Resour. Manag. 28, 2695–2730. https://doi.org/10.1007/s11269-014-0637-8.
Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5, 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.
Karimi, P., Bastiaanssen, W.G., 2015. Spatial evapotranspiration, rainfall and land use data in water accounting – part 1: review of the accuracy of the remote sensing data. Hydrol. Earth Syst. Sci. 19, 507–532. https://doi.org/10.5194/hess-19-507-2015.
Khan, M.S., Liaqat, U.W., Baik, J., Choi, M., 2018. Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. Agric. For. Meteorol. 252, 256–268. https://doi.org/10.1016/j.agrformet.2018.01.022.
Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011.
Konapala, G., Mishra, A., 2020. Quantifying climate and catchment control on hydrological drought in the continental United States. Water Resour. Res. 56 https://doi.org/10.1029/2018WR024620.
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, 935–952. https://doi.org/10.1175/JHM-D-19-0233.1.
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., Swenson, S., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011453.
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 https://doi.org/10.1029/2007WR006562.
Levizzani, V., Schmetz, J., Lutz, H., Kerkmann, J., Alberoni, P., Cervino, M., 2001. Precipitation estimations from geostationary orbit and prospects for METEOS Second Generation. Meteorol. Appl. 8, 23–41. https://doi.org/10.1017/S1350482701001037.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22. https://CRAN.R-project.org/doc/Rnews/.
Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., Li, X., Sang, Y.F., 2016. A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. J. Hydrol. 538, 82–95. https://doi.org/10.1016/j.jhydrol.2016.04.006.
Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50, 1131–1151. https://doi.org/10.1002/2013WR014581.
Maggioni, V., Massari, C., 2018. On the performance of satellite precipitation products in riverine flood modeling: a review. J. Hydrol. 558, 214–224. https://doi.org/10.1016/j.jhydrol.2018.01.039.
Maidment, R.I., Grimes, D., Allan, R.P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R., Black, E., 2014. The 30 year TAMS, African rainfall climatology and time series (TARCAT) data set. J. Geophys. Res. Atmos. 119, 10–619. https://doi.org/10.1002/2014JD021927.
Maidment, R.I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., Allan, R.P., Stein, T., Nkonde, E., Senkunda, S., et al., 2017. A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. Sci. Data 4, 170063. https://doi.org/10.1038/sdata.2017.63.
Martens, B., Gonzalez Miralles, D., Lievens, H., Van Der Schalie, R., De Jeu, R.A., Fernández-Prieto, D., Beck, H.E., Dorigo, W., Verhoest, N., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.
Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., Verhoest, N.E., 2016. Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. Int. J. Appl. Earth Obs. Geoinf. 48, 146–162. https://doi.org/10.1016/j.jag.2015.09.012.
Massari, C., Crow, W., Brocca, L., 2017. An assessment of the performance of global rainfall estimates without ground-based observations. Hydrol. Earth Syst. Sci. 21, 4347. https://doi.org/10.5194/hess-21-4347-2017.
McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., Wood, E.F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geosci. Model Dev. 9, 283–305. https://doi.org/10.5194/gmd-9-283-2016.
Miralles, D.G., Brutsaert, W., Dolman, A., Gash, J.H., 2020. On the use of the term “evapotranspiration”. Water Resour. Res. 56 https://doi.org/10.1029/2020WR028055 e2020WR028055.

Miralles, D.G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., Dolman, A., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15, 453–469. https://doi.org/10.5194/hess-15-453-2011.

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., et al., 2016. The WACMOS-ET project – part 2: evaluation of global terrestrial evaporation data sets. Hydrol. Earth Syst. Sci. 20, 823–842. https://doi.org/10.5194/hess-20-823-2016.

Mohamed, Y., Savenije, H., Bastiaanssen, W., Hurk, B., 2006. New lessons on the Sudd hydrology learned from remote sensing and climate modeling. Hydrol. Earth Syst. Sci. 10, 507–518. https://doi.org/10.5194/hess-10-507-2006.

Molle, F., Gaafar, I., El-Agha, D.E., Rap, E., 2018. The Nile delta's water and salt balances and implications for management. Agric. Water Manag. 197, 110–121. https://doi.org/10.1016/j.agwat.2017.11.016.

Moreira, A.A., Ruhoff, A.L., Roberti, D.R., de Arruda Souza, V., da Rocha, H.R., de Paiva, R.C.D., 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, 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, 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.

Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J., Guo, Z., et al., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophys. Res. Lett. 38 https://doi.org/10.1029/2010GL046230.

Nashwan, M.S., Shahid, S., Dewan, A., Ismail, T., Alias, N., 2020. Performance of five high resolution satellite-based precipitation products in arid region of Egypt: an evaluation. Atmos. Res. 236, 104809. https://doi.org/10.1016/j.atmosres.2019.104809.

Nauss, T., Kokhanovsky, A.A., 2006. Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data. Atmos. Chem. Phys. 6, 5031–5036. https://doi.org/10.5194/acp-6-5031-2006.

Nofal, E., Amer, M., El-Didy, S., Fekry, A.M., 2015. Delineation and modeling of seawater intrusion into the Nile Delta Aquifer: a new perspective. Water Sci. 29, 156–166. https://doi.org/10.1016/j.wsj.2015.11.003.

Novella, N.S., Thiaw, W.M., 2013. African Rainfall Climatology version 2 for famine early warning systems. J. Appl. Meteorol. Climatol. 52, 588–606. https://doi.org/10.1175/JAMC-D-11-0238.1.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., et al., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrol. Earth Syst. Sci. 24, 1485–1509. https://doi.org/10.5194/hess-24-1485-2020.

Pelgrum, H., Miltenburg, I., Cheema, M., Klaasse, A., Bastiaanssen, W., 2010. ETLook a novel continental evapotranspiration algorithm. Remote Sensing and Hydrology Symposium. Jackson Hole, Wyoming, USA, pp. 120–123.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and Random Forests for ecological prediction. Ecosystems 9, 181. https://doi.org/10.1007/s10021-005-0054-1.

Price, J.C., 1990. Using spatial context in satellite data to infer regional scale evapotranspiration. IEEE Trans. Geosci. Remote Sens. 28, 940–948. https://doi.org/10.1109/36.58983.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rodell, M., Famiglietti, J., 1999. Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. Water Resour. Res. 35, 2705–2723. https://doi.org/10.1029/1999WR900141.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., et al., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394. https://doi.org/10.1175/BAMS-85-3-381.

Romilly, T.G., Gebremichael, M., 2011. Evaluation of satellite rainfall estimates over Ethiopian river basins. Hydrol. Earth Syst. Sci. 15, 1505–1514. https://doi.org/10.5194/hess-15-1505-2011.

Rui, H., Beaudoing, H., Loeser, C., 2018. README Document for NASA GLDAS Version 2 Data Products. Technical Report. Goddart Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA.

Running, S.W., Mu, Q., Zhao, M., Moreno, A., 2019. MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3 and Year-end Gap-filled MOD16A2GF/A3GF) NASA Earth Observing System MODIS Land Algorithm (For Collection 6). Technical Report.

Ryu, Y., Baldocchi, D.D., Kobayashi, H., Van Ingen, C., Li, J., Black, T.A., Beringer, J., Van Gorsel, E., Knohl, A., Law, B.E., et al., 2011. Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. Glob. Biogeochem. Cycles 25. https://doi.org/10.1029/2011GB004053.

Sakumura, C., Bettadpur, S., Bruinsma, S., 2014. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. Geophys. Res. Lett. 41, 1389–1397. https://doi.org/10.1002/2013GL058632.

Save, H., 2020. CSR GRACE and GRACE-FO RL06 Mascon Solutions v02. https://doi.org/10.15781/cgq9-nh24.

Scanlon, B.R., Zhang, Z., Save, H., Wiese, D.N., Landerer, F.W., Long, D., Longuevergne, L., Chen, J., 2016. Global evaluation of new GRACE mascon products for hydrologic applications. Water Resour. Res. 52, 9412–9429. https://doi.org/10.1002/2016WR019494.

Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrol. Process. 21, 2075–2080. https://doi.org/10.1002/hyp.6825.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor. Appl. Climatol. 115, 15–40. https://doi.org/10.1007/s00704-013-0860-x.

Schneider, U., Becker, A., Finger, P., Rustemeier, E., Ziese, M., 2020. GPCC Full Data Monthly Product Version 2020 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges Built on GTS-Based and Historical Data. https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025.

Schneider, U., Ziese, M., Meyer-Christoffer, A., Finger, P., Rustemeier, E., Becker, A., 2016. The new portfolio of global precipitation data products of the Global Precipitation Climatology Centre suitable to assess and quantify the global water cycle and resources. Proc. Int. Assoc. Hydrol. Sci. 374, 29. https://doi.org/10.5194/piahs-374-29-2016.

Schoppa, L., Disse, M., Bachmair, S., 2020. Evaluating the performance of random forest for large-scale flood discharge simulation. J. Hydrol. 590, 125531. https://doi.org/10.1016/j.jhydrol.2020.125531.

Senay, G., Leake, S., Nagler, P., Artan, G., Dickinson, J., Cordova, J., Glenn, E., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrol. Process. 25, 4037–4049. https://doi.org/10.1002/hyp.8379.

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: a new parameterization for the SSEB approach. JAWRA J. Am. Water Resour. Assoc. 49, 577–591. https://doi.org/10.1111/jawr.12057.

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., Taylor, R.G., Jones, D., Longuevergne, L., Owor, M., Tindimugaya, C., 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, 4533–4549. https://doi.org/10.5194/hess-21-4533-2017.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K., 2018. Satellite remote sensing for water resources management: potential for supporting sustainable development in data-poor regions. Water Resour. Res. 54, 9724–9758. https://doi.org/10.1029/2017WR022437.

Shokr, M.E., 2020. Environmental applications of remote sensing in Egypt: a review and an outlook. In: Elbeih, S.F., Negm, A.M., Kostianoy, A. (Eds.), Environmental Remote Sensing in Egypt. Springer, Cham. https://doi.org/10.1007/978-3-030-39593-3_4.

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrol. Earth Syst. Sci. 20, 2611–2628. https://doi.org/10.5194/hess-20-2611-2016.

Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82, 2635–2670. https://doi.org/10.2166/wst.2020.369.

Stisen, S., Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. Hydrol. Process. 24, 879–891. https://doi.org/10.1002/hyp.7529.
Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys. 56, 79–107. https://doi.org/10.1002/2017RG000574.
Swain, A., 2011. Challenges for water sharing in the Nile Basin: changing geo-politics and changing climate. Hydrol. Sci. J. 56, 687–702. https://doi.org/10.1080/02626667.2011.577037.
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, 503–505. https://doi.org/10.1126/science.1099192.
Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R.P., Stringer, M., Chadwick, R., Kayitakire, F., 2014. Extension of the TAMS, satellite-based rainfall monitoring over Africa and from 1983 to present. J. Appl. Meteorol. Climatol. 53, 2805–2822. https://doi.org/10.1175/JAMC-D-14-0016.1.
Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A., 2013. Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. J. Hydrol. 499, 324–338. https://doi.org/10.1016/j.jhydrol.2013.07.012.
Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., De Roo, A., 2012. Validation of satellite-based precipitation products over sparsely gauged African river basins. J. Hydrometeorol. 13, 1760–1783. https://doi.org/10.1175/JHM-D-12-032.1.
Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., Van Beek, L., Uhlenbrook, S., 2014. Comparison of different evaporation estimates over the African continent. Hydrol. Earth Syst. Sci. 18, 193–212. https://doi.org/10.5194/hess-18-193-2014.
Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11, 910. https://doi.org/10.3390/w11050910.
Villarini, G., Mandapaka, P.V., Krajewski, W.F., Moore, R.J., 2008. Rainfall and sampling uncertainties: a rain gauge perspective. J. Geophys. Res. Atmos. 113. https://doi.org/10.1029/2007JD009214.
Wahr, J., Molenaar, M., Bryan, F., 1998. Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. J. Geophys. Res. Solid Earth 103, 30205–30229. https://doi.org/10.1029/98JB02844.
Wang, J., Bras, R., 2011. A model of evapotranspiration based on the theory of maximum entropy production. Water Resour. Res. 47 https://doi.org/10.1029/2010WR009392.
Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. Rev. Geophys. 50 https://doi.org/10.1029/2011RG000373.
Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., Griensven, A.v., 2020. Can we trust remote sensing evapotranspiration products over africa? Hydrol. Earth Syst. Sci. 24, 1565–1586. https://doi.org/10.5194/hess-24-1565-2020.
Wiese, D., Yuan, D., Boening, C., Landerer, F., Watkins, M., 2019. JPL GRACE and GRACE-FO Mascon Ocean, Ice, and Hydrology Equivalent Water Height Coastal Resolution Improvement (CRI) Filtered Release 06 Version 02, Dataset.
Xie, P., Joyce, R., Wu, S., Yoo, S.H., Yarosh, Y., Sun, F., Lin, R., 2017. Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. J. Hydrometeorol. 18, 1617–1641. https://doi.org/10.1175/JHM-D-16-0168.1.
Xie, P., Rudolf, B., Schneider, U., Arkin, P.A., 1996. Gauge-based monthly analysis of global land precipitation from 1971 to 1994. J. Geophys. Res. Atmos. 101, 19023–19034. https://doi.org/10.1029/96JD01553.
Yang, T., Asanjan, A.A., Welles, E., Gao, X., Sorooshian, S., Liu, X., 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour. Res. 53, 2786–2812. https://doi.org/10.1002/2017WR020482.
Yuan, D., 2018. JPL Level-2 Processing Standards Document for Level-2 Product Release 06. Jet Propulsion Laboratory, California Institute of Technology.
Zambrano-Bigiarini, M., 2020a. hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. R Package Version 0.4-0. https://doi.org/10.5281/zenodo.840087. http://hzambran.github.io/hydroGOF/.
Zambrano-Bigiarini, M., 2020b. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R Package Version 0.6-0. https://doi.org/10.5281/zenodo.83964. https://github.com/hzambran/hydroTSM.
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, 1295. 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, 1–15. https://doi.org/10.1038/s41598-019-51666-z.
Zeitoun, M., Mirumachi, N., Warner, J., Kirkegaard, M., Casc ao, A., 2020. Analysis for water conflict transformation. Water Int. 45, 365–384. https://doi.org/10.1080/02508060.2019.1607479.
Zhang, K., Kimball, J.S., Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. Wiley Interdiscip. Rev. Water 3, 834–853. https://doi.org/10.1002/wat2.1168.
Zhang, Y., Kong, D., Gan, R., Chiew, F.H., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens. Environ. 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031.
Zhu, H., Wheeler, M.C., Sobel, A.H., Hudson, D., 2014. Seamless precipitation prediction skill in the tropics and extratropics from a global model. Mon. Weather Rev. 142, 1556–1569. https://doi.org/10.1175/MWR-D-13-00222.1.