

Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies

Moctar Dembélé^{a,*}, Natalie Ceperley^{a,e}, Sander J. Zwart^b, Elga Salvadore^{c,d}, Gregoire Mariethoz^a, Bettina Schaefli^{a,e}

^a Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, 1015 Lausanne, Switzerland

^b International Water Management Institute, PMB CT 112 Cantonnements, Accra, Ghana

^c IHE Delft Institute for Water Education, Delft, 2611AX, The Netherlands

^d Department of Hydrology and Hydraulic Engineering, Faculty of Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium

^e Institute of Geography (GIUB), University of Bern, 3012 Bern, Switzerland



ARTICLE INFO

Keywords:

Actual evaporation
Satellite remote sensing
Reanalysis
Model parametrization
Hydrological processes
Spatial patterns
Multi-variable calibration
Multi-objective function

ABSTRACT

Twelve actual evaporation datasets are evaluated for their ability to improve the performance of the fully distributed mesoscale Hydrologic Model (mHM). The datasets consist of satellite-based diagnostic models (MOD16A2, SSEBop, ALEXI, CMRSET, SEBS), satellite-based prognostic models (GLEAM v3.2a, GLEAM v3.3a, GLEAM v3.2b, GLEAM v3.3b), and reanalysis (ERA5, MERRA-2, JRA-55). Four distinct multivariate calibration strategies (basin-average, pixel-wise, spatial bias-accounting and spatial bias-insensitive) using actual evaporation and streamflow are implemented, resulting in 48 scenarios whose results are compared with a benchmark model calibrated solely with streamflow data. A process-diagnostic approach is adopted to evaluate the model responses with in-situ data of streamflow and independent remotely sensed data of soil moisture from ESA-CCI and terrestrial water storage from GRACE. The method is implemented in the Volta River basin, which is a data scarce region in West Africa, for the period from 2003 to 2012.

Results show that the evaporation datasets have a good potential for improving model calibration, but this is dependent on the calibration strategy. All the multivariate calibration strategies outperform the streamflow-only calibration. The highest improvement in the overall model performance is obtained with the spatial bias-accounting strategy (+29%), followed by the spatial bias-insensitive strategy (+26%) and the pixel-wise strategy (+24%), while the basin-average strategy (+20%) gives the lowest improvement. On average, using evaporation data in addition to streamflow for model calibration decreases the model performance for streamflow (-7%), which is counterbalanced by the increase in the performance of the terrestrial water storage (+11%), temporal dynamics of soil moisture (+6%) and spatial patterns of soil moisture (+89%). In general, the top three best performing evaporation datasets are MERRA-2, GLEAM v3.3a and SSEBop, while the bottom three datasets are MOD16A2, SEBS and ERA5. However, performances of the evaporation products diverge according to model responses and across climatic zones. These findings open up avenues for improving process representation of hydrological models and advancing the spatiotemporal prediction of floods and droughts under climate and land use changes.

1. Introduction

Assessing the spatiotemporal variability of hydrological processes is the crux of effective water resource management. Global warming is expected to intensify (i.e., accelerate) the hydrological cycle, thus increasing or decreasing evaporation depending on places (Donat et al., 2016; Famiglietti and Rodell, 2013; Huntington, 2006).

Evaporation is a dominant flux of the water cycle (Martens et al., 2018; Oki and Kanae, 2006). It represents the exchange of water and energy between terrestrial ecosystems and the atmosphere. Therefore, evaporation can be used as a proxy for moisture availability and its consumption rate (He et al., 2019; Joiner et al., 2018; Van der Ent et al., 2010). The basic theories and estimation methods of evaporation are widely documented (e.g., Chen and Liu, 2020; Liou and Kar, 2014; McMahon et al., 2013; Pan et al., 2019; Zhang et al., 2016). Accurate representation of evaporation in hydrological models is important for studies that link climate and land use change because evaporation is the central flux that defines land-atmosphere interactions (Fisher et al., 2017; Mueller et al., 2011). Reliably modelling evaporation is essen-

* Corresponding author.

E-mail address: moctar.dembele@unil.ch (M. Dembélé).

List of symbols

<i>Symbols description</i>	
α	Spatial location measure
β	Bias measure
γ	Variability measure
λ	Latent heat of vaporization of water
Φ_{BA}	Objective function for case BA
Φ_{E_a}	Objective function for E_a
Φ_{PW}	Objective function for case PW
Φ_Q	Objective function for case Q
Φ_{SB}	Objective function for case SB
Φ_{SP}	Objective function for case SP
D_E	Euclidian distance
E_a	Actual evaporation
E_{KG}	Kling-Gupta efficiency
E_p	Potential evaporation
E_{ref}	Reference evaporation
E_{RMS}	Root mean squared error
E_{SP}	Spatial pattern efficiency
F_{DS}	Dynamical scaling function
I_{LA}	Leaf area index
P	Precipitation
Q	Streamflow
r	Pearson correlation coefficient
R_a	Extraterrestrial radiation
r_s	Spearman correlation coefficient
S_t	Terrestrial water storage anomaly
S_u	Soil moisture
T_{avg}	Average air temperature
T_{max}	Maximum air temperature
T_{min}	Minimum air temperature

tial in (semi-)arid catchments, where evaporation represents the dominant outgoing water flux. In such environments, a classical hydrological model calibration on streamflow alone could lead to important model biases. More generally, for catchments with strong anthropogenic influences on the water cycle (e.g., irrigation schemes, dams, etc.) or in data scarce regions, the inclusion of evaporation data in model calibration has the potential to better constrain a hydrological model rather than streamflow alone (Becker et al., 2019; Jiang and Wang, 2019). Accordingly, calibrating hydrological models solely based on streamflow is not sufficient to guarantee an accurate representation of the hydrological system because streamflow is the result of several inter-linked processes, thereby it masks spatial heterogeneity (Tobin and Bennett, 2017; Wambura et al., 2018). This limitation in model implementation has been overcome by the advent of multivariate calibration techniques using satellite-based datasets, which offers models a chance for better spatial heterogeneity (Efstratiadis and Koutsoyiannis, 2010; Rakovec et al., 2016).

In the quest to improve process representation in hydrological models, large scale distributed hydrological modelling faces the challenging requirement of spatially explicit observational datasets for model setup and performance evaluation (Clark et al., 2017; Fatici et al., 2016; Hrachowitz and Clark, 2017). Correspondingly, Satellite Remote Sensing (SRS) and reanalysis datasets of various hydrological processes have been used as input data (Beck et al., 2017; Maggioni and Massari, 2018) or as calibration and evaluation data for hydrological models (Koppa and Gebremichael, 2020; McCabe et al., 2017). Evaporation estimates from SRS are increasingly used in multivariate calibration of hydrological models. Because it is a key indicator of surface water availability, evaporation is an essential source of information for better constraining the spatiotemporal representation of processes in hydrological

models (Bai and Liu, 2018; Cui et al., 2019; Talsma et al., 2018). The increasing availability and diversity of gridded evaporation datasets has triggered many evaluation and comparison studies (Long et al., 2014; Vinukollu et al., 2011b), which highlight significant differences between the datasets and thereby indicate underlying uncertainty in the evaporation estimates (Baik et al., 2018; López et al., 2017). The uncertainty stems from the strong variability of bio-geophysical variables that drive evaporation (e.g., albedo, net radiation, surface roughness and temperature) and the diversity of the model structures, model parametrizations and input datasets used to estimate evaporation (Badgley et al., 2015; Wang and Dickinson, 2012; Zhang et al., 2020). Therefore, the choice and use of SRS evaporation data in hydrological modelling should be done cautiously, particularly in catchments with strong anthropogenic influences (Senkondo et al., 2019; Yang et al., 2016). Generally, the following four approaches are adopted to evaluate gridded evaporation products:

- (i) analysis of the variance between several products (e.g., Jimenez et al., 2011; Khan et al., 2018; Mueller et al., 2011; Senkondo et al., 2019; Trimbauer et al., 2014);
- (ii) point-to-pixel comparison with ground-based measurements (e.g., Chen et al., 2014; McCabe et al., 2015; Michel et al., 2016; Ramoelo et al., 2014; Velpuri et al., 2013);
- (iii) hydrological consistency by water balance calculation (e.g., Liu et al., 2016; McCabe et al., 2008; Miralles et al., 2016; Wang et al., 2018; Weerasinghe et al., 2020); and
- (iv) assessing the ability of evaporation datasets in improving the parameter estimation of hydrological models (e.g., Demirel et al., 2018; Immerzeel and Droogers, 2008; Jiang et al., 2020; Pomeon et al., 2018; Winsemius et al., 2008).

Assessing the uncertainty of evaporation estimates at large-scale is challenging due to the limited availability of ground-based measurements (Bhattarai et al., 2019; Ceperley et al., 2017). The uncertainty in evaporation varies in space and according to climate regions (Blatchford et al., 2020; Vinukollu et al., 2011a). In evaluating the contribution of gridded evaporation datasets to hydrological model calibration, some limitations can be denoted in previous studies. Most previous studies only use or compare few evaporation datasets (e.g., Kunnath-Poovakka et al., 2016; Vervoort et al., 2014), and rarely (if any) investigate the use of reanalysis datasets (i.e., retrospective analysis; cf. Bosilovich et al., 2008), which are an important source of spatial evaporation estimates (Feng et al., 2019). Usually, a lumped or semi-distributed model is used (e.g., Odusanya et al., 2019; Rientjes et al., 2013), which does not harness the full potential of the gridded evaporation datasets that is their spatial patterns (Armstrong et al., 2019; Stisen et al., 2018). Most studies do not test different model calibration strategies, with some exceptions that use a semi-distributed model (e.g., Herman et al., 2018; Rajib et al., 2018). Finally, a few studies use a bias-insensitive metric to focus only on the spatial patterns of gridded evaporation products (Dembélé et al., 2020a; Koch et al., 2018).

This study aims to fill current knowledge gaps by evaluating the utility of nine satellite-based (prognostic and diagnostic) and three reanalysis evaporation datasets in improving the performance of a distributed hydrological model using four distinct calibration strategies. This study does not intend to quantify the intrinsic accuracy of the evaporation products nor determine whether a product is better than the others in terms of absolute values. Rather it strives to evaluate their ability to improve the simulations of a distributed hydrological model when used as a calibration variable. Besides the high number and diversity of gridded evaporation datasets evaluated, the novelty of this study is the implementation of four distinct model calibration strategies with a fully distributed hydrological model, the evaluation of the model responses with multiple variables (i.e., streamflow, soil moisture and terrestrial water storage) to test evaporation error propagation on other hydrological processes, and the application of the experiment in a large basin spread across four eco-climatic zones with considerable anthropogenic

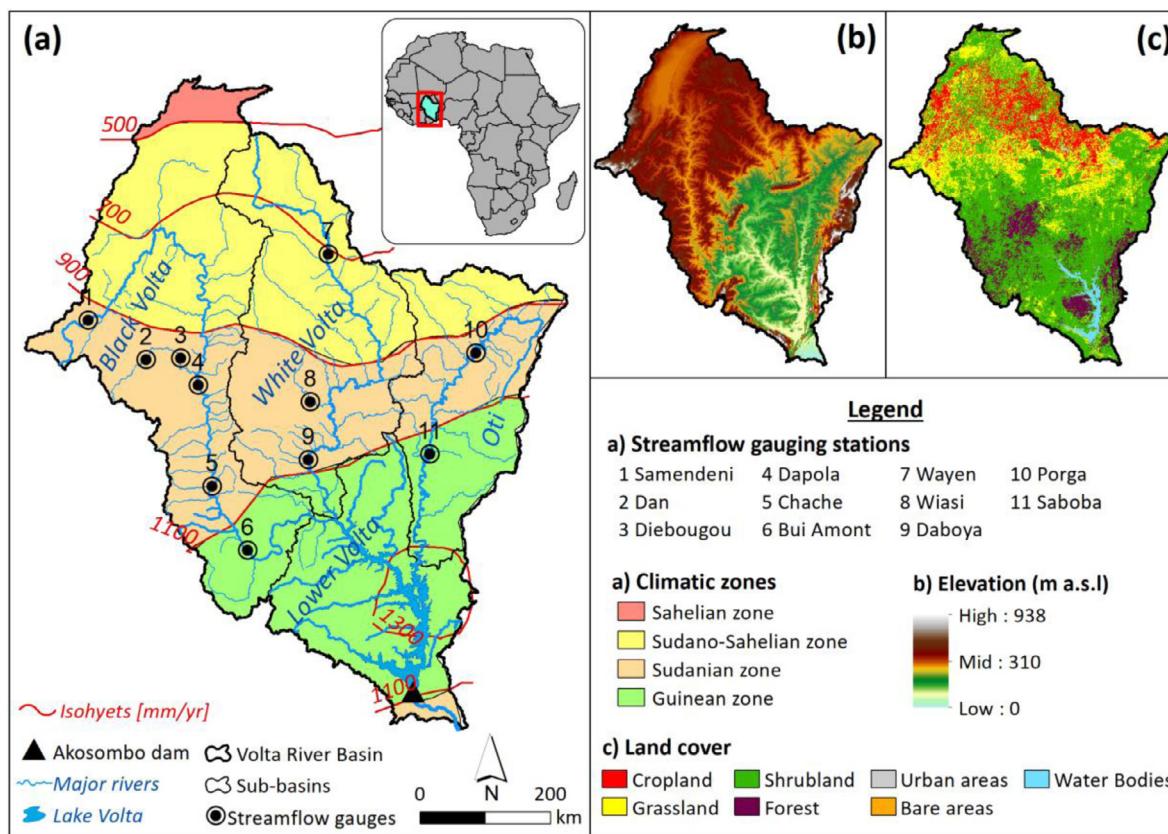


Fig. 1. Physical and hydroclimatic characteristics of the Volta River basin.

influence. This study strives to answer two inter-related research questions. Firstly, what is the ability of satellite and reanalysis evaporation datasets to improve the overall predictive skill of a fully distributed hydrological model? Secondly, how important is the model calibration strategy in improving the representation of hydrological processes? The proposed research is carried out in the Volta River basin located in West Africa, using the mesoscale Hydrologic Model (mHM) over a period of ten consecutive years (2003–2012).

2. Material and methods

2.1. Study area

The transboundary Volta River Basin (VRB) is shared among six countries of West Africa (i.e., Burkina Faso, Ghana, Togo, Mali, Benin and Côte d'Ivoire) and covers approximately 415,600 km² (Fig. 1). Most of the basin lies below 400 m a.s.l. with a predominantly flat topography (Fig. 1b). The Volta River flows north-south over a distance of 1,850 km and drains into the Atlantic Ocean at the Gulf of Guinea after transiting through Lake Volta (8,502 km²) formed by the Akosombo dam (7.94 Mm³) (Dembélé et al., 2019; Dembélé et al., 2020a; Williams et al., 2016). The basin area is composed of grassland interspersed with shrubs and trees (75%), cropland (13%), forest (9%), water bodies (2%) and bare land and settlements (1%) (Fig. 1c).

The VRB extends over four eco-climatic zones (i.e., Sahelian, Sudano-Sahelian, Sudanian and Guinean) characterized by increasing vegetation density and receiving increasing precipitation from north to south (Fig. 1a; Table 1). The information provided in Table 1 is obtained from the global aridity index database (Trabucco and Zomer, 2018), and the WFDEI meteorological data (see Table 2; Weedon et al., 2014) for the period 1979–2016. Fig. 1 and Table 1 are adapted from Dembélé et al. (2020a).

The climate is driven by the Inter-Tropical Convergence Zone (ITCZ) and varies from sub-humid in the south to semi-arid in the north (Dembélé et al., 2019; FAO/GIEWS, 1998; Mul et al., 2015). Actual evaporation varies between 70% and 90% of annual rainfall in the basin.

2.2. Distributed hydrological model

The fully distributed mesoscale Hydrologic Model (mHM) is a conceptual model that simulates dominant hydrological processes (e.g., evaporation, soil moisture and discharge) per grid cell in the modelling domain (Kumar et al., 2013; Samaniego et al., 2010). Samaniego et al. (2011) provide a schematic representation of the processes accounted for in mHM. A multiscale routing model based on the Muskingum-Cunge method (Cunge, 1969) is used for the routing of the total grid-generated runoff through the river network (Thober et al., 2019). The sub-grid variability of the basin physical characteristics (e.g., soil texture, land cover) is accounted for with a multiscale parameter regionalization technique (Samaniego et al., 2017). The model parameters (e.g., hydraulic conductivity, soil porosity) are linked to the basin physical characteristics via pedo-transfer functions and global parameters. Thirty-six global parameters (cf. Pokhrel et al., 2008) are tuned during model calibration for this study (Table S6). mHM was found reliable for modelling hydrological processes in the VRB (Dembélé et al., 2020a). The version 5.9 of mHM is used in this study.

Reference evaporation (E_{ref}) is calculated following the method of Hargreaves and Samani (1985), which was found to be reliable for semi-arid regions like the VRB (Bai et al., 2016; Er-Raki et al., 2010; Gao et al., 2017). E_{ref} is formulated as follows:

$$E_{ref} = \kappa \frac{R_a}{\lambda} (T_{avg} + 17.8) (T_{max} - T_{min})^{0.5} \quad (1)$$

where R_a (MJ/m²/day) is the extraterrestrial radiation computed based on the latitude of the location and the day of the year (Allen et al., 1998),

Table 1

Characteristics of the four eco-climatic zones in the Volta River basin with the Aridity Index (AI). The annual mean value with the range ([min-max]) is provided for Precipitation (P) and average air Temperature (T_{avg}).

Eco-climatic zones	Climate class	AI (-)	P (mm/year)	T_{avg} (°C)
Sahelian zone	Arid	0.16 [0.12-0.20]	570 [470-610]	29.3 [29.0-29.7]
Sudano-Sahelian zone	Semi-arid	0.29 [0.16-0.43]	790 [570-980]	28.6 [28.0-29.3]
Sudanian zone	Semi-arid/Dry sub-humid	0.47 [0.33-0.98]	1010 [890-1290]	28.1 [26.4-28.8]
Guinean zone	Dry sub-humid/Humid	0.70 [0.49-1.22]	1190 [1030-1420]	27.6 [26.0-28.6]

Table 2

Overview of the input datasets. CHIRPS: Climate Hazards Group InfraRed Precipitation with Station data; ESA CCI SM: European Space Agency Climate Change Initiative Soil Moisture; GIMMS: Global Inventory Modelling and Mapping Studies; GLiM: Global Lithological Map; GMTED: Global Multi-resolution Terrain Elevation Data; GRACE: Gravity Recovery and Climate Experiment; WFDEI: WATCH Forcing Data methodology applied to ERA-Interim data

Variables	Products	Spatial resolution	Temporal resolution	References
Model setup				
<u>Meteorological data</u>				
Rainfall	CHIRPS v2.0	0.05°	daily	Funk et al. (2015) http://chg.geog.ucsb.edu/data/chirps/
Temperature (average, minimum and maximum)	WFDEI	0.5°	hourly	Weedon et al. (2014) http://www.eu-watch.org/data_availability
<u>Morphological data</u>				
Terrain characteristics (elevation, slope, aspect, flow direction and flow accumulation)	GMTED 2010	225 m (0.0021°)	static	Danielson and Gesch (2011) https://topotools.cr.usgs.gov/
Soil properties (horizon depth, bulk density, sand and clay content)	SoilGrids	250 m (0.0023°)	static	Hengl et al. (2017) https://www.isric.org/explore/soilgrids
Geology	GLiM v1.0	0.5°	static	Hartmann and Moosdorf (2012) https://doi.pangaea.de/10.1594/PANGAEA.788537
Land use land cover	Globcover 2009	300 m (0.0028°)	static	Bontemps et al. (2011) http://due.esrin.esa.int/page_globcover.php
Phenology (leaf area index)	GIMMS	8 km (0.0833°)	bimonthly	Tucker et al. (2005), Zhu et al. (2013) http://cliveg.bu.edu/modismisr/lai3g-fpar3g.html
Model calibration/evaluation				
<u>In-situ data</u>				
Streamflow (Q)	-	point	daily	Multiple organizations (see acknowledgements)
<u>Satellite data</u>				
Terrestrial water storage anomaly (S_t)	GRACE TellUS v5.0	1°	monthly	Tapley et al. (2004), Swenson (2012) https://grace.jpl.nasa.gov/
Surface soil moisture (S_u)	ESA CCI SM v4.2	0.25°	daily	Dorigo et al. (2017) https://www.esa-soilmoisture-cci.org/
Actual evaporation (E_a)				See Table 3

$\lambda = 2.45$ MJ/kg is the latent heat of vaporization of water. The unit of radiation is converted into equivalent water evaporation in mm/day with the ratio R_a/λ . The differences in advection or vapor transfer effect are compensated by the constant $\kappa = 0.0023$, and T_{avg} , T_{max} and T_{min} represent the daily average, maximum and minimum air temperature in degrees Celcius (°C) at a given location.

Potential evaporation (E_p) is calculated by adjusting E_{ref} with a dynamical scaling function (F_{DS}) based on leaf area index (Allen et al., 1998; Demirel et al., 2018), therefore accounting for vegetation-climate interactions (Bai et al., 2018; Birhanu et al., 2019; Jiao et al., 2017). E_p is formulated as follows:

$$E_p = F_{DS} \cdot E_{ref}, \text{ with} \quad (2)$$

$$F_{DS} = a + b(1 - e^{(c \cdot I_{LA})}) \quad (3)$$

where a is the intercept term, b is the vegetation dependent component, and c represents the degree of nonlinearity of the leaf area index (I_{LA}). The coefficients a , b , and c are determined through model calibration.

In this study, actual evaporation (E_a) is defined as the sum of transpiration and evaporation from interception, land and water bodies (Coenders-Gerrits et al., 2020; Shuttleworth, 1993). E_a is calculated as a fraction of E_p from soil layers depending on soil moisture availability and the rooting depth (Feddes et al., 1976). Soil moisture is estimated by a multi-layer infiltration capacity approach adopting a three-layer soil scheme (0-5, 5-30 and 30-100 cm depths). Terrestrial water storage at each grid cell is the sum of the surface and subsurface water storage (i.e., lakes, wetlands, soil moisture reservoirs, interflow and baseflow).

More information on the calculation of hydrological processes in mHM can be found in the work of Kumar (2010) and Samaniego et al. (2010).

2.3. Input datasets

The morphological datasets (i.e., elevation, slope, land cover, etc.) and meteorological datasets (i.e., rainfall and air temperature) used to set up and run the distributed model are described in Table 2. The meteorological datasets are selected based on their accuracy (Dembélé and Zwart, 2016), and their suitability to plausibly represent hydrological processes in the VRB (Dembélé et al., 2020b). All morphological data are resampled to a resolution of 1/512° (~200 m at the equator) using the nearest neighbor technique, while the meteorological data are resampled to 0.0625° (~7 km) using bilinear resampling. Due to the good ability of mHM in parameter transferability across scales (Dembélé et al., 2020a), the model is run at daily time step with a spatial discretization of 0.25° (~28 km) corresponding to 619 modelling grid cells in the basin, which attenuate the computational demand. In-situ streamflow data and gridded datasets of evaporation are used for model calibration (Section 2.4). The description of the evaporation datasets is provided in Section 2.3.

The quality control and the gap-filling of the streamflow data are described by Dembélé et al. (2019). In addition to in-situ streamflow, SRS datasets of soil moisture and terrestrial water storage are used to evaluate the model performance after calibration. The surface soil moisture (S_u) dataset is obtained from ESA CCI (Dorigo et al., 2017) and represents the first soil layer (i.e., 2-5 cm depth). The blended

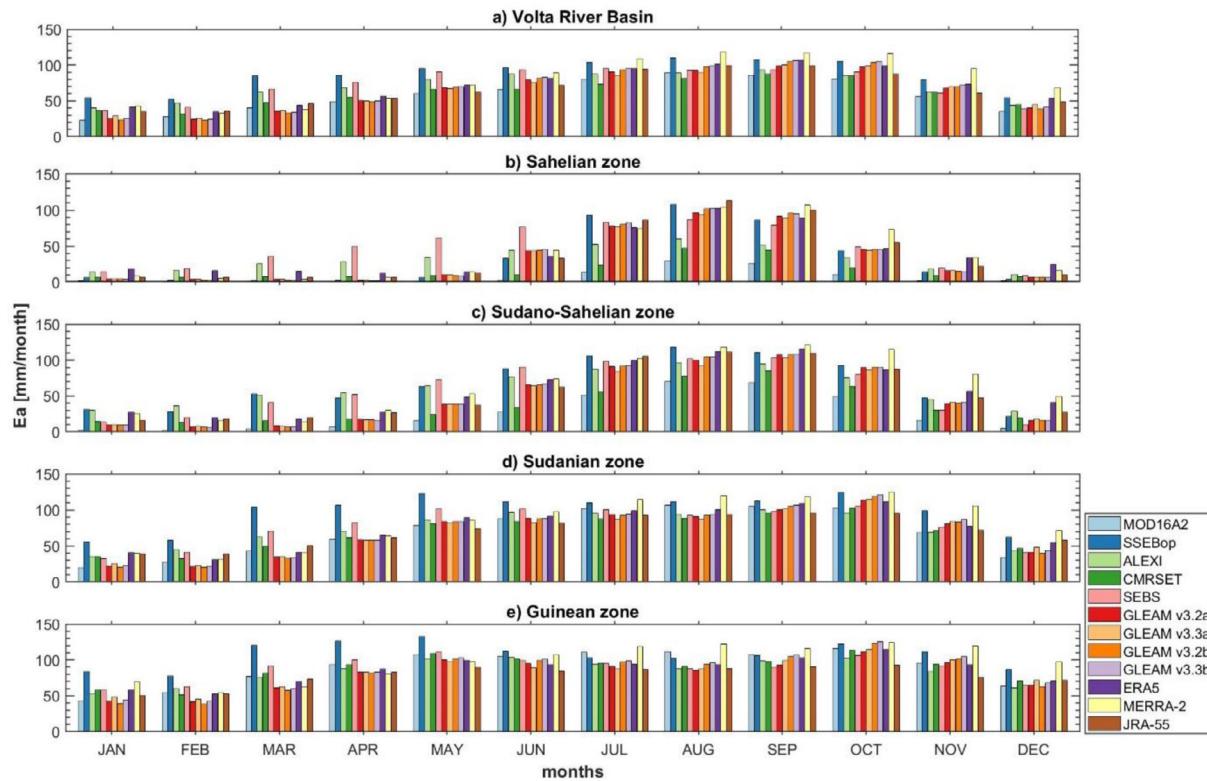


Fig. 2. Mean monthly total actual evaporation (E_a) of 12 gridded evaporation datasets averaged for the four climatic zones (sub-figures b–e) in the Volta River basin over the period 2003–2012.

product of both passive and active microwave products is used here (Gruber et al., 2017; Liu et al., 2012; Wagner et al., 2012). The terrestrial water storage (S_t) anomaly data (release RL05) is obtained from GRACE (Landerer and Swenson, 2012; Swenson, 2012). The ensemble arithmetic mean of different solutions from three processing centers (i.e., Center for Space Research at University of Texas, Geoforschungs Zentrum Potsdam and Jet Propulsion Laboratory) is used in this study because it has been shown to be more effective in reducing noise in the Earth's gravity signal than the individual solutions (Sakumura et al., 2014).

2.4. Evaporation datasets

Twelve gridded actual evaporation datasets including nine SRS-based products (MOD16A2, SSEBop, ALEXI, CMRSET, SEBS, GLEAM v3.2a, GLEAM v3.3a, GLEAM v3.2b, GLEAM v3.3b) and three reanalysis products (ERA5, MERRA-2, JRA-55) are evaluated in this study. Based on evaporation modelling approaches (Yilmaz et al., 2014), the SRS-based datasets can be further classified as diagnostic products (MOD16A2, SSEBop, ALEXI, CMRSET, SEBS) or prognostic products (GLEAM v3.2a, GLEAM v3.3a, GLEAM v3.2b, GLEAM v3.3b), while the reanalysis datasets are all prognostic products. A summary of the evaporation datasets is provided in Table 3, and more details on each dataset can be found in the corresponding references.

Four versions of the GLEAM product are evaluated. They differ in terms of input data used for their production and in terms of their spatiotemporal coverage (cf. Table 1 in Martens et al., 2017). The version v3.3 differs from the v3.2 in the following forcing datasets: surface radiation, near-surface air temperature and land cover maps. The versions v3.3a and v3.2a are produced with reanalysis, satellite and gauge-based datasets, while the versions v3.3b and v3.2b are mainly produced with satellite datasets.

Considerable differences can be observed both in the temporal dynamics and the spatial patterns of the 12 gridded evaporation datasets across the climatic zones in the VRB (Figs. 2 and 3).

2.5. Model calibration and evaluation

The modelling period extends from 2000 to 2012 with 3 years of model warm-up (2000–2002), 6 years for calibration (2003–2008), and 4 years for evaluation (2009–2012). Available daily in-situ streamflow datasets from 11 gauging locations are used for model calibration and evaluation, while monthly datasets of E_a (Table 3) are used for model calibration, and monthly datasets of S_u (ESA CCI) and S_t (GRACE) are used for model evaluation. All the E_a datasets are rescaled to 0.25° using bilinear interpolation to match the modelling spatial resolution, and sub-monthly data are aggregated to monthly resolution. Only the first soil layer of mHM is compared to the ESA CCI data, which represents the surface soil moisture.

First, a streamflow-only calibration is adopted as benchmark. Then, the contribution of evaporation datasets in improving hydrological model calibration is tested by simultaneously constraining the model with streamflow and each of the twelve gridded evaporation datasets using four calibration strategies for each (Fig. 4). Therefore, 48 scenarios (i.e., 12 datasets times 4 calibration strategies) are developed and compared to the benchmark calibration to evaluate the impact of different calibration strategies on model performance. The dynamically dimensioned search algorithm (Tolson and Shoemaker, 2007) is used for parameter estimation, using 5,000 iterations for each of the 48 scenarios and for the benchmark model calibration. The computational runtime is about 6 days for each of the 49 model simulations on a computer Intel Xeon Processor E5-2697 v3 with 64 GB of RAM.

2.5.1. Calibration on streamflow data – benchmark

The benchmark calibration (case Q) is elaborated by calibrating the hydrological model solely with streamflow (Q) data. The objective func-

Table 3

Gridded actual evaporation datasets. The characteristics of the datasets are those used in this study although the same datasets can be available from the data providers with different versions and spatiotemporal resolutions.

Datasets	Name/ Data portal	Spatial coverage	Spatial resolution	Temporal coverage	Temporal resolution	References
MOD16A2	Moderate Resolution Imaging Spectroradiometer (MODIS) Global Terrestrial Evapotranspiration Algorithm version 5 http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/MOD16A2_MONTHLY.MERRA_GMAO_1kmALB/	Global	0,0085° (~1 km)	2001-2014	Monthly	Mu et al. (2011)
SSEBop	Operational Simplified Surface Energy Balance https://edcintl.cr.usgs.gov/downloads/sciweb1/shared/fews/web/	Global	0.0083°(~1 km)	2003-2014	Monthly	Senay et al. (2007), Senay et al. (2013)
ALEXI	Atmosphere-Land Exchange Inverse ftp://ftp.wateraccounting.unesco-ihe.org/WaterAccounting/Data_Satellite/Evaporation/ALEXI/World/	70° N–60° S	0.05° (~5.6 km)	2003-2015	Monthly	Anderson et al. (1997), Anderson et al. (2007)
CMRSET	CSIRO MODIS Reflectance Scaling EvapoTranspiration http://remote-sensing.nci.org.au/u39/public/data/wirada/cmrset/	Global	0.05° (~5.6 km)	2001-2013	Monthly	Guerschman et al. (2009)
SEBS	Surface Energy Balance System ftp://ftp.wateraccounting.unesco-ihe.org/WaterAccounting/Data_Satellite/Evaporation/SEBS/SEBS/	40° N–40° S	0.05° (~5.6 km)	2001-2012	Monthly	Su (2002)
GLEAM v3.2a GLEAM v3.3a GLEAM v3.2b GLEAM v3.3b	Global Land Evaporation Amsterdam Model https://www.gleam.eu	Global	0.25° (~28 km)	1980-present	Daily	Martens et al. (2017), Miralles et al. (2011)
			50° N–50° S		2003- present	
ERA5	European Centre for Medium-range Weather Forecasts ReAnalysis 5 (ERA5) hourly data on single levels https://cds.climate.copernicus.eu/	Global	0.25°(~28 km)	1979- present	Hourly	Hersbach et al. (2018), Hersbach et al. (2020)
MERRA-2	Modern-Era Retrospective Analysis for Research and Applications 2 (Evaporation_land: M2TUNXLND_V5.12.4) https://disc.gsfc.nasa.gov/datasets/M2TUNXLND_V5.12.4/summary	Global	0.5° x 0.625°(~56 km)	1980- present	Hourly	Gelaro et al. (2017), Reichle et al. (2017)
JRA-55	Japanese 55 year ReAnalysis (JRA-55); evaporation: fcst_phy2m125 https://jra.kishou.go.jp/JRA-55/index_en.html	Global	1.25°(~140 km)	1959- present	3-hourly	Kobayashi et al. (2015), Harada et al. (2016)

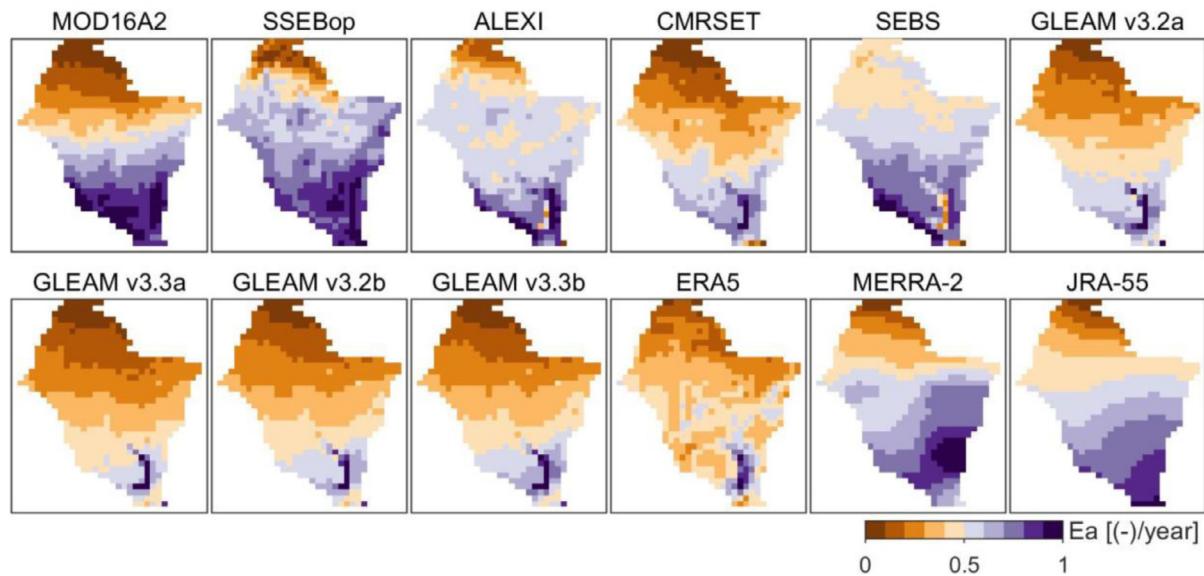


Fig. 3. Spatial distribution of total annual actual evaporation (E_a) for 12 gridded evaporation datasets in the Volta River basin averaged for the period 2003–2012. A min-max normalization of the values allows rescaling them between 0 and 1 to facilitate the intercomparison of the spatial patterns among the datasets.

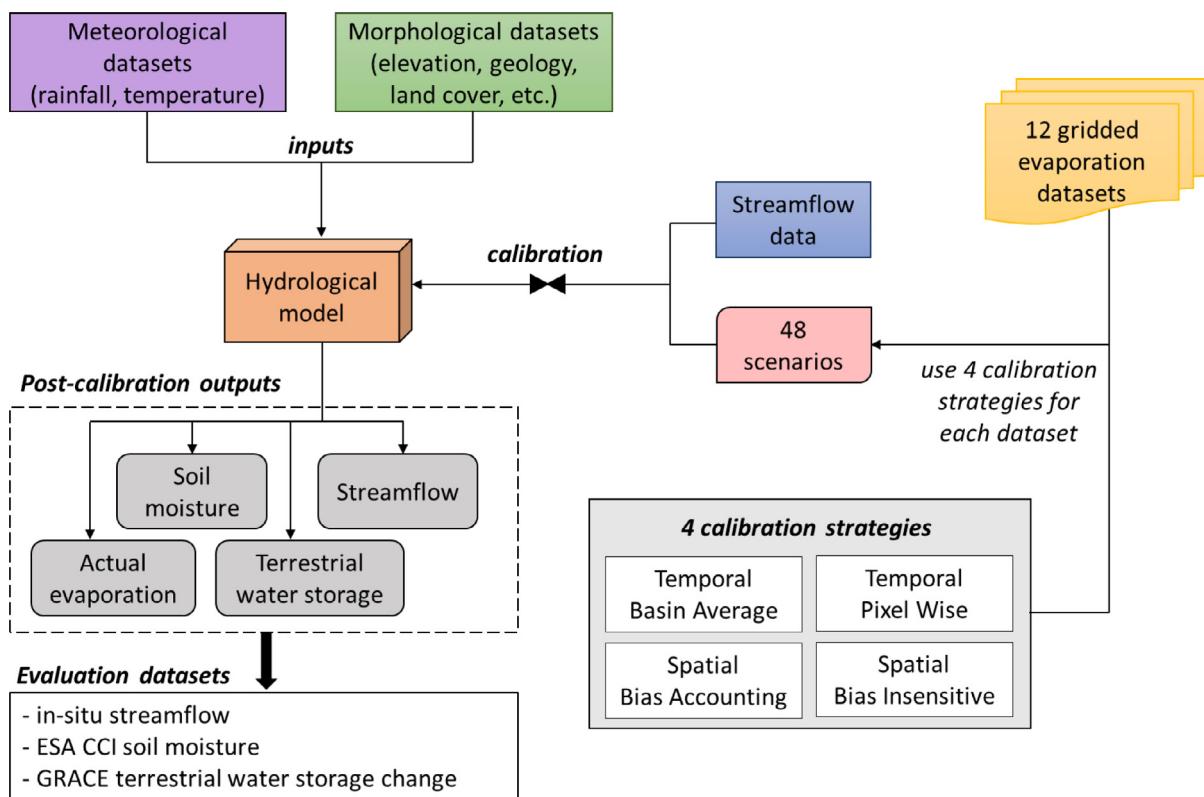


Fig. 4. Overview of the modelling approach to evaluate the reanalysis and satellite-based evaporation datasets.

tion for case Q (Φ_Q) to be minimized is obtained by calculating the average Kling-Gupta efficiency (E_{KG}) over the 11 gauging points for Q in the basin, and subtracting it from 1. Φ_Q ranges from its optimal value that is 0 to positive infinity, and is formulated as follows:

$$\Phi_Q = 1 - \left[\frac{1}{g} \sum_{i=1}^g E_{KG,i}(Q_{\text{mod},i}, Q_{\text{obs},i}) \right], \text{ with} \quad (4)$$

$$E_{KG} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad (5)$$

where g is the total number of streamflow gauging stations in the basin, E_{KG} is the modified Kling-Gupta efficiency (Kling et al., 2012) calculated for the observed ($Q_{\text{obs},i}$) and modelled ($Q_{\text{mod},i}$) streamflow of the i^{th} gauging point. E_{KG} is composed of the Pearson correlation coefficient (r), the bias term (β , i.e., the ratio of the means) and the variability term (γ , i.e., the ratio of the coefficients of variation, cf. Eq. 8). E_{KG} ranges from negative infinity to its optimal value that is 1. A model is better than the mean observed flow if $E_{KG} > -0.41$ (Knoben et al., 2019).

2.5.2. Multivariate calibration with streamflow and evaporation

2.5.2.1. Spatial pattern efficiency metric. A bias-insensitive and multi-component metric developed by Dembélé et al. (2020a) is used to quantify the degree of reproduction of the spatial patterns of hydrological processes. The proposed spatial pattern efficiency (E_{SP}) metric only considers the spatial pattern of the underlying variables and ignores their absolute values. E_{SP} is an integrated measure of the dynamics, the spatial variability, and the locational matching of grid cells between the modelled (X_{mod}) and observed (X_{obs}) variables. With X_{obs} and X_{mod} composed of n cells, E_{SP} is defined as follows:

$$E_{SP} = 1 - \sqrt{(r_s - 1)^2 + (\gamma - 1)^2 + (\alpha - 1)^2}, \text{ with} \quad (6)$$

$$r_s = 1 - \frac{6 \sum_i^n d_i^2}{n(n^2 - 1)}, \quad (7)$$

$$\gamma = \frac{\sigma_{mod}/\mu_{mod}}{\sigma_{obs}/\mu_{obs}} \text{ and} \quad (8)$$

$$\alpha = 1 - E_{RMS}(Z_{X_{mod}}, Z_{X_{obs}}) \quad (9)$$

where r_s is the Spearman rank-order correlation coefficient with d_i the difference between the ranks of the i^{th} cell of X_{mod} and X_{obs} , γ is the variability ratio (i.e., the ratio of the coefficients of variation), and α the spatial location matching term calculated as the root mean squared error (E_{RMS}) of the standardized values (z-scores, Z_X) of X_{mod} and X_{obs} (Dembélé et al., 2020a). E_{SP} ranges from negative infinity to its optimal value that is unity. $E_{SP} = 0$ when there is a moderate relationship between the ranks of the observed and modelled variables (i.e., $r_s = 0.55$), and $E_{SP} = -0.67$ when the ranks are not related (i.e., $r_s = 0$). More details on E_{SP} are provided by Dembélé et al. (2020a).

2.5.2.2. Multivariate calibration strategies. Four multivariate calibration strategies with distinct objective functions are proposed to simultaneously consider Q and E_a data as calibration variables. Each objective function (Eqs. 11, 15, 17 and 19) is formulated based on the Euclidean distance approach (Eq. 10), in which all elements are equally weighted (Khu and Madsen, 2005). The Euclidian distance (D_E) between two points X and Y of coordinates (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) in an n -dimensional space (Upton and Cook, 2014) is given by:

$$D_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (10)$$

The four multivariate calibration strategies differ from each other based on the formulation of the sub-objective function for E_a (i.e., Φ_{E_a}), while Φ_Q remains unchanged. The observed variable ($E_{a,obs}$) and the modelled variable ($E_{a,mod}$) of actual evaporation are represented each by a 3D array of dimension [$M \times N \times T$], with M the number of rows, N the number of columns, and T the number of time steps (Fig. 5). For this study, M represents 40 latitude rows, N represents 36 longitude columns and T represents 72 months (i.e. calibration period). The modelling domain has 1440 grid cells of which 619 are active (i.e. grid cells representing the basin area). The inactive grid cells are masked out during the calculation of the performance metrics.

In the following, E'_a is a 2D array of actual evaporation represented by all cells (i,j) in the spatial domain Ω , and E_a^{ij} is the time series of actual evaporation for a given cell at row i and column j . The four multivariate calibration strategies are defined as follows:

1- Temporal basin average (BA): the matching of the observed and modelled E_a is done on basin-averaged time series. The sub-objective function ($\Phi_{E_a,BA}$) to be minimized is obtained by calculating the E_{KG} for the observed and modelled time series of basin average actual evaporation ($\overline{E_{a,obs}}$ and $\overline{E_{a,mod}}$) and subtracting it from 1. The objective function (Φ_{BA}) is formulated as follows:

$$\Phi_{BA} = \sqrt{\Phi_Q^2 + \Phi_{E_a,BA}^2}, \text{ with} \quad (11)$$

$$\Phi_{E_a,BA} = 1 - E_{KG}\left(\overline{E_{a,mod}}, \overline{E_{a,obs}}\right), \text{ where} \quad (12)$$

$$\overline{E_{a,mod}}(t) = \frac{1}{M \times N} \sum_{t=1}^T \mathbf{E}_{a,mod}^t(i, j), \text{ and} \\ \forall i, j \in \Omega \quad (13)$$

$$\overline{E_{a,obs}}(t) = \frac{1}{M \times N} \sum_{t=1}^T \mathbf{E}_{a,obs}^t(i, j) \\ \forall i, j \in \Omega \quad (14)$$

2- Temporal pixel-wise (PW): the matching of the modelled and observed E_a is done individually on the time series of each grid cell in the basin. The E_{KG} is calculated for the observed and modelled time series of E_a at each grid cell in the basin, and the sub-objective function ($\Phi_{E_a,PW}$) to be minimized is the average of the E_{KG} calculated for all grids, subtracted from 1. The objective function (Φ_{PW}) is formulated as follows:

$$\Phi_{PW} = \sqrt{\Phi_Q^2 + \Phi_{E_a,PW}^2}, \text{ with} \quad (15)$$

$$\Phi_{E_a,PW} = 1 - \left[\frac{1}{M \times N} \sum_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N \\ \forall t \in T}} E_{KG}\left(E_{a,mod}^{ij}, E_{a,obs}^{ij}\right) \right] \quad (16)$$

3- Spatial bias-accounting (SB): the matching of the modelled and observed E_a is done for all pixels at each time step. The E_{KG} is calculated at each time step between all the pixels of the observed and modelled E_a . The sub-objective function ($\Phi_{E_a,SB}$) to be minimized is the average of the E_{KG} calculated for all time steps, subtracted from 1. The objective function (Φ_{SB}) is formulated as follows:

$$\Phi_{SB} = \sqrt{\Phi_Q^2 + \Phi_{E_a,SB}^2}, \text{ with} \quad (17)$$

$$\Phi_{E_a,SB} = 1 - \frac{1}{T} \sum_{t=1}^T E_{KG}\left(\mathbf{E}_{a,mod}^t(i, j), \mathbf{E}_{a,obs}^t(i, j)\right) \\ \forall i, j \in \Omega \quad (18)$$

4- Spatial bias-insensitive (SP): the sub-objective function ($\Phi_{E_a,SP}$) to be minimized is similarly calculated as for the SB calibration except that a bias-insensitive metric (i.e., E_{SP}) is used as skill score instead of E_{KG} . The objective function (Φ_{SP}) is formulated as follows:

$$\Phi_{SP} = \sqrt{\Phi_Q^2 + \Phi_{E_a,SP}^2}, \text{ with} \quad (19)$$

$$\Phi_{E_a,SP} = 1 - \frac{1}{T} \sum_{t=1}^T E_{SP}\left(\mathbf{E}_{a,mod}^t(i, j), \mathbf{E}_{a,obs}^t(i, j)\right) \\ \forall i, j \in \Omega \quad (20)$$

All the objective functions (Φ_{BA} , Φ_{PW} , Φ_{SB} and Φ_{SP}) vary between their optimal value that is 0 and positive infinity. Except case SP that is a bias-insensitive approach, other calibration strategies consider the absolute values (i.e. raw data) of evaporation.

Observed and modelled actual Evaporation (E _a) data	Transformation for calibration strategies		
	- Temporal Pixel-wise (PW)	- Temporal basin average (BA) - Spatial bias-accounting (SB) - Spatial bias-insensitive (SP)	
 $E_{a,obs}$ latitude longitude time	<p>For each cell (i,j)</p> $E_{a,obs} = \begin{matrix} \text{---} \\ M \\ \text{---} \\ N \end{matrix} \quad E_{a,obs}^{ij} = \begin{matrix} \text{---} \\ 1 \quad \text{to} \quad T \\ \text{---} \end{matrix}$ $E_{a,obs}^{ij} = E_{a,obs}(i, j, :)$	<p>For each time step t</p> $E_{a,obs} = \begin{matrix} \text{---} \\ M \\ \text{---} \\ N \end{matrix} \quad E_{a,obs}^t = \begin{matrix} \text{---} \\ 1 \quad \text{to} \quad T \\ \text{---} \end{matrix}$ $\overline{E_{a,obs}}(t) = \text{mean}(E_{a,obs}^t)$ $\overline{E_{a,obs}} = \text{basin-averaged } E_{a,obs}$	
 $E_{a,mod}$ latitude longitude time	$E_{a,mod} = \begin{matrix} \text{---} \\ M \\ \text{---} \\ N \end{matrix} \quad E_{a,mod}^{ij} = \begin{matrix} \text{---} \\ 1 \quad \text{to} \quad T \\ \text{---} \end{matrix}$ $E_{a,mod}^{ij} = E_{a,mod}(i, j, :)$	$E_{a,mod} = \begin{matrix} \text{---} \\ M \\ \text{---} \\ N \end{matrix} \quad E_{a,mod}^t = \begin{matrix} \text{---} \\ 1 \quad \text{to} \quad T \\ \text{---} \end{matrix}$ $\overline{E_{a,mod}}(t) = \text{mean}(E_{a,mod}^t)$ $\overline{E_{a,mod}} = \text{basin-averaged } E_{a,mod}$	
Objective functions for E _a	$\Phi_{E_a_PW} = 1 - \left[\frac{1}{M \times N} \sum_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N \\ \forall t \in T}} E_{KG} \left(E_{a,mod}^{ij}, E_{a,obs}^{ij} \right) \right]$	$\Phi_{E_a_BA} = 1 - E_{KG} \left(\overline{E_{a,mod}}, \overline{E_{a,obs}} \right)$ $\Phi_{E_a_SB} = 1 - \frac{1}{T} \sum_{t=1}^T E_{KG} \left(E_{a,mod}^t(i, j), E_{a,obs}^t(i, j) \right)$ $\Phi_{E_a_SP} = 1 - \frac{1}{T} \sum_{t=1}^T E_{SP} \left(E_{a,mod}^t(i, j), E_{a,obs}^t(i, j) \right)$	

Fig. 5. Graphical and programming syntax description of the use of evaporation data in the four calibration strategies. The blue matrix represents the observed data while the brown matrix represents the modelled data. For the syntax, (i, j, :) means ith element of the first dimension (latitude), jth element of the second dimension (longitude) and all elements of the third dimension (time). Similarly, (:, :, t) means tth element of the third dimension and all elements of the first and second dimensions. The objective functions are calculated based on the Kling-Gupta efficiency (E_{KG}) and the spatial pattern efficiency (E_{SP}) metrics.

2.6. Model output evaluation

In addition to daily streamflow data, independent monthly datasets of satellite-based soil moisture (S_u) and terrestrial water storage (S_t) (Table 2) are used for model evaluation. The temporal dynamics of streamflow is evaluated with E_{KG} . Both the temporal dynamics and the spatial patterns of modelled S_u are evaluated using the Pearson correlation coefficient (r) and the spatial pattern efficiency metric (E_{SP}), while only the temporal dynamics of modelled S_t is assessed using r , due to the coarse spatial resolution of the GRACE data. The skill scores for the temporal dynamics are calculated for each pixel (or gauging point) in the basin, while spatial skill scores are calculated per time step.

3. Results

In general, the trend of the model performance (high vs. low scores) among the scenarios (i.e., evaporation datasets vs. calibration strategies) is conserved between the calibration and the evaluation periods for all variables. Therefore, the following results are presented for the entire simulation period that comprises the calibration and evaluation periods whose results are additionally provided in the supplementary materials.

3.1. Model performance for multiple hydrological processes in the VRB

The model performance for various hydrological processes in the VRB reveals the potential of SRS and reanalysis evaporation datasets to improve the model responses if the appropriate calibration strategy

is used (Fig. 6). Detailed results are provided in the supplementary material (Figures S1-S2).

For Q , the benchmark model (i.e., Q-only) yields a median E_{KG} of 0.69. In the multivariate calibration scenarios (i.e., $Q+E_a$), the E_{KG} of Q varies between 0.42 for SEBS with case PW to 0.73 for CMRSET with case SP. The best multivariate calibration strategy is the case SB with an average E_{KG} of 0.68 and 75% of the evaporation datasets producing a higher model performance than the benchmark, followed by case SP ($E_{KG} = 0.67$), case PW ($E_{KG} = 0.63$) and case BA ($E_{KG} = 0.60$). The top 3 best evaporation datasets for the average E_{KG} of Q over the calibration strategies are GLEAM v3.2b ($E_{KG} = 0.71$), GLEAM v3.3b ($E_{KG} = 0.71$) and GLEAM v3.3a ($E_{KG} = 0.70$), while the worst are ALEXI ($E_{KG} = 0.61$), SEBS ($E_{KG} = 0.56$) and ERA5 ($E_{KG} = 0.51$). The decrease in the model performance for Q in the multivariate calibration might be an artifact caused by equifinality (i.e., non-uniqueness of model parameters; Beven, 2006; Savenije, 2001) that occurred with the Q-only calibration, which gives more degrees of freedom for constraining the model parameter space Dembélé et al. (2020a). In fact, Figures S68-S71 show that the high performance for Q achieved with the Q-only calibration is obtained at the expense of poor performance for other hydrological processes, while the multivariate calibrations with $Q+E_a$ result in parameter sets that provide equivalent model performance for Q but higher model performance for S_t and S_u . These results support the findings of Dembélé et al. (2020a), thereby confirming the pitfalls of the Q-only calibration (Bouaziz et al., 2020; Stisen et al., 2018).

The evaporation datasets show a high potential to improve the temporal dynamics of modelled S_t as 79% of the multivariate calibration scenarios outperform the case Q ($r = 0.73$) with an average r of 0.81. The

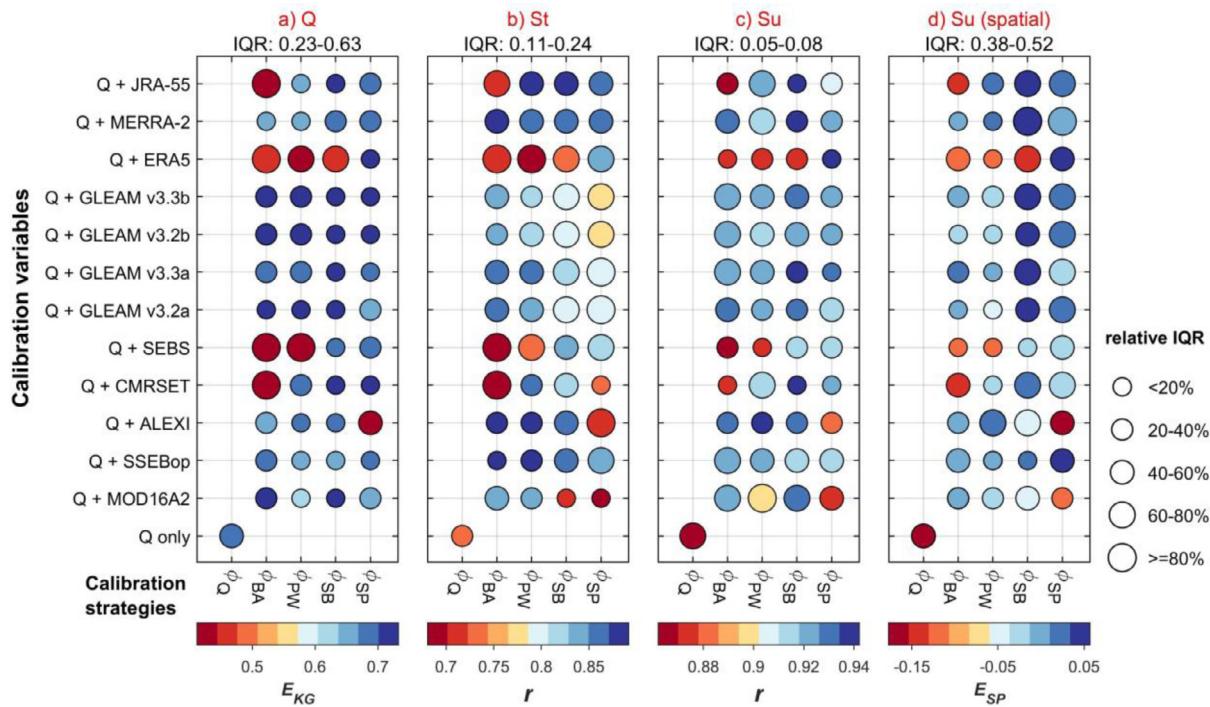


Fig. 6. Model performance in the entire simulation period (2003–2012) for the temporal dynamics (a, b, c) of streamflow (Q), terrestrial water storage (S_t) and soil moisture (S_u), and the spatial patterns (d) of S_u in the Volta River basin. The x-axis gives the objective functions for different model calibration strategies. The y-axis indicates the variables used for the model calibration. Q-only is the benchmark calibration. Circle color represents the median model performance obtained with 11 gauging points for Q , with 52 pixels for S_t , with 619 pixels for S_u (c) or with 120 months for S_u (d). The color bars show the skill score (i.e., E_{KG} , r , E_{SP}). Circle size represents model performance variability in terms of relative Interquartile Range (IQR) computed as $(IQR - IQR_{\min})/(IQR_{\max} - IQR_{\min})$. The IQR range for the 49 scenarios is given in the subplot titles. The best model is the bluest and smallest circle, while the worst model is the reddest and largest circle.

lowest average r of S_t among the calibration strategies is given by case SP ($r = 0.79$), but interestingly, it outperforms the case Q. The highest performances for median r of S_t are obtained with MERRA-2 ($r = 0.87$), SSEBop ($r = 0.87$) and ALEXI ($r = 0.84$) while lowest performances are given by SEBS ($r = 0.77$), MOD16A2 ($r = 0.76$) and ERA5 ($r = 0.74$).

The temporal dynamics of modelled S_u show a higher model performance than that of S_t , with an average r of 0.91 and 98% of the multivariate calibration scenarios that outperform the case Q ($r = 0.86$). The case BA performs similarly to case PW and case SP with an average r of 0.91 across evaporation datasets, and outperformed by case SB ($r = 0.93$). The highest r of S_u is 0.93, and it is obtained with GLEAM v3.3a, MERRA-2 and GLEAM v3.2a, while the weakest scores are obtained with MOD16A2 ($r = 0.91$), SEBS ($r = 0.89$) and ERA5 ($r = 0.89$).

The representation of the spatial patterns of S_u improves for all the multivariate scenarios as compared to the case Q. However, case SB ($E_{SP} = 0.0$) has the highest average performance across the evaporation datasets, while case BA ($E_{SP} = -0.04$) has the lowest performance. The best evaporation datasets for the simulation of the spatial patterns of S_u considering all the calibration strategies are SSEBop ($E_{SP} = 0.02$), MERRA-2 ($E_{SP} = 0.01$) and GLEAM v3.3a ($E_{SP} = 0.01$), while the worst are MOD16A2 ($E_{SP} = -0.04$), SEBS ($E_{SP} = -0.07$) and ERA5 ($E_{SP} = -0.09$).

In general, it is observed that the model performances for S_t and S_u improve for most of the multivariate calibration scenarios. Among the multivariate calibration strategies, case SB gives the best results considering the average model performance for all variables (Q , S_t and S_u). In Fig. 7b, the highest average relative change in model performance by multivariate calibration strategies as compared to the Q-only calibration is obtained with case SB is (+29%), followed by case SP (+26%), case PW (+24%) and case BA (+20%). Consequently, all grid-based model calibration strategies outperform the basin-average calibration (case BA), which gives the lowest average model performance. These results highlight the value of calibrating hydrological models on the full extent of

gridded evaporation datasets. It is noted that, in most of the scenarios, calibrating the model only on the spatial patterns (case SP) of the evaporation datasets, thereby ignoring their absolute values, improves the predictive skill of the model with a higher performance than that of case Q and case BA (Fig. 6). With these findings, case SP (i.e., only spatial patterns of evaporation datasets) can be preferred to case SB (i.e., absolute values of evaporation datasets) for the calibration of hydrological models because the propagation of the errors of the evaporation estimates into the modelling process can be case specific and depends on the model structure. This observation also draw attention on the adequacy of the model for a given experiment (Addor and Melsen, 2019). Moreover, for a given evaporation dataset, the spatial variation in biases in the estimates can lead to contrasting performance across regions (Jung et al., 2019; Nicholson, 2000).

The spatial patterns of ALEXI, followed by those of MOD16A2 and CMRSET, were the least informative for the model calibration in case SP (Fig. 6). Considering all hydrological processes and model calibration strategies, the top three best performing evaporation datasets are MERRA-2, GLEAM v3.3a and SSEBop, while the bottom three datasets are MOD16A2, SEBS and ERA5 (Figs. 6 and 7a). However, it is noteworthy that they outperform the Q-only calibration when used in case SP, meaning that only their spatial patterns improve the model performance. In general, the versions 3.3 of GLEAM show a slightly higher model performance than the versions 3.2.

3.2. Impact of calibration strategies on spatial patterns

Different model calibration strategies result in different spatial patterns of modelled E_a and S_u , as shown for selected scenarios in Fig. 8. Additional maps of E_a , S_u and S_t for other scenarios are provided in the supplementary material (Figures S58–59, Figures S65–66 and Figures S38–39), along with those obtained with the Q-only calibration (Figure

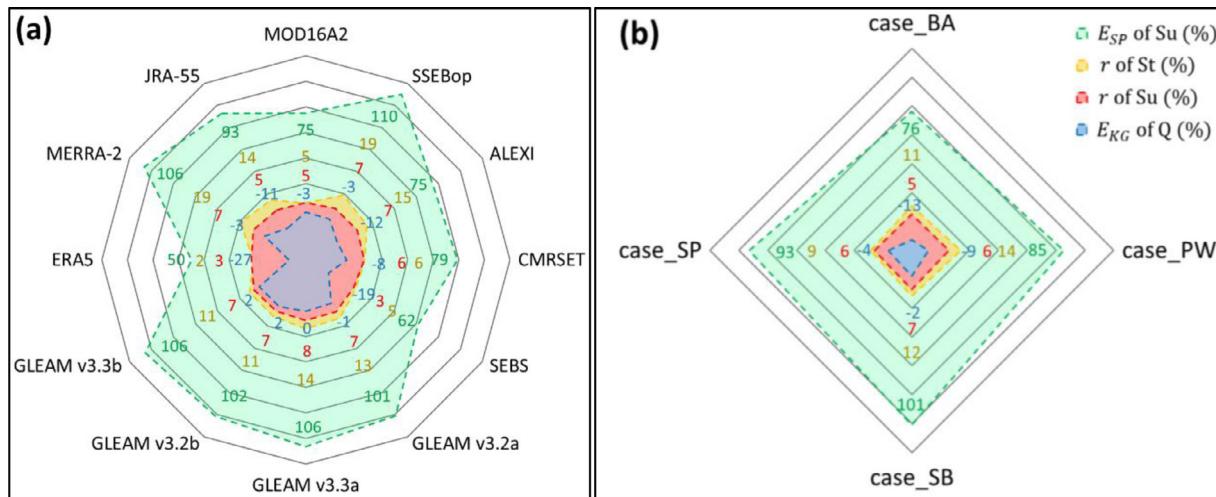


Fig. 7. Relative change in the hydrological model performance as compared to the Q-only calibration strategy when (a) adopting a multivariate calibration strategy ($Q+E_a$) with twelve evaporation datasets, or (b) using four different calibration strategies. The values on the line from each vertex to the center of the polygons give the relative difference in model performance as compared to the Q-only calibration. The change in model performance is given in percentage of the performance metrics (i.e., E_{KG} , r , E_{SP}) for streamflow (Q), terrestrial water storage (S_t), and soil moisture (S_u). For instance in (a), using SSEbop in multivariate calibration increases E_{SP} of S_u by 110%, while in (b), the calibration strategy case BA decreases E_{KG} of Q by -13%, as compared to the Q-only calibration.

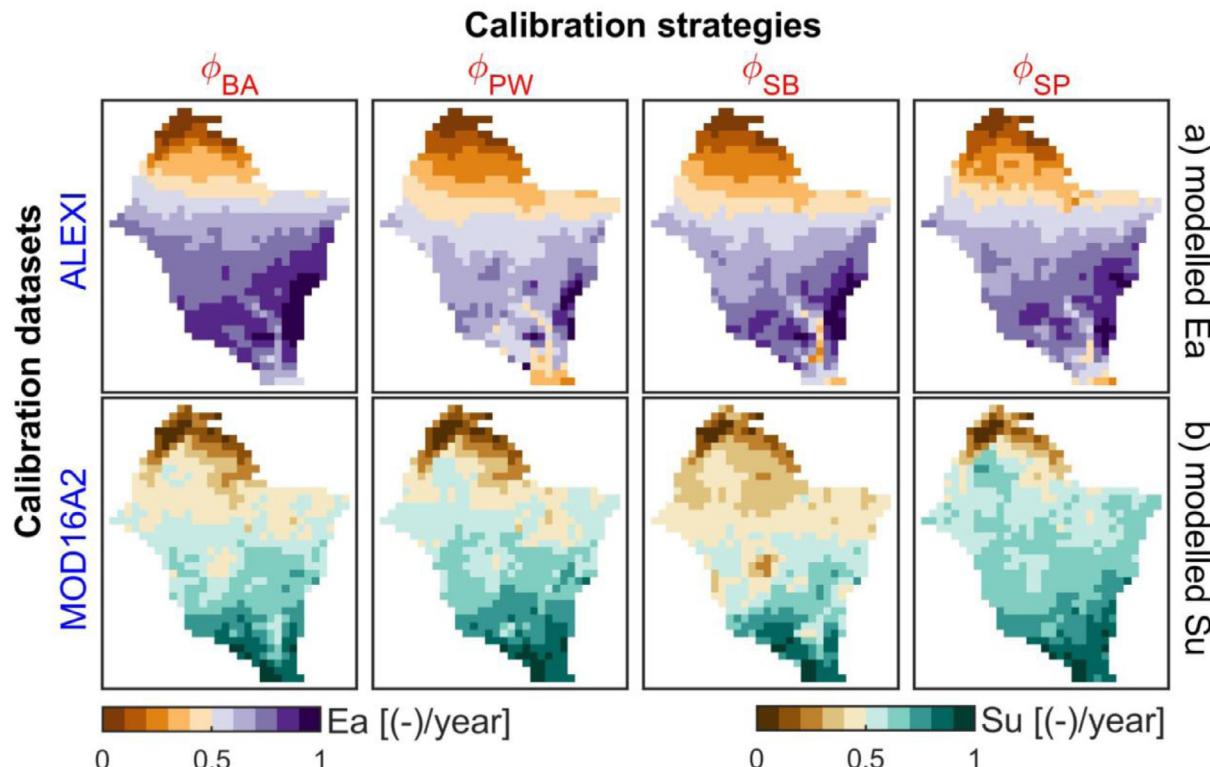


Fig. 8. Maps of long-term (2003-2012) average of annual (a) actual evaporation (E_a) and (b) soil moisture (S_u) obtained by calibrating the mHM model with two evaporation datasets (y-axis, blue font) and different calibration strategies (x-axis, red font). A min-max normalization of the values allows rescaling them between 0 and 1 to emphasize the spatial patterns and use a unique color scale.

S12). In general, the south to north gradient of increasing aridity observed with modelled E_a and S_u is well depicted for all the calibration strategies. However, considerable mismatches in the variability of the patterns are observed among the calibration strategies. Such discrepancies in spatial patterns have implications for water resources assessment including water accounting, flood and drought monitoring and prediction (AghaKouchak et al., 2015; Klemas, 2014; Teng et al., 2017; West et al., 2019). Knowing when flood or drought events occur is important,

but knowing the spatial extent of the event is crucial for deploying efficient adaptation and mitigation strategies (Brunner et al., 2020; Diaz et al., 2019; He et al., 2020). Consequently, improving the representation of the spatial patterns of hydrological processes should be a key consideration in modelling with spatially distributed models. The comparison of the maps of modelled E_a (Fig. 8a) with the reference ALEXI dataset (Fig. 3) reveals that the spatial patterns of E_a in the case BA show the highest mismatch with the reference, thereby unveiling the

Variables	VRB				Sahelian zone			Sudano-Sahelian zone			Sudanian zone			Guinean zone			
	Temporal dynamics		Spatial patterns		Temporal dynamics		Spatial patterns		Temporal dynamics		Spatial patterns		Temporal dynamics		Spatial patterns		
	Q	St	Su	Su	St	Su	Su	St	Su	Su	St	Su	Su	St	Su	Su	
Performance metrics	EKG	r	r	Esp	r	r	Esp	r	r	Esp	r	r	Esp	r	r	Esp	
Q + Evaporation datasets	Q only	0.694	0.730	0.863	-0.177	0.729	0.910	-0.358	0.725	0.884	-0.349	0.747	0.856	-0.398	0.597	0.777	-0.495
	MOD16A2	0.676	0.763	0.908	-0.044	0.702	0.934	-0.372	0.750	0.923	-0.351	0.818	0.907	-0.357	0.686	0.831	-0.410
	SSEBop	0.673	0.868	0.920	0.018	0.714	0.924	-0.461	0.845	0.933	-0.331	0.925	0.925	-0.384	0.827	0.855	-0.378
	ALEXI	0.611	0.840	0.919	-0.043	0.734	0.936	-0.485	0.829	0.932	-0.338	0.893	0.921	-0.363	0.783	0.857	-0.474
	CMRSET	0.639	0.775	0.914	-0.037	0.684	0.936	-0.441	0.768	0.924	-0.353	0.833	0.916	-0.379	0.753	0.861	-0.361
	SEBS	0.560	0.768	0.892	-0.068	0.702	0.914	-0.480	0.786	0.903	-0.348	0.829	0.893	-0.416	0.717	0.840	-0.292
	GLEAM v3.2a	0.690	0.823	0.925	0.002	0.611	0.938	-0.465	0.784	0.936	-0.338	0.893	0.930	-0.348	0.803	0.863	-0.387
	GLEAM v3.3a	0.696	0.833	0.928	0.011	0.604	0.938	-0.457	0.799	0.938	-0.332	0.894	0.931	-0.369	0.803	0.864	-0.403
	GLEAM v3.2b	0.711	0.810	0.920	0.003	0.631	0.939	-0.488	0.784	0.932	-0.326	0.878	0.923	-0.348	0.777	0.852	-0.415
	GLEAM v3.3b	0.709	0.811	0.923	0.010	0.647	0.938	-0.482	0.785	0.934	-0.336	0.879	0.925	-0.341	0.774	0.852	-0.414
Calibration strategies	ERA5	0.508	0.741	0.889	-0.089	0.724	0.919	-0.434	0.761	0.900	-0.356	0.797	0.883	-0.405	0.683	0.835	-0.297
	MERRA-2	0.677	0.868	0.927	0.011	0.657	0.945	-0.484	0.835	0.940	-0.334	0.927	0.932	-0.366	0.832	0.864	-0.447
	JRA-55	0.616	0.835	0.909	-0.013	0.673	0.928	-0.431	0.815	0.920	-0.326	0.885	0.911	-0.376	0.780	0.851	-0.380

Fig. 9. Model performance of streamflow (Q), terrestrial water storage (S_t) and soil moisture (S_u) using various evaporation datasets and model calibration strategies with the mHM model. The results are shown for the four climatic zones in the Volta River Basin (VRB) over the simulation period (2003–2012). Each score for a given evaporation product represents the average over the scores obtained with 4 calibration strategies, while the score is the average over scenarios with 12 evaporation datasets for each calibration strategy. The skill scores of the temporal dynamics are obtained with the Kling-Gupta efficiency (E_{KG}) for Q and the Pearson's correlation coefficient (r) for S_t and S_u . The spatial pattern efficiency (E_{SP}) is used to assess the spatial representation of S_u . The benchmark model performance is given for the Q-only calibration as a reference. The skill scores are ranked from the highest (bluest) to the lowest (reddest).

potential pitfalls of the basin-average calibration. The highest improvement of the spatial patterns of S_u is obtained with case SB (+101%), followed by case SP (+93%) and case PW (+85%), while case BA (+76%) provides the least improvement (Fig. 7b).

3.3. Analysis per climatic zone

The analysis of the model performance according to climatic zones is done for all hydrological processes except Q because of the few gauging stations, which are unevenly distributed across the VRB (cf. Fig. 1). The results reveal contrasting model performances across climatic zones (Fig. 9). Additional results are provided in the supplementary material (Figures S3-S11).

In average, the model performance at predicting the temporal dynamics of S_t and the spatial patterns of S_u is higher in the intermediate climatic regions (i.e., Sudano-Sahelian and Sudanian zones) than in the driest and wettest regions (i.e., Sahelian and Guinean zones) of the VRB. In terms of the temporal dynamics of S_u , the model performance decreases slightly from the driest to the wettest regions.

In general, the multivariate calibration scenarios with evaporation datasets lead to a higher model performance for all hydrological processes in all climatic zones, as compared to the Q-only calibration (Fig. 9, Figure S13). Although the GLEAM products globally perform well, they are the least effective at predicting the temporal dynamics of S_t in the Sahelian zone. The top three best evaporation datasets for improving the average model performance across climatic zones are SSEBop, MERRA-2 and ALEXI for the temporal dynamics of S_t , MERRA-2, GLEAM v3.3a and GLEAM v3.2a for the temporal dynamics of S_u , and GLEAM v3.3b, GLEAM v3.3a and GLEAM v3.2a for the spatial patterns of S_u . Similar to results obtained at the entire VRB scale (Section 3.1), MOD16A2, SEBS and ERA5 still show the lowest contribution in improving the model performance across different climatic zones. More details on the ranking of the evaporation datasets are provided in the supplementary material (Tables S1-S5; Figures S14, S21 and S40).

Contrary to the basin scale analysis, case SP is the least efficient calibration strategy per climatic zones. This result can be justified by the fact that the model calibration on the spatial patterns of evaporation datasets is done at the scale of the VRB. Consequently, the spatial variability of hydrological processes at large scale is not representative of

the climatic zones where the patterns are more homogenous. For regions with strong spatial variability, sub-region model calibration on spatial patterns can be a way forward in overcoming the pitfalls of domain-wide calibration, thereby ultimately resulting in a higher model performance.

4. Discussions

The results presented are primarily valid for the VRB because errors in evaporation estimates are known to vary according to region (Hartanto et al., 2017; Sörensson and Ruscica, 2018). However, the innovative model evaluation approach proposed in this study is not location-specific and can be applied to other regions using a grid-based hydrological model. The robust evaluation approach for evaporation datasets is based on multiple hydrological processes, and responds to the recent call of the scientific community for process-oriented diagnostics of earth system models (Maloney et al., 2019; Melsen et al., 2016). The same methodology can be deployed for the evaluation of other spatial observational datasets such as gridded soil moisture estimates using robust metrics (Dong et al., 2019), and subsequently be used to estimate their value for hydrological modelling. Moreover, this study highlights the potential of satellite and reanalysis evaporation datasets to improve the representation of various hydrological processes, which might guide modellers in choosing the adequate product for their applications, and support the data developers in their quest of improving global estimates of evaporation (McCabe et al., 2019). It must be noted that the overall ranking in Fig. 9 does not systematically determine whether a dataset is good or bad, rather it shows which evaporation product provides the highest or lowest model performance for a given hydrological flux or state variable. Only the skill scores allow a judgement on the ability of a given dataset to yield a good model performance.

There might be uncertainties related to the rescaling of the evaporation datasets from their native spatial resolutions to that of the hydrological modelling resolution. However, for a fair comparison, all the datasets should be used at the same spatiotemporal resolution. The high performance of the GLEAM datasets is most likely due to the integration of soil moisture information in the calculation of actual evaporation. Therefore, GLEAM is expected to perform well for hydrological modelling. Surprisingly, MERRA-2 ranks among the best evaporation

datasets notwithstanding its coarse spatial resolution. The high performance of MERRA-2 can be explained by its high temporal resolution, which might compensate its lower spatial resolution. In general, there is no clear tendency of SRS datasets to outperform the reanalysis datasets, and vice versa. Thus, besides their use as forcing data, reanalysis datasets represent a valuable source of information for the calibration of hydrological models. Satellite data of soil moisture and terrestrial water storage used for model evaluation in this study are not free of errors. However, at large scale and in poorly gauged regions, they are the only source of data that can be used for spatial model evaluation (Peters-Lidard et al., 2019).

Overall, this study strives to provide solutions to some of the current challenges in hydrology (i.e., modelling methods, uncertainty and spatial variability; Blöschl et al., 2019). The proposed methodology represents an innovative way to use satellite and reanalysis datasets to improve process representation in hydrological models (Clark et al., 2015; Peters-Lidard et al., 2017), develop hydrological water accounting (Hunink et al., 2019) and advance prediction in ungauged basins (Hrachowitz et al., 2013; Zhang et al., 2019).

Future studies should test the advantages of a multi-scale calibration framework that accounts for both domain-wide and sub-domain spatial heterogeneity, which might lead to better prediction of spatial patterns across climatic zones at large scale. Calibration strategies with multiple non-commensurable variables as well as spatial patterns is a way forward in advancing process representation in hydrological models (Dembélé et al., 2020a; Nijzink et al., 2018; Zink et al., 2018). More importantly, the choice of the calibration strategy or the objective function is determinant for high model performance, mainly under a changing environment (Fowler et al., 2018; Schaeefli et al., 2010). Uncertainties in the structure of the mHM model might influence the modelled hydrological processes. For instance, the calculation of actual evaporation (Section 2.1) might be subject to a double accounting of stress from both the leaf area index and soil moisture. However, in such case, the sources of stress can be proportionally adjusted by the model parameters during the calibration. Further work should investigate the impact of different model structures on the methodology proposed in this study, and assess the parameter sensitivity of mHM depending on calibration strategies and variables, which is beyond the scope of the current study. Besides improving model calibration, satellite and reanalysis datasets can play an important role in identifying deficiencies in model structures and contribute to model improvement (Hulsman et al., 2020). As the accuracy of the satellite and reanalysis datasets depends on the quality of the input meteorological datasets used for their production, it is also important to assess the impact of precipitation datasets on evaporation modelling (Dembélé et al., 2020b; Mao and Wang, 2017; Or and Lehmann, 2019).

Finally, an ensemble product that merges different evaporation datasets is a potential way forward in reducing regional uncertainties and thereby improving global estimates (da Motta Paca et al., 2019; Jiménez et al., 2018). Such advances in evaporation product development can facilitate prediction in ungauged basins using earth observations.

5. Conclusion

Four model calibration strategies are used to evaluate twelve satellite and reanalysis datasets in the large transboundary Volta River basin located in West Africa. The experiment is done with the mHM model over the period 2003–2012. The key findings can be summarized as follows:

- Satellite and reanalysis datasets can improve the predictive skill of the hydrological model if the appropriate calibration strategy is used.
- Overall, MERRA-2, GLEAM v3.3a and SSEBop individually provide the highest contribution in improving the model performance.

- Model calibration on the full extent of the gridded evaporation datasets result in a higher model performance than calibration on basin-average estimates.
- Using only the spatial patterns of gridded evaporation data for model calibration, and not their absolute values, yields higher model performance than classical approaches based on basin average evaporation signal or based only on streamflow.
- Contrasting spatial patterns of soil moisture are obtained depending on the modelling scenarios, with differences in the model performances according to climatic zones.

These findings contribute to solving current challenges related to large-scale hydrological modelling and provide avenues for improving process representation with the use of increasingly available satellite and reanalysis datasets. Improving the representation of the spatial patterns of hydrological processes should be a key consideration in modelling with spatially distributed models, which would allow better predicting floods and droughts. Moreover, the results provide insights to the developers of the evaporation datasets and might serve of guidance for future developments. However, a replication of the proposed methodology to evaluate evaporation datasets should be applied in other regions with different hydro-climatic conditions, and with different hydrological and land surface models. Future work should also investigate the possibility of prediction in ungauged basins solely from earth observation datasets, which are increasingly and readily becoming available.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Moctar Dembélé: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Natalie Ceperley:** Conceptualization, Funding acquisition, Writing - review & editing. **Sander J. Zwart:** Conceptualization, Funding acquisition, Writing - review & editing. **Elga Salvadore:** Conceptualization, Resources, Writing - review & editing. **Gregoire Mariethoz:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing. **Bettina Schaeefli:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Acknowledgments and Data

We thank the providers of the datasets used in this study (see Tables 2 and 3). We are grateful to the streamflow data providers (DGRE of Burkina Faso, HSD of Ghana and the Volta Basin Authority). Moctar Dembélé was supported by the Swiss Government Excellence Scholarship (2016.0533 / Burkina Faso / OP) and the Doc.Mobility fellowship (SNF, P1LAP2_178071) of the Swiss National Science Foundation. Bettina Schaeefli and Natalie Ceperley were supported by a research grant of the Swiss National Science Foundation (SNF, PP00P2_157611). We thank the three anonymous reviewers for their constructive feedback. The entire modelling database can be freely accessed at <http://doi.org/10.5281/zenodo.3800416>

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.advwatres.2020.103667](https://doi.org/10.1016/j.advwatres.2020.103667).

References

- Addor, N., Melsen, L., 2019. Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resour. Res.* 55 (1), 378–390. <https://doi.org/10.1029/2018WR022295>.
- AghaKouchak, A., Farahmand, A., Melton, F.S., Teixeira, J., Anderson, M.C., Wardlow, B.D., Hain, C.R., 2015. Remote sensing of drought: Progress, challenges and opportunities. *Rev. Geophys.* 53 (2), 452–480. <https://doi.org/10.1002/2014rg000456>.
- Allen, R.G., L.S. Pereira, D.Raes, and M. Smith (1998), Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper56, 326 pp, <http://academic.uprm.edu/abe/backup2/tomas/fao%2056.pdf>.
- Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ.* 60 (2), 195–216. [https://doi.org/10.1016/S0034-4257\(96\)00215-5](https://doi.org/10.1016/S0034-4257(96)00215-5).
- Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *J. Geophys. Res. Atmos.* 112 (D10). <https://doi.org/10.1029/2006JD007506>.
- Armstrong, R.N., Pomeroy, J.W., Martz, L.W., 2019. Spatial variability of mean daily estimates of actual evaporation from remotely sensed imagery and surface reference data. *Hydrol. Earth Syst. Sc.* 23 (12), 4891–4907. <https://doi.org/10.5194/hess-23-4891-2019>.
- Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. *J. Hydrometeorol.* 16 (4), 1449–1455. <https://doi.org/10.1175/JHM-D-14-0040.1>.
- Bai, P., Liu, X., Yang, T., Li, F., Liang, K., Hu, S., Liu, C., 2016. Assessment of the influences of different potential evapotranspiration inputs on the performance of monthly hydrological models under different climatic conditions. *J. Hydrometeorol.* 17 (8), 2259–2274. <https://doi.org/10.1175/JHM-D-15-0202.1>.
- Bai, P., Liu, X., 2018. Intercomparison and evaluation of three global high-resolution evapotranspiration products across China. *J. Hydrol.* 566, 743–755. <https://doi.org/10.1016/j.jhydrol.2018.09.065>.
- Bai, P., Liu, X., Zhang, Y., Liu, C., 2018. Incorporating vegetation dynamics noticeably improved performance of hydrological model under vegetation greening. *Sci. Total Environ.* 643, 610–622. <https://doi.org/10.1016/j.scitotenv.2018.06.233>.
- Baik, J., Liaqat, U.W., Choi, M., 2018. Assessment of satellite-and reanalysis-based evapotranspiration products with two blending approaches over the complex landscapes and climates of Australia. *Agricul. For. Meteorol.* 263, 388–398. <https://doi.org/10.1016/j.agrformet.2018.09.007>.
- Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., 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. Sc.* 21 (12), 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>.
- Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., Schüth, C., 2019. Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived ET data. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2019.123944>.
- Beven, K., 2006. A manifesto for the equifinality thesis. *J. Hydrol.* 320 (1-2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>.
- Bhattarai, N., Mallick, K., Stuart, J., Vishwakarma, B.D., Niraula, R., Sen, S., Jain, M., 2019. An automated multi-model evapotranspiration mapping framework using remotely sensed and reanalysis data. *Remote Sens. Environ.* 229, 69–92. <https://doi.org/10.1016/j.rse.2019.04.026>.
- Birhanu, D., Kim, H., Jang, C., 2019. Effectiveness of introducing crop coefficient and leaf area index to enhance evapotranspiration simulations in hydrologic models. *Hydro. Process.* <https://doi.org/10.1002/hyp.13464>.
- 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.* <https://doi.org/10.1002/hyp.13791>.
- Blöschl, G., Bierkens, M.F., Chamberl, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H., Sivapalan, M., 2019. Twenty-three Unsolved Problems in Hydrology (UPH)-a community perspective. *Hydrol. Sci. J.* <https://doi.org/10.1080/02626667.2019.1620507>.
- Bontemps, S., P.Defourny, E. V.Bogaert, O.Arino, V.Kalogirou, and J. R.Perez (2011), GLOBCOVER 2009-Products description and validation report, 53 pp, http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf.
- Bosilovich, M.G., Chen, J., Robertson, F.R., Adler, R.F., 2008. Evaluation of global precipitation in reanalyses. *J. Appl. Meteorol. Clim.* 47 (9), 2279–2299. <https://doi.org/10.1175/2008jams1921.1>.
- Bouaziz, L.J.E., et al., 2020. Behind the scenes of streamflow model performance. *Hydrol. Earth Syst. Sci. Discuss.* 2020, 1–38. <https://doi.org/10.5194/hess-2020-176>.
- Brunner, M.I., Melsen, L.A., Wood, A.W., Rakovec, O., Mizukami, N., Knoben, W.J.M., Clark, M.P., 2020. Flood hazard and change impact assessments may profit from re-thinking model calibration strategies. *Hydrol. Earth Syst. Sci. Discuss.* 2020, 1–19. <https://doi.org/10.5194/hess-2020-192>.
- Ceperley, N.C., Mande, T., van de Giesen, N., Tyler, S., Yacouba, H., Parlange, M.B., 2017. Evaporation from cultivated and semi-wild Sudanian Savanna in west Africa. *Hydrol. Earth Syst. Sc.* 21, 4149–4167. <https://doi.org/10.5194/hess-21-4149-2017>.
- Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sens. Environ.* 237, 111594. <https://doi.org/10.1016/j.rse.2019.111594>.
- Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., Li, X., Li, X., Liu, S., Ma, Z., Miyata, A., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *Remote Sens. Environ.* 140, 279–293. <https://doi.org/10.1016/j.rse.2013.08.045>.
- Clark, M.P., et al., 2015. Improving the representation of hydrologic processes in Earth System Models. *Water Resour. Res.* 51 (8), 5929–5956. <https://doi.org/10.1002/2015wr017096>.
- Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R.N., Cai, X.T., Wood, A.W., Peters-Lidard, C.D., 2017. The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. *Hydrol. Earth Syst. Sc.* 21 (7), 3427–3440. <https://doi.org/10.5194/hess-21-3427-2017>.
- Coenders-Gerrits, M., Schipperoort, B., Jiménez-Rodríguez, C., 2020. *Evaporative Processes on Vegetation: An Inside Look*. In: Van, J.T., Gutmann, Stan, E., Friesen, J. (Eds.), *Precipitation Partitioning by Vegetation*. Springer, Switzerland, pp. 35–48.
- Cui, X., Guo, X., Wang, Y., Wang, X., Zhu, W., Shi, J., Lin, C., Gao, X., 2019. Application of remote sensing to water environmental processes under a changing climate. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2019.04.078>.
- Cunge, J.A., 1969. On the subject of a flood propagation computation method (Muskingum method). *J. Hydraul. Res.* 7 (2), 205–230. <https://doi.org/10.1080/00221686909500264>.
- Paca, da Motta, H., V., Espinoza-Dávalos, G.E., Hessels, T.M., Moreira, D.M., Comair, G.F., Bastiaanssen, W.G., 2019. The spatial variability of actual evapotranspiration across the Amazon River Basin based on remote sensing products validated with flux towers. *Ecol. Process.* 8 (1), 6. <https://doi.org/10.1186/s13717-019-0158-8>.
- Danielson, J. J., and D. Geshc (2011), Global multi-resolution terrain elevation data 2010 (GMTED2010)Rep. 2331-1258, 34 pp, US Geological Survey, <https://doi.org/10.3133/ofr20111073>.
- Dembélé, M., Zwart, S.J., 2016. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *Int. J. Remote Sens.* 37 (17), 3995–4014. <https://doi.org/10.1080/01431161.2016.1207258>.
- Dembélé, M., Oriani, F., Tumbulto, J., Mariéthoz, G., Schaeffli, B., 2019. Gap-filling of daily streamflow time series using Direct Sampling in various hydroclimatic settings. *J. Hydrol.* 569, 573–586. <https://doi.org/10.1016/j.jhydrol.2018.11.076>.
- Dembélé, M., Hrachowitz, M., Savenije, H.H., Mariéthoz, G., Schaeffli, B., 2020a. Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite datasets. *Water Resour. Res.* <https://doi.org/10.1029/2019WR026085>.
- Dembélé, M., Schaeffli, B., van de Giesen, N., Mariéthoz, G., 2020b. *Suitability of 17 rainfall and temperature gridded datasets for large-scale hydrological modelling*, 2020. *Hydrol. Earth Syst. Sci. Discuss.*, West Africa, pp. 1–39.
- Demirel, M.C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., Stisen, S., 2018. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrol. Earth Syst. Sc.* 22 (2), 1299–1315. <https://doi.org/10.5194/hess-22-1299-2018>.
- Diaz, V., Perez, G.A.C., Van Lanen, H.A., Solomatine, D., Varouchakis, E.A., 2019. Characterisation of the dynamics of past droughts. *Sci. Total Environ.*, 134588. <https://doi.org/10.1016/j.scitotenv.2019.134588>.
- Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A., Maher, N., 2016. More extreme precipitation in the world's dry and wet regions. *Nat. Clim. Change* 6 (5), 508–513. <https://doi.org/10.1038/nclimate2941>.
- Dong, J., Crow, W., Reichle, R., Liu, Q., Lei, F., Cosh, M.H., 2019. A Global Assessment of Added Value in the SMAP Level 4 Soil Moisture Product Relative to Its Baseline Land Surface Model. *Geophys. Res. Lett.* 46 (12), 6604–6613. <https://doi.org/10.1029/2019GL083398>.
- Dorigo, W., et al., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* 203, 185–215. <https://doi.org/10.1016/j.rse.2017.07.001>.
- Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrolog. Sci. J.* 55 (1), 58–78. <https://doi.org/10.1080/02626660903526292>.
- Er-Raki, S., Chehbouni, A., Khabba, S., Simonneaux, V., Jarlan, L., Ouldbba, A., Rodriguez, J., Allen, R., 2010. Assessment of reference evapotranspiration methods in semi-arid regions: can weather forecast data be used as alternate of ground meteorological parameters? *J. Arid Environ.* 74 (12), 1587–1596. <https://doi.org/10.1016/j.jaridenv.2010.07.002>.
- Famiglietti, J.S., Rodell, M., 2013. Water in the balance. *Science* 340 (6138), 1300–1301. <https://doi.org/10.1126/science.1236460>.
- FAO/GIEWS (1998), Sahel Weather And Crop Situation 1998, 8 pp, Food and Agriculture Organization (FAO) / Global Information and Early Warning System (GIEWS), <http://www.fao.org/docrep/004/x0059e/x0059e00.htm>.
- Faticchi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J. Hydrol.* 537, 45–60. <https://doi.org/10.1016/j.jhydrol.2016.03.026>.
- Feddes, R.A., Kowalik, P., Kolinskamalinka, K., Zaradny, H., 1976. Simulation of Field Water-Uptake by Plants Using a Soil-Water Dependent Root Extraction Function. *J. Hydrol.* 31 (1-2), 13–26. [https://doi.org/10.1016/0022-1694\(76\)90017-2](https://doi.org/10.1016/0022-1694(76)90017-2).
- Feng, T., Su, T., Zhi, R., Tu, G., Ji, F., 2019. Assessment of actual evapotranspiration variability over global land derived from seven reanalysis datasets. *Int. J. Climatol.* 39 (6), 2919–2932. <https://doi.org/10.1002/joc.5992>.
- Fisher, J.B., 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 (4), 2618–2626. <https://doi.org/10.1002/2016wr020175>.
- Fowler, K., Peel, M., Western, A., Zhang, L., 2018. Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function. *Water Resour. Res.* 54 (5), 3392–3408. <https://doi.org/10.1029/2017WR022466>.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., 2015. The climate hazards infrared precipitation with

- stations—a new environmental record for monitoring extremes. *Scientific Data* 2, 150066. <https://doi.org/10.15780/G2RP4Q>.
- Gao, F., Feng, G., Ouyang, Y., Wang, H., Fisher, D., Adeli, A., Jenkins, J., 2017. Evaluation of reference evapotranspiration methods in arid, semiarid, and humid regions. *JAWRA J. Am. Water Resour. Assoc.* 53 (4), 791–808. <https://doi.org/10.1111/1752-1688.12530>.
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Climate* 30 (14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gruber, A., Dorigo, W.A., Crow, W., Wagner, W., 2017. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE T. Geosci. Remote* 55 (12), 6780–6792. <https://doi.org/10.1109/Tgrs.2017.2734070>.
- Guerschman, J.P., Van Dijk, A.I., Mattersdorf, G., Beringer, J., Hutley, L.B., Leuning, R., Pipunic, R.C., Sherman, B.S., 2009. Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. *J. Hydrol.* 369 (1-2), 107–119. <https://doi.org/10.1016/j.jhydrol.2009.02.013>.
- Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., Onoda, H., Onogi, K., Miyaoka, K., Takahashi, K., 2016. The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability. *J. Meteorol. Soc. Jpn. Ser. II* 94 (3), 269–302. <https://doi.org/10.2151/jmsj.2016-015>.
- Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Appl. Eng. Agricult.* 1 (2), 96–99. <https://doi.org/10.13031/2013.26773>.
- Hartanto, I.M., Van Der Kwast, J., Alexandridis, T.K., Almeida, W., Song, Y., van Andel, S., Solomatine, D., 2017. Data assimilation of satellite-based actual evapotranspiration in a distributed hydrological model of a controlled water system. *Int. J. Appl. Earth Observ. Geoinform.* 57, 123–135. <https://doi.org/10.1016/j.jag.2016.12.015>.
- Hartmann, J., Moosdorf, N., 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochem. Geophys. Geosy.* 13 (12). <https://doi.org/10.1029/2012gc004370>.
- He, M., Kimball, J.S., Yi, Y., Running, S.W., Guan, K., Moreno, A., Wu, X., Maneta, M., 2019. Satellite data-driven modeling of field scale evapotranspiration in croplands using the MOD16 algorithm framework. *Remote Sens. Environ.* 230, 111201. <https://doi.org/10.1016/j.rse.2019.05.020>.
- He, X., Pan, M., Wei, Z., Wood, E.F., Sheffield, J., 2020. A Global Drought and Flood Catalogue from 1950 to 2016. *Bull. Am. Meteorol. Soc.* 2020. <https://doi.org/10.1175/BAMS-D-18-0269.1>.
- Hengl, T., et al., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12 (2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>.
- Herman, M.R., Nejadhashemi, A.P., Abouali, M., Hernandez-Suarez, J.S., Daneshvar, F., Zhang, Z., Anderson, M.C., Sadeghi, A.M., Hain, C.R., Sharifi, A., 2018. Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability. *J. Hydrol.* 556, 39–49. <https://doi.org/10.1016/j.jhydrol.2017.11.009>.
- Hersbach, H., et al., 2018. Operational global reanalysis: progress, future directions and synergies with NWP. *ERA Report Series* 27, ECMWF, Reading. European Centre for Medium Range Weather Forecasts, UK <https://www.ecmwf.int/node/18765>.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., 2020. The ERA5 global reanalysis. *Quart. J. R. Meteorol. Soc.* <https://doi.org/10.1002/qj.3803>.
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., 2013. A decade of Predictions in Un-gauged Basins (PUB)—a review. *Hydrol. Sci. J.* 58 (6), 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>.
- Hrachowitz, M., Clark, M.P., 2017. HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrol. Earth Syst. Sc.* 21 (8), 3953–3973. <https://doi.org/10.5194/hess-21-3953-2017>.
- Hulsmans, P., Savenije, H.H.G., Hrachowitz, M., 2020. Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement. *Hydrol. Earth Syst. Sci. Discuss.* 2020, 1–26. <https://doi.org/10.5194/hess-2020-191>.
- Hunink, J., Simons, G., Suárez-Almiñana, S., Solera, A., Andreu, J., Giuliani, M., Zamberletti, P., Grillakis, M., Koutoulis, A., Tsanis, I., 2019. A simplified water accounting procedure to assess climate change impact on water resources for agriculture across different European river basins. *Water* 11 (10), 1976. <https://doi.org/10.3390/w11101976>.
- Huntington, T.G., 2006. Evidence for intensification of the global water cycle: review and synthesis. *J. Hydrol.* 319 (1-4), 83–95. <https://doi.org/10.1016/j.jhydrol.2006.07.003>.
- Immerzeel, W.W., Droogers, P., 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. *J. Hydrol.* 349 (3-4), 411–424. <https://doi.org/10.1016/j.jhydrol.2007.11.017>.
- Jiang, D., Wang, K., 2019. The Role of Satellite-Based Remote Sensing in Improving Simulated Streamflow: A Review. *Water* 11 (8), 1615. <https://doi.org/10.3390/w11081615>.
- Jiang, L., Wu, H., Tao, J., Kimball, J.S., Alfieri, L., Chen, X., 2020. Satellite-Based Evapotranspiration in Hydrological Model Calibration. *Remote Sensing* 12 (3), 428. <https://doi.org/10.3390/rs12030428>.
- Jiao, Y., Lei, H.M., Yang, D.W., Huang, M.Y., Liu, D.F., Yuan, X., 2017. Impact of vegetation dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology coupled model. *J. Hydrol.* 551, 116–131. <https://doi.org/10.1016/j.jhydrol.2017.05.060>.
- Jimenez, C., Prigent, C., Mueller, B., Seneviratne, S.I., McCabe, M., Wood, E.F., Rossow, W., Balsamo, G., Betts, A., Dirmeyer, P., 2011. Global intercompari- son of 12 land surface heat flux estimates. *J. Geophys. Res. Atmos.* 116 (D2). <https://doi.org/10.1029/2010JD014545>.
- Jiménez, C., Martens, B., Miralles, D.M., Fisher, J.B., Beck, H.E., Fernández-Prieto, D., 2018. Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. *Hydrol. Earth Syst. Sc.* 22 (8). <https://doi.org/10.5194/hess-22-4513-2018>.
- Joiner, J., Yoshida, Y., Anderson, M., Holmes, T., Hain, C., Reichle, R., Koster, R., Middleton, E., Zeng, F.-W., 2018. Global relationships among traditional reflectance vegetation indices (NDVI and NDII), evapotranspiration (ET), and soil moisture variability on weekly timescales. *Remote Sens. Environ.* 219, 339–352. <https://doi.org/10.1016/j.rse.2018.10.020>.
- Jung, H.C., Getirana, A., Arsenault, K.R., Holmes, T.R., McNally, A., 2019. Uncertainties in Evapotranspiration Estimates over West Africa. *Remote Sensing* 11 (8), 892. <https://doi.org/10.3390/rs11080892>.
- 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. *Agricul. For. Meteorol.* 252, 256–268. <https://doi.org/10.1016/j.agrmet.2018.01.022>.
- Khu, S.T., Madsen, H., 2005. Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resour. Res.* 41 (3). <https://doi.org/10.1029/2004WR003041>.
- Klemas, V., 2014. Remote sensing of floods and flood-prone areas: an overview. *J. Coast. Res.* 31 (4), 1005–1013. <https://doi.org/10.2112/JCOASTRES-D-14-00160.1>.
- 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>.
- Knoben, W.J., Freer, J.E., Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol. Earth Syst. Sc.* 23 (10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn. Ser. II* 93 (1), 5–48. <https://doi.org/10.2151/jmsj.2015-001>.
- Koch, J., Demirel, M.C., Stisen, S., 2018. The SPAcial EFFiciency metric (SPEAF): multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geosci. Model. Dev.* 11 (5), 1873–1886. <https://doi.org/10.5194/gmd-11-1873-2018>.
- Koppa, A., Gebremichael, M., 2020. Improving the Applicability of Hydrologic Models for Food-Energy-Water Nexus Studies Using Remote Sensing Data. *Remote Sensing* 12 (4), 599. <https://doi.org/10.3390/rs12040599>.
- Kumar, R., 2010. Distributed Hydrologic Model Parameterization: Application in a Mesoscale River Basin, edited Helmholtz Centre for Environmental Research - UFZ <https://www.ufz.de/index.php?en=2093&ufzPublicationIdentifier=11006>.
- Kumar, R., Samaniego, L., Attinger, S., 2013. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resour. Res.* 49 (1), 360–379. <https://doi.org/10.1029/2012wr012195>.
- Kunnath-Poovakka, A., Ryu, D., Renzullo, L., George, B., 2016. The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction. *J. Hydrol.* 535, 509–524. <https://doi.org/10.1016/j.jhydrol.2016.02.018>.
- Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* 48 (4). <https://doi.org/10.1029/2011wr011453>.
- Liou, Y.-A., Kar, S.K., 2014. Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—A review. *Energies* 7 (5), 2821–2849. <https://doi.org/10.3390/en7052821>.
- 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>.
- Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* 123, 280–297. <https://doi.org/10.1016/j.rse.2012.03.014>.
- Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.* 50 (2), 1131–1151. <https://doi.org/10.1002/2013WR014581>.
- López, O., Houborg, R., McCabe, M.F., 2017. Evaluating the hydrological consistency of evaporation products using satellite-based gravity and rainfall data. *Hydrol. Earth Syst. Sc.* 21 (1), 323–343. <https://doi.org/10.5194/hess-21-323-2017>.
- 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>.
- Maloney, E.D., Gettelman, A., Ming, Y., Neelin, J.D., Barrie, D., Mariotti, A., Chen, G.-C., Coleman, D.R., Kuo, Y.-H., Singh, B., 2019. Process-Oriented Evaluation of Climate and Weather Forecasting Models. *Bull. Am. Meteorol. Soc.* 100 (9), 1665–1686. <https://doi.org/10.1175/BAMS-D-18-0042.1>.
- Mao, Y., Wang, K., 2017. Comparison of evapotranspiration estimates based on the surface water balance, modified Penman-Monteith model, and reanalysis data sets for continental China. *J. Geophys. Res. Atmos.* 122 (6), 3228–3244. <https://doi.org/10.1002/2016JD026065>.
- Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernandez-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model. Dev.* 10 (5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>.
- Martens, B., Waegeman, W., Dorigo, W.A., Verhoest, N.E., Miralles, D.G., 2018. Terrestrial evaporation response to modes of climate variability. *NPJ Climate Atmos. Sci.* 1 (1), 1–7. <https://doi.org/10.1038/s41612-018-0053-5>.

- McCabe, M., Wood, E.F., Wójcik, R., Pan, M., Sheffield, J., Gao, H., Su, H., 2008. Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies. *Remote Sens. Environ.* 112 (2), 430–444. <https://doi.org/10.1016/j.rse.2007.03.027>.
- McCabe, M., A.Ershadi, C.Jimenez, D. G.Miralles, D.Michel, and E. F.Wood (2015), The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data, <https://doi.org/10.5194/gmd-9-283-2016>.
- McCabe, M.F., et al., 2017. The Future of Earth Observation in Hydrology. *Hydrol. Earth Syst. Sc.* 21 (7), 3879–3914. <https://doi.org/10.5194/hess-21-3879-2017>.
- McCabe, M.F., Miralles, D.G., Holmes, T.R., Fisher, J.B., 2019. Advances in the remote sensing of terrestrial evaporation, edited Multidisciplinary Digital Publishing Institute.
- McMahon, T., Peel, M., Lowe, L., Srikanthan, R., McVicar, T., 2013. Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. *Hydrol. Earth Syst. Sc.* 17 (4), 1331–1363. <https://doi.org/10.5194/hess-17-1331-2013>.
- Melsen, L.A., Teuling, A.J., Torfs, P.J., Uijlenhoet, R., Mizukami, N., Clark, M.P., 2016. HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrol. Earth Syst. Sc.* 20 (3), 1069–1079. <https://doi.org/10.5194/hess-20-1069-2016>.
- Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M.F., Fisher, J.B., Mu, Q., 2016. The WACMOS-ET project-Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. *Hydrol. Earth Syst. Sc.* 20 (2), 803–822. <https://doi.org/10.5194/hess-20-803-2016>.
- Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sc.* 15 (2), 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., 2016. The WACMOS-ET project-Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrol. Earth Syst. Sc.* 20 (2), 823–842. <https://doi.org/10.5194/hess-20-823-2016>.
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* 115 (8), 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>.
- Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J., Guo, Z., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.* 38 (6). <https://doi.org/10.1029/2010GL046230>.
- Mul, M., Obuobie, E., Appoh, R., Kankam-Yeboah, K., Bekoe-Obeng, E., Amisigo, B., Logah, F.Y., Ghansah, B., McCartney, M., 2015. Water resources assessment of the Volta River BasinRep. 9290908297. International Water Management Institute (IWMI) 82 pp http://www.iwmi.cgiar.org/Publications/Working_Papers/working/wor166.pdf.
- Nicholson, S., 2000. Land surface processes and Sahel climate. *Rev. Geophys.* 38 (1), 117–139. <https://doi.org/10.1029/1999RG900014>.
- Nijzink, R.C., et al., 2018. Constraining Conceptual Hydrological Models With Multiple Information Sources. *Water Resour. Res.* 54 (10), 8332–8362. <https://doi.org/10.1029/2017wr021895>.
- Odusanya, A.E., Mehdi, B., Schürz, C., Oke, A.O., Awokola, O.S., Awomeso, J.A., Adejuwon, J.O., Schulz, K., 2019. Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data-sparse catchment in southwestern Nigeria. *Hydrol. Earth Syst. Sc.* 23 (2), 1113–1144. <https://doi.org/10.5194/hess-23-1113-2019>.
- Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. *Science* 313 (5790), 1068–1072. <https://doi.org/10.1126/science.1128845>.
- Or, D., Lehmann, P., 2019. Surface evaporative capacitance: How soil type and rainfall characteristics affect global-scale surface evaporation. *Water Resour. Res.* 55 (1), 519–539. <https://doi.org/10.1029/2018WR024050>.
- Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., 2019. Evaluation of global terrestrial evapotranspiration by state-of-the-art approaches in remote sensing, machine learning, and land surface models. *Hydrol. Earth Syst. Sci. Discuss.* 1–51. <https://doi.org/10.5194/hess-2019-409>.
- Peters-Lidard, C.D., Clark, M., Samaniego, L., Verhoest, N.E.C., van Emmerik, T., Uijlenhoet, R., Achieng, K., Franz, T.E., Woods, R., 2017. Scaling, Similarity, and the Fourth Paradigm for Hydrology. *Hydrol. Earth Syst. Sc.* 21 (7), 3701–3713. <https://doi.org/10.5194/hess-2016-695>.
- Peters-Lidard, C.D., Hossain, F., Leung, L.R., McDowell, N., Rodell, M., Tapiador, F.J., Turk, F.J., Wood, A., 2019. 100 years of progress in hydrology. *Meteorol. Monographs.* <https://doi.org/10.1175/AMSMONOGRAPH-D-18-0019.1>.
- Pokhrel, P., Gupta, H.V., Wagener, T., 2008. A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resour. Res.* 44 (12). <https://doi.org/10.1029/2007wr006615>.
- Pomeon, T., Diekkruger, B., Kumar, R., 2018. Computationally Efficient Multivariate Calibration and Validation of a Grid-Based Hydrologic Model in Sparsely Gauged West African River Basins. *Water* 10 (10), 1418. <https://doi.org/10.3390/w10101418>.
- Rajib, A., Evanson, G.R., Golden, H.E., Lane, C.R., 2018. Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. *J. Hydrol.* 567, 668–683. <https://doi.org/10.1016/j.jhydrol.2018.10.024>.
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schafer, D., Schron, M., Samaniego, L., 2016. Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins. *J. Hydrometeorol.* 17 (1), 287–307. <https://doi.org/10.1175/Jhm-D-15-0054.1>.
- Ramoelo, A., Majoz, N., Mathieu, R., Jovanovic, N., Nickless, A., Dzikiti, S., 2014. Validation of global evapotranspiration product (MOD16) using flux tower data in the African savanna, South Africa. *Remote Sensing* 6 (8), 7406–7423. <https://doi.org/10.3390/rs6087406>.
- Reichle, R.H., Liu, Q., Koster, R.D., Draper, C.S., Mahanama, S.P., Partyka, G.S., 2017. Land surface precipitation in MERRA-2. *J. Climate* 30 (5), 1643–1664. <https://doi.org/10.1175/JCLI-D-16-0570.1>.
- Rientjes, T., Muthuwatta, L.P., Bos, M., Booij, M.J., Bhatti, H., 2013. Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration. *J. Hydrol.* 505, 276–290. <https://doi.org/10.1016/j.jhydrol.2013.10.006>.
- Sakumura, C., Bettadpur, S., Bruinsma, S., 2014. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.* 41 (5), 1389–1397. <https://doi.org/10.1002/2013GL058632>.
- Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* 46 (5). <https://doi.org/10.1029/2008WR007327>.
- Samaniego, L., Kumar, R., Jackisch, C., 2011. Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data. *Hydrolog. Res.* 42 (5), 338–355. <https://doi.org/10.2166/nh.2011.156>.
- Samaniego, L., et al., 2017. Toward seamless hydrologic predictions across spatial scales. *Hydrol. Earth Syst. Sc.* 21 (9), 4323–4346. <https://doi.org/10.5194/hess-21-4323-2017>.
- Savenije, H.H., 2001. Equifinality, a blessing in disguise? *Hydrol. Process.* 15 (14), 2835–2838. <https://doi.org/10.1002/hyp.494>.
- Schaefli, B., Harman, C., Sivapalan, M., Schymanski, S., 2010. Hydrologic predictions in a changing environment: behavioral modeling. *Hydrol. Earth Syst. Sci. Discuss.* 7, 7779–7808. <https://doi.org/10.5194/hess-15-635-2011>.
- Senay, G.B., Budde, M., Verdin, J.P., Melesse, A.M., 2007. A coupled remote sensing and simplified surface energy balance approach to estimate actual evapotranspiration from irrigated fields. *Sensors* 7 (6), 979–1000. <https://doi.org/10.3390/s7060979>.
- 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 (3), 577–591. <https://doi.org/10.1111/jawr.12057>.
- Senkondo, W., Munishi, S.E., Tumbo, M., Nobert, J., Lyon, S.W., 2019. Comparing remotely-sensed surface energy balance evapotranspiration estimates in heterogeneous and data-limited regions: A case study of Tanzania's Kilombero Valley. *Remote Sensing* 11 (11), 1289. <https://doi.org/10.3390/rs1111289>.
- Shuttleworth, W., 1993. In: DR, M. (Ed.), edited by. McGraw-Hill New York, New York pp. 4.1–4.53.
- Sörensson, A.A., Ruscica, R.C., 2018. Intercomparison and uncertainty assessment of nine evapotranspiration estimates over South America. *Water Resour. Res.* 54 (4), 2891–2908. <https://doi.org/10.1002/2017WR021682>.
- Stisen, S., Koch, J., Sonnenborg, T.O., Refsgaard, J.C., Bircher, S., Ringgaard, R., Jensen, K.H., 2018. Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model. *Hydrol. Process.* 32 (17), 2654–2668. <https://doi.org/10.1002/hyp.13177>.
- Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sc.* 6 (1), 85–100. <https://doi.org/10.5194/hess-6-85-2002>.
- Swenson, S. C. (2012). GRACE monthly land water mass grids NETCDF release 5.0. Ver. 5.0. PO.DAAC, CA, USA, <https://doi.org/10.5067/TELND-NC005>, [Dataset accessed 2018.11.01].
- Talsma, C.J., Good, S.P., Jimenez, C., Martens, B., Fisher, J.B., Miralles, D.G., McCabe, M.F., Purdy, A.J., 2018. Partitioning of evapotranspiration in remote sensing-based models. *Agricultr. For. Meteorol.* 260, 131–143. <https://doi.org/10.1016/j.agrformet.2018.05.010>.
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.* 31 (9). <https://doi.org/10.1029/2004gl019920>.
- Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F.W., Dutta, D., Kim, S., 2017. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environ. Modell. Softw.* 90, 201–216. <https://doi.org/10.1016/j.envsoft.2017.01.006>.
- Thober, S., M.Cuntz, M.Kelbling, R.Kumar, J.Mai, and L. J. G. M. D. Samaniego (2019), The multiscale Routing Model mRM v1.0: simple river routing at resolutions from 1 to 50 km, 2019, 1-26, <https://doi.org/10.5194/gmd-12-2501-2019>.
- Tobin, K.J., Bennett, M.E., 2017. Constraining SWAT calibration with remotely sensed evapotranspiration data. *JAWRA J. Am. Water Resour. Assoc.* 53 (3), 593–604. <https://doi.org/10.1111/1752-1688.12516>.
- Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* 43 (1). <https://doi.org/10.1029/2005wr004723>.
- Trabucco, A., and R.Zomer (2018), Global Aridity Index and Potential Evapotranspiration (ETO) Climate Database v2. CGIAR Consortium for Spatial Information (CGIAR-CSL). Published online, available from the CGIAR-CSL GeoPortal at <https://cgciarcsli.community>, https://figshare.com/articles/Global_Aridity_Index_and_Potential_EvapotranspirationETO_Climate_Database_v2/7504448 [accessed 11.07.2019].
- 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. Sc.* 18 (1), 193–212. <https://doi.org/10.5194/hess-18-193-2014>.
- Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.* 26 (20), 4485–4498. <https://doi.org/10.1080/01431160500168686>.

- Upton, G., Cook, I., 2014. A dictionary of statistics, edited Oxford university press.
- Van der Ent, R.J., Savenije, H.H., Schaeefli, B., Steele-Dunne, S.C., 2010. Origin and fate of atmospheric moisture over continents. *Water Resour. Res.* 46 (9). <https://doi.org/10.1029/2010WR009127>.
- Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sens. Environ.* 139, 35–49. <https://doi.org/10.1016/j.rse.2013.07.013>.
- Vervoort, R.W., Miechels, S.F., van Ogtrop, F.F., Guillaume, J.H., 2014. Remotely sensed evapotranspiration to calibrate a lumped conceptual model: Pitfalls and opportunities. *J. Hydrol.* 519, 3223–3236. <https://doi.org/10.1016/j.jhydrol.2014.10.034>.
- Vinukollu, R.K., Meynadier, R., Sheffield, J., Wood, E.F., 2011a. Multi-model, multi-sensor estimates of global evapotranspiration: Climatology, uncertainties and trends. *Hydrolog. Process.* 25 (26), 3993–4010. <https://doi.org/10.1002/hyp.8393>.
- Vinukollu, R.K., Wood, E.F., Ferguson, C.R., Fisher, J.B., 2011b. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sens. Environ.* 115 (3), 801–823. <https://doi.org/10.1016/j.rse.2010.11.006>.
- Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., Ertl, M., 2012. Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture, paper presented at ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals).
- Wambura, F.J., Dietrich, O., Lischeid, G., 2018. Improving a distributed hydrological model using evapotranspiration-related boundary conditions as additional constraints in a data-scarce river basin. *Hydrolog. Process.* 32 (6), 759–775. <https://doi.org/10.1002/hyp.11453>.
- Wang, G., Pan, J., Shen, C., Li, S., Lu, J., Lou, D., Hagan, D.F., 2018. Evaluation of evapotranspiration estimates in the Yellow River Basin against the water balance method. *Water* 10 (12), 1884. <https://doi.org/10.3390/w10121884>.
- Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.* 50 (2). <https://doi.org/10.1029/2011RG000373>.
- Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P., 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.* 50 (9), 7505–7514. <https://doi.org/10.1002/2014WR015638>.
- Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., van Griensven, A., 2020. Can we trust remote sensing evapotranspiration products over Africa? *Hydrolog. Earth Syst. Sci.* 24 (3), 1565–1586. <https://doi.org/10.5194/hess-24-1565-2020>.
- West, H., Quinn, N., Horswell, M., 2019. Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities. *Remote Sens. Environ.* 232, 111291. <https://doi.org/10.1016/j.rse.2019.111291>.
- Williams, T.O., Mul, M.L., Biney, C.A., Smakhtin, V., 2016. The Volta River Basin: Water for food, economic growth and environment, edited Routledge 282 pp.
- Winsemius, H., Savenije, H., Bastiaanssen, W., 2008. Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins? *Hydrolog. Earth Syst. Sc.* 12 (6), 1403–1413. <https://doi.org/10.5194/hess-12-1403-2008>.
- Yang, Z., Zhang, Q., Yang, Y., Hao, X., Zhang, H., 2016. Evaluation of evapotranspiration models over semi-arid and semi-humid areas of China. *Hydrolog. Process.* 30 (23), 4292–4313. <https://doi.org/10.1002/hyp.10824>.
- Yilmaz, M.T., Anderson, M.C., Zaitchik, B., Hain, C.R., Crow, W.T., Ozdogan, M., Chun, J.A., Evans, J., 2014. Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin. *Water Resour. Res.* 50 (1), 386–408. <https://doi.org/10.1002/2013WR014194>.
- Zhang, B., Xia, Y., Long, B., Hobbins, M., Zhao, X., Hain, C., Li, Y., Anderson, M.C., 2020. Evaluation and comparison of multiple evapotranspiration data models over the contiguous United States: Implications for the next phase of NLDAS (NLDAS-Testbed) development. *Agricult. For. Meteorol.* 280, 107810. <https://doi.org/10.1016/j.agrmet.2019.107810>.
- Zhang, K., Kimball, J.S., Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdisciplin. Rev. Water* 3 (6), 834–853. <https://doi.org/10.1002/wat2.1168>.
- Zhang, Y., Chiew, F.H., Liu, C., Tang, Q., Xia, J., Tian, J., Kong, D., Li, C., 2019. Can remotely sensed actual evapotranspiration facilitate hydrological prediction in ungauged regions without runoff calibration? *Water Resour. Res.* e24394. <https://doi.org/10.1029/2019WR026236>.
- Zhu, Z.C., Bi, J., Pan, Y.Z., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S.L., Nemani, R.R., Myneni, R.B., 2013. Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011. *Remote Sensing* 5 (2), 927–948. <https://doi.org/10.3390/rs5020927>.
- Zink, M., Mai, J., Cuntz, M., Samaniego, L., 2018. Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature. *Water Resour. Res.* 54 (4), 2976–2998. <https://doi.org/10.1002/2017wr021346>.