Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets

Dembélé, Moctar; Hrachowitz, Markus; Savenije, Hubert H.G.; Mariéthoz, Grégoire; Schaefli, Bettina

DOI
10.1029/2019WR026085

Publication date
2020

Document Version
Final published version

Published in
Water Resources Research

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.
Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets

Moctar Dembélé, Markus Hrachowitz, Hubert H. G. Savenije, Grégoire Mariéthoz, and Bettina Schaeffli

1Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland, 2Faculty of Civil Engineering and Geosciences, Water Resources Section, Delft University of Technology, Delft, Netherlands, 3Now at Institute of Geography (GIUB), University of Bern, Bern, Switzerland

Abstract Hydrological model calibration combining Earth observations and in situ measurements is a promising solution to overcome the limitations of the traditional streamflow-only calibration. However, combining multiple data sources in model calibration requires a meaningful integration of the data sets, which should harness their most reliable contents to avoid accumulation of their uncertainties and mislead the parameter estimation procedure. This study analyzes the improvement of model parameter selection by using only the spatial patterns of satellite remote sensing data, thereby ignoring their absolute values. Although satellite products are characterized by uncertainties, their most reliable key feature is the representation of spatial patterns, which is a unique and relevant source of information for distributed hydrological models. We propose a novel multivariate calibration framework exploiting spatial patterns and simultaneously incorporating streamflow and three satellite products (i.e., Global Land Evaporation Amsterdam Model [GLEAM] evaporation, European Space Agency Climate Change Initiative [ESA CCI] soil moisture, and Gravity Recovery and Climate Experiment [GRACE] terrestrial water storage). The Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature data set is used for model evaluation. A bias-insensitive and multicomponent spatial pattern matching metric is developed to formulate a multiobjective function. The proposed multivariate calibration framework is tested with the mesoscale Hydrologic Model (mHM) and applied to the poorly gauged Volta River basin located in a predominantly semiarid climate in West Africa. Results of the multivariate calibration show that the decrease in performance for streamflow (−7%) and terrestrial water storage (−6%) is counterbalanced with an increase in performance for soil moisture (+105%) and evaporation (+26%). These results demonstrate that there are benefits in using satellite data sets, when suitably integrated in a robust model parametrization scheme.

1. Introduction

One of the key challenges in hydrological modeling (Beven, 2019a; Singh, 2018) is the reliable representation of the spatiotemporal variability of natural processes, to which the footprint of human activity is often superimposed. In most places, available in situ observations are not sufficient to capture the spatiotemporal heterogeneity of dominant hydrological processes (AghaKouchak et al., 2015; Hrachowitz & Clark, 2017). With the upswing in development of distributed hydrological models (DHMs) that offer spatially explicit predictions as an essential tool for decision making (Fatichi et al., 2016; Kampf & Burges, 2007; Paniconi & Putti, 2015; Semenova & Beven, 2015), there is a growing interest in the plausibility of their spatial patterns (Ko et al., 2019; Koch et al., 2018; Stisen et al., 2018; Wealands et al., 2005; Zink et al., 2018).

Most commonly, hydrological models are calibrated using streamflow data alone (Becker et al., 2019; Yassin et al., 2017). The streamflow signal represents an integrated response of the hydrological system to a set of natural drivers (e.g., climate and landscape) and anthropogenic influences (e.g., deforestation and reservoirs) occurring upstream of the measurement’s location (Koch et al., 2015; Rientjes et al., 2013). Although streamflow is key to understanding the temporal dynamics of a system, it does not disclose much information on the system-internal spatial heterogeneity of the hydrological processes (McDonnell et al., 2007; Rajib et al., 2018). It therefore has little discriminatory power to constrain the feasible parameter space.
of a distributed model, i.e., the boundary flux or closure problem (Beven, 2006b). Consequently, a spatially DHM calibrated only on streamflow is very unlikely able to reproduce a reliable spatiotemporal representation of other hydrological fluxes and states (Birhanu et al., 2019; Clark et al., 2016; Grayson & Bloschl, 2001; Hrachowitz et al., 2014; Livneh & Lettenmaier, 2012; Minville et al., 2014), even if a multiscale parameter regionalization (MPR) scheme is used (Rakovec et al., 2016). Mismatches between temporal and spatial patterns should therefore be expected when comparing hydrological model outputs to other distributed observational data sets (Vereecken et al., 2008; Xu et al., 2014).

For a few decades, satellite remote sensing (SRS) has opened up new avenues for the development of spatial hydrology (Cui et al., 2018; Engman & Gurney, 1991; Lettenmaier et al., 2015; McCabe et al., 2017; Mendoza et al., 2002; Pasetto et al., 2018; Schmugge et al., 2002). The increasing and unprecedented availability of SRS data at increasingly finer spatial and temporal resolutions has triggered the development of large-domain water management applications including flood and drought monitoring (Hapuarachchi et al., 2011; Klemas, 2014; Revilla-Romero et al., 2015; Senay et al., 2015; Sheffield et al., 2012; Su et al., 2017; Teng et al., 2017; Wu et al., 2014). The use of SRS data in water resources monitoring is promising, and it has led to an increasing number of studies on a variety of topics in hydrology, including precipitation, evaporation, and soil moisture estimation (Cañazave et al., 2016; Chen & Wang, 2018; Cui et al., 2019; National Academies of Sciences, Engineering, and Medicine, 2019; Schultz & Engman, 2012). SRS data complement in situ hydrometeorological data (Balsamo et al., 2018), which are typically scarce and whose unavailability hinders the understanding of environmental systems (Tang et al., 2009). This aspect is particularly relevant for developing countries where research for development initiatives have been increasing in the recent years (Montanari et al., 2015).

Besides direct use of SRS data for water resources monitoring and management (Cui et al., 2019; Sheffield et al., 2018), an increasing body of literature addresses the question of how these data sets can be used to improve hydrological modeling (Baroni et al., 2019; Clark et al., 2015; Guntner, 2008; Liu et al., 2012; Nijzink et al., 2018; Paniconi & Putti, 2015; Parajka et al., 2009). The scientific community has, in fact, long been advocating the use of spatial data for DHM evaluation (Beven & Feyen, 2002; Grayson & Bloschl, 2001; Koch et al., 2015; Refsgaard, 2001; Wealands et al., 2005). SRS data sets have the potential to improve models either via data assimilation (Leroux et al., 2016; Tangdamrongsub et al., 2017; Tian et al., 2017) or via calibration (Bai et al., 2018; Li et al., 2018; Rientjes et al., 2013). In this context, data assimilation is used to update the states of a given model, e.g., to compensate for model structural deficiencies (Spaaks & Boutsen, 2013). For parameter estimation (i.e., model calibration) with SRS data, the existing approaches consist in using SRS data alone or in combination with in situ data, usually streamflow data (Immerzeel & Droogers, 2008; Li et al., 2018; Rajib, Evenson, et al., 2018; Wambura et al., 2018). Calibration of hydrological models without concomitant streamflow data remains challenging, and attempts to do so have only shown limited success (Nijzink et al., 2018; Silvestro et al., 2015; Sutanudjaja et al., 2014; Wanders et al., 2014).

The simultaneous calibration of hydrological models with streamflow and different combinations of complementary data from SRS is increasingly discussed in recent literature (Stisen et al., 2018). Multivariate (i.e., multiple variables) parameter estimation (Efstratiadis & Koutsoyiannis, 2010) can substantially reduce the feasible model and parameter space and lead to more realistic internal model dynamics and related hydrological signatures (Clark et al., 2017; Shafii & Tolson, 2015), which can ultimately enhance the overall representation of catchment functioning (Bergström et al., 2002; Rakovec et al., 2016). Furthermore, and intimately linked to the above, multivariate calibration strategies can considerably reduce equifinality (i.e., nonidentifiable model parameters in inverse modeling approaches; Beven, 2006a; Savenije, 2001) and reduce prediction uncertainty (Fenicia et al., 2008; Fovet et al., 2015; Gupta et al., 1998; Gupta et al., 2008; Hrachowitz et al., 2014; Schoups et al., 2005). However, important open questions remain with respect to the combination of SRS data with streamflow data for model parameter estimation. While some studies observed a significant improvement in the representation of model outputs after SRS data incorporation (Chen et al., 2017; Leroux et al., 2016; Werth et al., 2009; Yassin et al., 2017), others found minor changes or even major deteriorations (Stisen et al., 2018; Tangdamrongsub et al., 2017; Tobin & Bennett, 2017). Such apparently contradictory conclusions are case study specific and need to be understood as resulting from model structures, model parametrizations, and trade-offs between improving water balance estimates and related streamflow dynamics and better representing other hydrological fluxes and states (Euser et al., 2013; Koppa et al., 2019;
Yassin et al., 2017). More generally, the key challenge results from the integration of several data sources (SRS or in situ) in parameter estimation, which can be attributed to conflicting information from different types of SRS data. Nonetheless, multivariate parameter estimation with SRS data remains promising, especially when streamflow data availability is limited or the data quality is questionable.

Although SRS data are more accessible with higher spatiotemporal resolution compared to in situ observations, they are generally not direct measurements of hydrological processes, which adds a level of uncertainty to any SRS-based parameter estimation study (Ehlers et al., 2018; Knoche et al., 2014; Ma et al., 2018). However, they provide spatial information on hydrological processes, which makes them a unique and relevant information source for spatially distributed representations of the system in models (Stisen et al., 2018; Wambura et al., 2018). For instance, many studies report different model performances when using different satellite-based products as input (e.g., precipitation; Pomeon et al., 2018; Thiemig et al., 2013) or as calibration variables (e.g., evaporation, soil moisture, and terrestrial water storage; Bai, Liu, & Liu, 2018; Nijzink et al., 2018). Nevertheless, for a given region, different products can give considerably different absolute values of a specific variable while they may exhibit plausible and similar spatial patterns (Beck et al., 2017; Dembele & Zwart, 2016). Additionally, retaining only the spatial pattern information of SRS data can substantially mitigate the uncertainty resulting from the fact that they are not direct observations, as long as their relative values are used rather than their absolute values (Mendiguren et al., 2017; Wambura et al., 2018).

In the context of using SRS data for DHM calibration, the simultaneous use of more than one SRS product to constrain several hydrological state or flux variables is uncommon (Clark et al., 2017; Lopez et al., 2017; Nijzink et al., 2018), as is the incorporation of spatial pattern in the calibration scheme using bias-insensitive metrics (Demirel et al., 2018; Zink et al., 2018). Using different variables from SRS products simultaneously in parameter estimation is in general not straightforward (Rajib et al., 2018; Silvestro et al., 2015; Tian et al., 2017) because they all have limitations (e.g., spatiotemporal resolutions and accuracy), which can lead to significant trade-offs in multivariate calibration (Koppa et al., 2019).

In light of the above, we propose to test a novel multivariate calibration strategy in which a DHM will be trained to simultaneously reproduce spatial patterns, i.e., relative spatial differences, of three variables from different SRS products describing different components of the hydrological system (i.e., evaporation, soil moisture, and terrestrial water storage), as well as in situ observations of streamflow. The proposed calibration framework combines simultaneously four noncommensurable variables and a new bias-insensitive metric for spatial pattern representation, which as a whole is different from previous studies (e.g., Demirel et al., 2018; Koppa et al., 2019; Nijzink et al., 2018; Rakovec et al., 2016; Zink et al., 2018) and therefore makes the novelty of this study. The following research hypotheses are tested:

1. Building upon previous work (e.g., Demirel et al., 2018; Rakovec et al., 2016; Zink et al., 2018), we assume that simultaneously calibrating a DHM on four noncommensurable variables and spatial patterns of satellite data considerably improves the predictive skill of the model, even for a DHM integrating a MPR scheme.
2. Our new bias-insensitive metric based on pixel-by-pixel locational matching can be used to improve the calibration of a DHM on observed spatial patterns of hydrological processes even in the presence of strong climatic gradients.

The overall goal of this study is to improve the spatial representation of dominant hydrological processes of a DHM without significantly deteriorating the streamflow signal and reproducing plausible dynamics of the hydrological system using spatial pattern information from SRS data sets. Such improvement will be an asset for spatial hydrology and large-domain water management applications (e.g., water accounting, drought monitoring, and flood prediction) and might subsequently lead to advances in prediction in ungauged basins (Blöschl et al., 2013; Hrachowitz et al., 2013; Sivapalan, 2003) with the use of readily accessible SRS data (Butler, 2014; Wulder & Coops, 2014). This work embraces the fourth paradigm for hydrology (i.e., data-intensive science, Peters-Lidard et al., 2017) and contributes to solving some of the issues (e.g., spatial variability and modeling methods) recently identified as the 23 unsolved problems in hydrology in the 21st century (Blöschl et al., 2019). The proposed multivariate calibration framework is tested with the mesoscale Hydrologic Model (mHM), with a case study in the poorly gauged Volta River basin (VRB) in West Africa.
2. Study Area

The transboundary VRB is the study area. It covers approximately 415,600 km² across six countries of West Africa. Figure 1 shows the physical and hydroclimatic characteristics of the VRB. The climate is characterized by a south-north gradient of increasing aridity and varies from subhumid in the south to semiarid in the north (Dembélé et al., 2019). Climate is driven by the Intertropical Convergence Zone, and four ecoclimatic zones (i.e., Sahelian, Sudano-Sahelian, Sudanian, and Guinean) can be identified (Figure 1a) based on the average annual precipitation and agricultural features (Food and Agriculture Organization/Global Information and Early Warning System, 1998; Mul et al., 2015). The characteristics of the four ecoclimatic zones are given in Table 1. Actual evaporation exceeds 80% of annual rainfall in the basin (Andreini et al., 2000; De Condappa & Lemoalle, 2009).

The topography is predominately flat as 95% of the relief is below 400 m above sea level (Figure 1b). The drainage system is composed of four sub-basins known as Black Volta (152,800 km²), White Volta (113,400 km²), Oti (74,500 km²), and Lower Volta (74,900 km²). The Volta River flows over 1,850 km and transits in the Lake Volta formed by the Akosombo dam before draining into the Atlantic Ocean at the Gulf of Guinea (Williams et al., 2016). Land cover (Figure 1c) is dominated by savannah formed by grassland interspersed with shrubs and trees covering about 88% of the basin area. Other land cover types include forest (9%), water bodies (2%), and bare land and settlements (1%).

![Figure 1. Physical and hydroclimatic characteristics of the Volta River basin.](image)

### Table 1

<table>
<thead>
<tr>
<th>Ecoclimatic zones</th>
<th>Climate class</th>
<th>$AI$ (−)</th>
<th>$P$ (mm/year)</th>
<th>$T_{\text{avg}}$ (°C)</th>
<th>$T_{\text{min}}$ (°C)</th>
<th>$T_{\text{max}}$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sudano-Sahelian</td>
<td>Semiard</td>
<td>0.29 [0.16–0.43]</td>
<td>790 [570–980]</td>
<td>29 [28–29]</td>
<td>20 [20–21]</td>
<td>36 [35–37]</td>
</tr>
</tbody>
</table>

Note. The annual mean value (with min-max range in brackets) is given for each variable. The aridity index (United Nations Environment Programme, 1997) is obtained from the global aridity index database (Trabucco & Zomer, 2018), and the WFDEI data (Weedon et al., 2014) are used for the long-term (1979–2016) estimation of annual precipitation and air temperature. $AI$ = aridity index; $P$ = precipitation, $T_{\text{avg}}$ = average air temperature; $T_{\text{min}}$ = minimum air temperature; $T_{\text{max}}$ = maximum air temperature.
3. Data Sets

The data sets used to set up and run the distributed model for the 2000–2012 period include the basin morphological data (elevation, slope, land cover, etc.) and meteorological data (i.e., rainfall and air temperature). In situ streamflow data and complementary data from SRS are used to calibrate and to evaluate the model performance. A description of the data sets with their characteristics and their sources is given in Table 2.

Concerning the SRS products, the terrestrial water storage ($S_t$) anomaly data derived from changes in surface mass, which is related to the Earth's gravity field, is obtained from the Gravity Recovery and Climate Experiment (GRACE; Landerer & Swenson, 2012; Tapley et al., 2004). Over land, $S_t$ is the sum of snow, ice, surface water, soil moisture, and groundwater. The data release RL05 (Swenson, 2012) is used in this study. It is a simple arithmetic mean of different solutions from three processing centers: Jet Propulsion Laboratory, Center for Space Research at University of Texas, and Geoforschungs Zentrum Potsdam.

### Table 2

**Overview of the Modeling Data Sets**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Product</th>
<th>Spatial resolution</th>
<th>Temporal resolution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model setup</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meteorological data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td>CHIRPS v2.0</td>
<td>0.05°</td>
<td>Daily</td>
<td>Funk et al. (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://chg.geog.ucsb.edu/data/chirps/">http://chg.geog.ucsb.edu/data/chirps/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weedon et al. (2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.eu-watch.org/data_availability">http://www.eu-watch.org/data_availability</a></td>
</tr>
<tr>
<td>Temperature (average, minimum, and maximum)</td>
<td>WFDEI</td>
<td>0.5°</td>
<td>Hourly</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphological data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrain characteristics (elevation, slope, aspect, flow direction, and flow accumulation)</td>
<td>GMTED 2010</td>
<td>225 m (0.0021°)</td>
<td>Static</td>
<td>Danielson and Gesch (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://topotools.cr.usgs.gov/">https://topotools.cr.usgs.gov/</a></td>
</tr>
<tr>
<td>Soil properties (horizon depth, bulk density, and sand and clay content)</td>
<td>SoilGrids</td>
<td>250 m (0.0023°)</td>
<td>Static</td>
<td>Hengl et al. (2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://www.isric.org/explore/soilgrids">https://www.isric.org/explore/soilgrids</a></td>
</tr>
<tr>
<td>Geology</td>
<td>GLiM v1.0</td>
<td>0.5°</td>
<td>Static</td>
<td>Hartmann and Moosdorf (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://doi.pangaea.de/10.1594/PANGAEA.78837">https://doi.pangaea.de/10.1594/PANGAEA.78837</a></td>
</tr>
<tr>
<td>Land use land cover</td>
<td>Globcover 2009</td>
<td>300 m (0.0028°)</td>
<td>Static</td>
<td>Bontemps et al. (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://due.esrin.esa.int/page.globcover.php">http://due.esrin.esa.int/page.globcover.php</a></td>
</tr>
<tr>
<td>Phenology (leaf area index)</td>
<td>GIMMS</td>
<td>8 km (0.0833°)</td>
<td>Bimonthly</td>
<td>Tucker et al. (2005), Zhu et al. (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://cliveg.bu.edu/modismisr/lai3g-fpar3g.html">http://cliveg.bu.edu/modismisr/lai3g-fpar3g.html</a></td>
</tr>
<tr>
<td>Model calibration/evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In situ data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streamflow</td>
<td>-</td>
<td>Point</td>
<td>Daily</td>
<td>Multiple organizations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(see the Acknowledgements section)</td>
</tr>
<tr>
<td>Complementary satellite products</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrestrial water storage anomaly</td>
<td>GRACE TellUS v5.0</td>
<td>1°</td>
<td>Monthly</td>
<td>Tapley et al. (2004), Landerer and Swenson (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://grace.jpl.nasa.gov/">https://grace.jpl.nasa.gov/</a></td>
</tr>
<tr>
<td>Surface soil moisture</td>
<td>ESA CCI SM v4.2</td>
<td>0.25°</td>
<td>Daily</td>
<td>Dorigo et al. (2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://www.esa-soilmoisture-cci.org/">https://www.esa-soilmoisture-cci.org/</a></td>
</tr>
<tr>
<td>Actual evaporation</td>
<td>GLEAM v3.2a</td>
<td>0.25°</td>
<td>Daily</td>
<td>Martens et al. (2017), Miralles et al. (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://www.gleam.eu/">https://www.gleam.eu/</a></td>
</tr>
<tr>
<td>Land surface temperature (only for model evaluation)</td>
<td>MYD11A2 v6</td>
<td>1 km (0.0083°)</td>
<td>8-day</td>
<td>Wan et al. (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="https://lpdaac.usgs.gov/products/myd11a2v006/">https://lpdaac.usgs.gov/products/myd11a2v006/</a></td>
</tr>
</tbody>
</table>

**Note.** 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; GLEAM = Global Land Evaporation Amsterdam Model; 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.
Sakumura et al. (2014) found this ensemble mean product more effective in reducing noise in the Earth’s gravity signal compared to the individual products. As the original baseline for GRACE-derived $S_t$ anomaly data is the period 2004–2009, the $S_t$ data is converted to a new baseline corresponding to the modeling period (2003–2012) used in this study, by averaging each grid point over the new baseline and subtracting that value from all time steps (National Aeronautics and Space Administration, 2019).

The surface soil moisture ($S_s$) data for a soil layer depth of 2–5 cm is obtained from European Space Agency Climate Change Initiative (ESA CCI; Dorigo et al., 2017). The combined product used in this study is a blended product of both active and passive microwave products derived from scatterometer (European Remote-Sensing Satellite (ERS), Active Microwave Instrument (AMI) and Advanced Scatterometer (ASCAT)) and radiometer (Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSM/I), Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E), WindSat, Advanced Microwave Scanning Radiometer 2 (AMSR2), and Soil Moisture and Ocean Salinity (SMOS)) retrievals (Liu et al., 2012; Wagner et al., 2012). The merging algorithm of the combined product version 4.2 is described by Gruber et al. (2017).

Actual evaporation ($E_a$) data are obtained from the Global Land Evaporation Amsterdam Model (GLEAM) land surface model that uses satellite data as input (Martens et al., 2017; Miralles et al., 2011). It separately estimates the components of terrestrial evaporation (i.e., transpiration, bare soil evaporation, open-water evaporation, interception loss, and sublimation) based on the fraction of land cover types (i.e., bare soil, low vegetation, tall vegetation, and open water) before aggregating them for each grid cell. In GLEAM, potential evaporation ($E_p$) is calculated based on the Priestley and Taylor (1972) equation and thereafter converted into transpiration or bare soil evaporation using a stress factor, which is a parameter that accounts for environmental conditions limiting evaporation. The stress factor is estimated from microwave vegetation optical depth (i.e., water content in vegetation) and root-zone soil moisture that is calculated with a multilayer water balance algorithm. The fraction of open-water evaporation is assumed to equal $E_p$. The Gash (1979) analytical model further refined by Valente et al. (1997) is used to calculate rainfall interception by forests.

Land surface temperature ($T_s$) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument of the National Aeronautics and Space Administration satellites are used as independent data for model evaluation, and they are not used during model calibration. The daytime product from the Aqua platform is used because that satellite passes over our study region around 13:45 in local time, corresponding to the highest daily temperature period with a clear-sky coverage (Wan et al., 2015).

For a full description of the data sets, the reader is referred to the corresponding data references in Table 2.

4. Experimental Design

4.1. Hydrological Model

mHM (version 5.8) is a spatially explicit (i.e., fully distributed) conceptual model based on numerical approximations of dominant hydrological processes per grid cell in the modeling domain (Kumar et al., 2013; Samaniego et al., 2010). The following processes can be represented: canopy interception, snow accumulation and melting, soil moisture dynamics, infiltration and surface runoff, evaporation, subsurface storage and discharge generation, deep percolation and baseflow, and discharge attenuation and flood routing. The total grid-generated runoff is routed to the neighboring downstream cell following the river network using a multiscale routing model (Thober et al., 2019) based on the Muskingum-Cunge method (Cunge, 1969). mHM uses an MPR technique (Samaniego et al., 2017) to account for the subgrid variability of the basin physical characteristics (e.g., topography, soil texture, geology, and land cover properties). Pedotransfer functions and global parameters are used to link the model parameters (e.g., hydraulic conductivity and soil porosity) to the basin physical characteristics. There are at most 53 global or super parameters (Pokhrel et al., 2008), which are time- and space-invariant parameters tuned during the calibration procedure. The model uses three different scales of spatial discretization (i.e., grid cell resolution) of the modeling domain corresponding to operation levels. The first scale is used for morphologic data (L0), the second is dedicated to fluxes and states calculation (L1), and the third scale is considered for the meteorological data (L2). L1 should be a multiple of L0 and a submultiple of L2. Three land use and land cover classes are
considered by mHM: forests (e.g., coniferous and deciduous), permeable areas (e.g., grasslands, croplands, and bare soils), and impervious areas (e.g., urban and built-up areas, water bodies, and consolidated soils).

### 4.2. Model Implementation

For the current setup of the mHM model, reference evaporation \( (E_{\text{ref}}) \) is calculated based on the Hargreaves and Samani (1985) method, which only requires temperature data. Following Demirel et al. (2018), a dynamically scaling function \( (F_{\text{DS}}) \) is used to calculate potential evaporation \( (E_p) \) to account for vegetation-climate interactions and improve spatial estimation of \( E_p \) (Bai et al., 2018; Jiao et al., 2017). The approach is based on the concept of \( E_p \) calculation from crop coefficients and \( E_{\text{ref}} \) (Allen et al., 1998; Birhanu et al., 2019). Here the \( F_{\text{DS}} \) (equation (1)) plays the role of a spatially varying crop coefficient, and it is estimated based on leaf area index \( (I_{\text{LA}}) \) data. It is formulated as follows:

\[
F_{\text{DS}} = a + b\left(1 - e^{cI_{\text{LA}}}ight),
\]

where \( a \) is the intercept term representing uniform scaling, \( b \) represents the vegetation-dependent component, and \( c \) describes the degree of nonlinearity in the \( I_{\text{LA}} \) dependency. The coefficients \( a, b, \) and \( c \) are determined during model calibration.

\[
E_p = F_{\text{DS}} \times E_{\text{ref}}.
\]

Total actual evaporation (i.e., all forms of evaporation including transpiration, \( E_a \)) is estimated as a fraction of \( E_p \) from soil layers and depends on the fraction of vegetation roots and soil moisture availability (Feddes et al., 1976). Soil moisture is calculated following a multilayer infiltration capacity approach using a three-layer soil scheme. The depths of the soil layers are 5, 30, and 100 cm. Terrestrial water storage in mHM is calculated per grid cell by summing up all the subsurface water storage (i.e., reservoirs generating soil moisture, interflow, and baseflow) and the surface water storage in sealed areas. There is no snowfall in the VRB.

In this study, 36 global parameters are determined through model calibration (see supporting information Table S2). The full description of the model and the formulation of the hydrological processes can be found in the work of Samaniego et al. (2010). All morphological data are resampled to a resolution of 1/512° (~200 m at the equator) and the meteorological data to 0.0625° (~7 km). The nearest neighbor technique is used to resample categorical data, while bilinear resampling is used for continuous data. The model is run at daily time step with a spatial resolution of 0.25° (~28 km), corresponding to 619 modeling grid cells in the basin, but taking into account the subgrid variability of the morphological data using the MPR technique (Samaniego et al., 2010; Samaniego et al., 2017). Every modeling grid cell (L1) contains 16 meteorological grid cells (L2) and 16,384 morphological grid cells (L0).

### 4.3. Model Calibration and Evaluation Strategies

The modeling period spans from 2000 to 2012 and consists of 3 years (2000–2002) model warm-up period, 6 years (2003–2008) calibration period, and 4 years (2009–2012) evaluation period. Based on data availability and quality in the VRB (Dembélé et al., 2019), 11 streamflow gauging stations are chosen to have a good coverage of the river network (Figure 1), and the calibration is done on them simultaneously to obtain a single-parameter set for the whole VRB. The domain-wide calibration, which was proven to give similar performance as the domain-split calibration (Mizukami et al., 2017), is preferred here because of the limited number of streamflow stations and for seamless spatial pattern representation across sub-basins (see Figure 1).

Two main calibration approaches are adopted to evaluate the benefit of including spatial patterns in multivariate parameter estimation with SRS data. The first approach is the streamflow-only calibration, and the second approach uses multiple SRS data sets in addition to streamflow. In both cases, the formulation of the objective functions follows the Euclidian distance approach in which all elements are equally weighted (Khu & Madsen, 2005); see equation (A1).

#### 4.3.1. Calibration on Streamflow—Benchmark

The first calibration approach is the benchmark calibration case (case Q) where the hydrological model is constrained with in situ streamflow (Q) data only. An objective function \( \Phi_Q \) (equation (3)) combines the Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) of streamflow \( (E_{\text{NS}}) \) and the Nash-Sutcliffe efficiency of
the logarithm of streamflow ($\text{ENSlog}$) presented in equations (A2) and (A3), and it is formulated as Euclidean distance such that it has to be minimized:

$$
\Phi_Q = \frac{1}{g} \sum_g \sqrt{(1-\text{EN})^2 + (1-\text{ENSlog})^2},
$$

where $g$ is the number of streamflow gauging stations present within the modeling domain. $\Phi_Q$ is obtained by equally weighing the streamflow gauging stations, and it ranges from 0, its ideal value, to $+\infty$.

### 4.3.2. Calibration on Multiple Variables With Spatial Patterns

#### 4.3.2.1. Spatial Pattern Efficiency Metric

The degree of reproduction of the spatial patterns of $E_a$ and $S_a$ is quantified with a proposed pattern matching metric, denoted $E_{SP}$. The development of $E_{SP}$ is motivated by the need for simplicity and robustness in pattern matching with respect to existing metrics (cf. Koch et al., 2015; Koch et al., 2018). It is a multicomponent metric formulated in a way functionally equivalent to the Kling-Gupta efficiency (cf. equation (A6); Gupta et al., 2009; Kling et al., 2012) and the SPAtial Efficiency metric (SPAEM, Demirel et al., 2018; Koch et al., 2018). It simultaneously assesses the matching of the spatial distribution of grid cells, the relative variance, and the strength of the monotonic relationship between the observed and estimated variables. Therefore, $E_{SP}$ is a bias-insensitive metric that focuses on the patterns of the variables rather than their magnitudes. Considering a modeled variable ($X_{mod}$) and an observed variable ($X_{obs}$) of $n$ elements, $E_{SP}$ is defined as follows:

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

where $r_s$ is the Spearman rank-order correlation coefficient with $d$ the difference between the ranks of $X_{mod}$ and $X_{obs}$, $\gamma$ is the variability ratio (i.e., ratio of coefficients of variation) that assesses the similarity in the dispersion of the probability distributions of $X_{mod}$ and $X_{obs}$, with $\mu$ and $\sigma$ representing the mean and the standard deviation, respectively, and $\alpha$ the spatial location matching term calculated as the root-mean-square error ($E_{RMS}$) of the standardized values ($z$-scores, $Z_Q$) of $X_{mod}$ and $X_{obs}$ (see equations (A4) and (A5)). The z-scores is a standardization of the scale of a distribution that facilitates its comparison with another distribution. The $z$-scores identify and describe the exact location of each observation in a distribution (Gravetter & Wallnau, 2013). For a given variable with values represented spatially as a 2-D matrix, the $z$-scores represent the number of standard deviations the value in each grid cell is from the population mean (Oyana & Margai, 2015). Consequently, forcing the $z$-scores of $X_{mod}$ and $X_{obs}$ to be equal (i.e., minimizing their $E_{RMS}$) corresponds to matching their grid cell locations (i.e., spatial patterns). Finally, $E_{SP}$ is formulated such that it ranges from $-\infty$ to 1, which is its optimal value. Contrarily to SPAEM, $E_{SP}$ does not require any user-defined parameter (i.e., number of bins in SPAEM), and it uses a nonparametric correlation coefficient (i.e., $r_s$), which limits its sensitivity to outliers as opposed to the Pearson correlation coefficient (Legates & McCabe, 1999; Pool et al., 2018; Spearman, 1904) used in SPAEM. A comparison of $E_{SP}$ to SPAEM is provided in the supporting information (Figure S40 and Table S3).

#### 4.3.2.2. Multivariate Calibration Strategies

In contrast to the first calibration strategy, which only considers $Q$ as target variable (case $Q$), the second calibration strategy involves multiple variables (case MV). The potential improvement of the modeled fluxes and states is estimated by constraining the parameter estimation with a simultaneous combination of three variables from SRS products ($S_a$, $S_{n}$, and $E_a$), in addition to $Q$. Here the spatial patterns of $E_a$ and $S_n$ are used in the multivariate calibration, while the temporal dynamics of $S_a$ are averaged over the whole basin due to
Regions where little or no stream data are available. Consequently, four additional objective functions (Table 3) are used for multivariate calibration cases without $E_a$ (case MV-$E_a$), without $S_u$ (case MV-$S_u$), without $S_t$ (case MV-$S_t$), and without streamflow (case MV-$Q$).

### Table 3: Variants of Multivariate Calibration Cases in the Leave-One-Out Approach

<table>
<thead>
<tr>
<th>Calibration case</th>
<th>Calibration variable</th>
<th>Objective function</th>
<th>Specificity</th>
<th>Equation number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case MV-$E_a$</td>
<td>Q, S_t, S_u</td>
<td>$\Phi_{MV-E_a} = \sqrt{\Phi_Q^2 + \Phi_{S_t}^2 + \Phi_{S_u}^2}$</td>
<td>No direct constraint on evaporation</td>
<td>(12)</td>
</tr>
<tr>
<td>Case MV-$S_u$</td>
<td>Q, S_t, E_a</td>
<td>$\Phi_{MV-S_u} = \sqrt{\Phi_Q^2 + \Phi_{S_t}^2 + \Phi_{E_a}^2}$</td>
<td>No specific constraint on surface soil moisture</td>
<td>(13)</td>
</tr>
<tr>
<td>Case MV-$S_t$</td>
<td>Q, S_u, E_a</td>
<td>$\Phi_{MV-S_t} = \sqrt{\Phi_Q^2 + \Phi_{S_u}^2 + \Phi_{E_a}^2}$</td>
<td>No direct constraint on deep subsurface processes</td>
<td>(14)</td>
</tr>
<tr>
<td>Case MV-$Q$</td>
<td>S_t, S_u, E_a</td>
<td>$\Phi_{MV-Q} = \sqrt{\Phi_S^2 + \Phi_{S_u}^2 + \Phi_{E_a}^2}$</td>
<td>Only satellite-based variables with no direct constraint on streamflow</td>
<td>(15)</td>
</tr>
</tbody>
</table>

The relatively coarse spatial resolution of the GRACE data. The temporal dynamics of $S_t$ per grid cell is also assessed during model evaluation. The multivariate objective function $\Phi_{MV}$ (equation (8)) is defined as follows:

$$\Phi_{MV} = \sqrt{\Phi_Q^2 + \Phi_{S_t}^2 + \Phi_{S_u}^2}$$ (8)

$$\Phi_{E_a} = 1 - \frac{1}{T} \sum_t E_{SP}(E_{a,mod}(t), E_{a,obs}(t))$$ (9)

$$\Phi_{S_u} = 1 - \frac{1}{T} \sum_t E_{SP}(S_{u,mod}(t), S_{u,obs}(t))$$ (10)

$$\Phi_{S_t} = E_{RMS}(Z_{S_t,mod}(t), Z_{S_t,obs}(t))$$ (11)

where $t$ is the number of time steps of the calibration period. The subobjective functions $\Phi_{E_a}$ and $\Phi_{S_u}$ are based on the $E_{SP}$ (equation (4)) of modeled and observed $E_a$ and $S_u$, while $\Phi_{S_t}$ denotes the root-mean-square error ($E_{RMS}$) of the $z$ scores of the modeled and observed basin-averaged $S_t$ anomaly.

Consequently, $\Phi_Q$ ensures a reliable prediction of streamflow signatures (i.e., high and low flows), $\Phi_{E_a}$ and $\Phi_{S_u}$ serve to improve the spatial patterns of the modeled $E_a$ and $S_u$, while $\Phi_{S_t}$ constrains the temporal dynamics of the modeled $S_t$, which should contribute to a better prediction of the water balance at monthly and annual scales. In fact, $\Phi_{E_a}$ and $\Phi_{S_u}$ are calculated such that the spatial pattern efficiencies of $E_a$ and $S_u$ are determined over the grid cells at each monthly time step, before averaging them over the calibration period, while $\Phi_{S_t}$ is calculated for the basin-averaged $S_t$ over the calibration period. Note that in the $E_{RMS}$ metric (equation (A5)), $n$ denotes the number of grid cells in the spatial domain when calculating $\Phi_{E_a}$ and $\Phi_{S_u}$ (i.e., $\alpha$ in $E_{SP}$), while it corresponds to the length of the calibration period (i.e., $n = t$) in the calculation of $\Phi_{S_t}$. All constituents of $\Phi_{MV}$ (i.e., $\Phi_Q$, $\Phi_{E_a}$, $\Phi_{S_u}$, and $\Phi_{S_t}$) vary in the same range from 0 to $\pm\infty$. Therefore, $\Phi_{MV}$ has the same range of values with an optimal value of 0. The dynamically dimensioned search algorithm (Tolson & Shoemaker, 2007) is used for parameter estimation using 3,000 iterations. Daily streamflow data are used for model calibration, while monthly SRS data are preferred to avoid errors due to potential time lags among satellite sensor measurements and model simulations.

#### 4.3.2.3. Contribution of Individual Variables to Multivariate Calibration

Assessing the individual contribution of variables used in multivariate calibration is rarely done, although it can help quantify trade-offs in modeling flux and states variables (Koppa et al., 2019). Here the contribution of each SRS data set to the multivariate calibration case is investigated with a leave-one-out approach. The procedure consists in removing one SRS data type from the calibration case MV and evaluating the predictive skill of the model. In addition, multivariate calibration without streamflow data, and thus exclusively based on SRS data, is tested to determine the potential of SRS data for hydrological model calibration in regions where little or no streamflow data are available. Consequently, four additional objective functions (Table 3) are used for multivariate calibration cases without $E_a$ (case MV-$E_a$), without $S_u$ (case MV-$S_u$), without $S_t$ (case MV-$S_t$), and without streamflow (case MV-$Q$).
4.4. Postcalibration Model Evaluation

The predictive skill of the model is evaluated by assessing the transferability of the global parameters across temporal periods and spatial scales obtained by the abovementioned calibration strategies. First, the temporal transferability is evaluated following a split-sample test that consists in assessing performances for a period that is different from the calibration period (Klemes, 1986). Secondly, spatial scale transferability is evaluated by using different grid cell (i.e., pixel) sizes as modeling resolution (Kumar et al., 2013; Samaniego et al., 2010). The global parameters of the model for all calibration cases are obtained for a resolution of 0.25° (~28 km, i.e., 619 pixels in the basin), and the same parameters are used to run the model without recalibration at four different finer scales: 0.125° (~14 km, i.e., 2,320 pixels), 0.0625° (~7 km, i.e., 8,974 pixels), 0.03125° (~3.5 km, i.e., 35,231 pixels), and 0.015625° (~1.75 km, i.e., 139,494 pixels). The evaluation data for model parameter transferability are streamflow for streamflow, using the Kling-Gupta efficiency ($E_{KG}$), and fine-scale $T_s$ data to evaluate $E_u$ and $S_u$ using $r_s$, while no high-resolution data are available for $S_t$ evaluation. $T_s$ is used as proxy data for the evaluation of $S_u$ and $E_u$ because past studies found significant negative correlation between $T_s$ and $S_u$ (Kumar et al., 2013; Lakshmi et al., 2003; Wang et al., 2007) and a control of $T_s$ over $E_u$ (Boni et al., 2001; Lakshmi, 2000).

Following Biondi et al. (2012), supplemental skill metrics different from those used in model calibration are computed for a thorough model evaluation because every metric has its own limitations (Fowler et al., 2018; Knoben et al., 2019; Santos et al., 2018; Schaefl & Gupta, 2007). In addition to $E_{NS}$, $E_{NSlog}$, $E_SP$, $E_{RMS}$, and $r_s$, the $E_{KG}$ is reported for model evaluation.

5. Results and Discussions

The following section presents and discusses the results of model performances for different variables used in the calibration procedure. The results refer to the evaluation period when analyzing the results if not clearly specified. However, both calibration and evaluation results are presented in figures. Hereafter, the SRS data sets are called reference data, as they are not direct observations. Detailed results on model performances for each of the four climatic zones in the VRB are provided as supporting information (Figures S17–S33).

5.1. Model Performance for Streamflow

The model performance for streamflow ($Q$) at the 11 gauging stations is given in Figure 2. For the calibration period, the mean $E_{KG}$ is 0.67 ($E_{NS} = 0.71$, $E_{NSlog} = 0.72$) for the model calibration with only $Q$ data (i.e., case...
The performance of $Q$ in the calibration period decreases when multiple variables are used to constrain the parameter search. The mean $E_{KG}$ is 0.55 ($E_{NS} = 0.57$, $E_{NSlog} = 0.66$) for the multivariate calibration (i.e., case MV), corresponding to a decrease of 18% compared to that in case $Q$.

Regarding the other multivariate calibration cases, the best performance with respect to $Q$ is obtained in case MV-St with a mean $E_{KG}$ of 0.65 ($E_{NS} = 0.67$, $E_{NSlog} = 0.70$), which represents a slight decrease of 3% compared to that in case $Q$, while the weakest performance is given by case MV-Q with a mean $E_{KG}$ of 0.19 ($E_{NS} = 0.33$, $E_{NSlog} = -0.38$). Differences in measurement scales between $Q$ data and satellite products (i.e., river dimensions vs. pixel size) can justify a low performance for case MV-Q (see section 5.5). In general, all calibration cases give a good timing of $Q$ with a mean correlation coefficient of $r > 0.79$, but they underestimate it with a mean bias of $\beta < 0.85$, except case MV-Q, which shows overestimation with a positive bias ($\beta = 1.19$). They all show a higher variability of $Q$ than the observed data, with mean $\gamma > 1.04$, except case Q ($\gamma = 0.98$) and case MV-Q ($\gamma = 0.40$), which produce a lower variability. A subset of the hydrographs of three stations from different climatic zones are depicted in Figure 3. More statistics on the model performance along with the complete hydrographs and flow signatures (i.e., flow duration curves and seasonal streamflow) are provided in the supporting information (Table S1 and Figures S1–S11).

During the evaluation period, as compared to the calibration period, the model performance for the mean $E_{KG}$ decreases by 2% for case Q (from 0.67 to 0.66) and case MV-St (from 0.65 to 0.64) and 90% for case MV-Q (from 0.19 to 0.02), while it increases by 7% for case MV-Ea (from 0.61 to 0.65), 12% for case MV-Su (from 0.56 to 0.62), and 14% for case MV (from 0.55 to 0.63). Considering the mean $E_{KG}$, case MV performs less well than case Q by 11% on average, which means 18% less during the calibration and 4% less during the evaluation period. The deterioration of streamflow performance in a multivariate calibration setting is also reported in previous studies (Bai, Liu, & Liu, 2018; Livneh & Lettenmaier, 2012; Pomeon et al., 2018; Rakovec et al., 2016). However, this is largely an artefact of Type I error (i.e., falsely accepting poor models; Beven, 2010) induced by the $Q$ only calibration, resulting in inconsistency in the representation of processes (Gupta et al., 2012; Hrachowitz et al., 2014). In addition, the performance of $Q$ slightly increases when $E_{a}$ (+7%) or $S_{i}$ (+11%) are left out of the multivariate calibration with case MV during the evaluation period. Therefore, as shown in Figure 2, the combinations $Q + S_{i} + S_{u}$ (i.e., case MV-Ens) and $Q + S_{i} + E_{u}$ (i.e., case MV-St) are the best for streamflow prediction, while $Q + S_{i} + E_{a}$ (i.e., case MV-Su) performs similar to $Q + S_{i} + E_{u}$ (i.e., case MV).
5.2. Model Performance for Terrestrial Water Storage

The statistics for the monthly terrestrial water storage ($S_t$) anomalies are given in Figure 4. The results per climatic zone is provided in the supporting information (Figure S19). A similar trend in skill scores (i.e., $\text{ERMS}$ and $r$) among all calibration cases is observed in the calibration and evaluation periods with weaker scores during evaluation.

The evaluation period is characterized by a substantial improvement from model case $Q$ (median $\text{ERMS} = 8.41$ cm, $r = 0.73$) to case MV ($\text{ERMS} = 7.38$ cm, $r = 0.81$). In general, all multivariate calibration cases reproduce the variability in $S_t$ better than did case $Q$. Previous studies also reported improvement of $S_t$ prediction in multivariate settings (Chen et al., 2017; Livneh & Lettenmaier, 2012; Werth et al., 2009). The lowest performance increase is observed when $E_a$ ($\text{ERMS} = 8.21$ cm, $r = 0.76$) or $S_t$ ($\text{ERMS} = 8.18$ cm, $r = 0.78$) is removed from the multivariate setting. The best prediction is obtained with case MV‐$Q$, yielding median $\text{ERMS}$ of $6.65$ cm and $r$ of $0.84$.

Figure 5a shows the climatology of the basin‐averaged $S_t$ for all models. The full monthly time series (Figure S17) and the climatological trends (Figure S18) per climatic zone are provided in the supporting information. The temporal dynamics of the normalized GRACE‐derived $S_t$ is well reproduced by all models ($r > 0.89$) with different degrees of underestimation from September to March, which is a period with little or no rainfall, and slight overestimation from April to June, which is the beginning of the rainy season. All models fit well the period July–August, which is the wettest period of the rainy season. In general, case MV‐$E_a$ shows the highest deviation from the satellite signal ($r = 0.90$) followed by case $Q$ ($r = 0.92$). Removing spatial patterns of $E_a$ from the calibration leads to an $S_t$ overestimation during the rainy season and an underestimation during the dry season. The same trend can be observed when only $Q$ is used for model calibration (i.e., case $Q$). The $S_t$ simulation improves in the multivariate calibration including $Q$ (i.e., case MV, $r = 0.97$), but the best match is obtained when $Q$ is left out (i.e., case MV‐$Q$, $r = 0.99$). When GRACE‐derived $S_t$ is excluded from the parameter estimation (i.e., case MV‐$S_t$), the model still performs well for $S_t$ climatology with $r = 0.96$. Consequently, $E_a$ is the most critical variable for predicting the $S_t$ signal in the proposed multivariate calibration setting, while $S_u$ is less critical, probably because the GRACE‐derived $S_t$ signal already accounts for $S_u$ (Li et al., 2012).

5.3. Model Performance for Soil Moisture

The climatology of the basin‐averaged $S_u$ for all models. The full monthly time series (Figure S17) and the climatological trends (Figure S18) per climatic zone are provided in the supporting information. The temporal dynamics of the normalized GRACE‐derived $S_u$ is well reproduced by all models ($r > 0.89$) with different degrees of underestimation from September to March, which is a period with little or no rainfall, and slight overestimation from April to June, which is the beginning of the rainy season. All models fit well the period July–August, which is the wettest period of the rainy season. In general, case MV‐$E_a$ shows the highest deviation from the satellite signal ($r = 0.90$) followed by case $Q$ ($r = 0.92$). Removing spatial patterns of $E_a$ from the calibration leads to an $S_u$ overestimation during the rainy season and an underestimation during the dry season. The same trend can be observed when only $Q$ is used for model calibration (i.e., case $Q$). The $S_u$ simulation improves in the multivariate calibration including $Q$ (i.e., case MV, $r = 0.97$), but the best match is obtained when $Q$ is left out (i.e., case MV‐$Q$, $r = 0.99$). When GRACE‐derived $S_u$ is excluded from the parameter estimation (i.e., case MV‐$S_u$), the model still performs well for $S_u$ climatology with $r = 0.96$. Consequently, $E_a$ is the most critical variable for predicting the $S_u$ signal in the proposed multivariate calibration setting, while $S_u$ is less critical, probably because the GRACE‐derived $S_u$ signal already accounts for $S_u$ (Li et al., 2012).
reference $S_u$ for all months, while case MV and case MV-Q have similar performances ($r \approx 0.94$) but with a better fit of the reference $S_u$ in the recession limb. It can be inferred that $Q$ is the most critical variable for $S_u$ reproduction during the rising limb, while $S_t$ and satellite $S_u$ improve the simulation during the recession limb. Case $Q$ outperforms all multivariate calibration cases when soil water content increases, and it underperforms them when the maximum water content is reached and starts decreasing. The overall lowest performance is given by case $Q$ ($r = 0.92$), followed by case MV-$Q$ ($r = 0.93$), suggesting that $Q$ alone is not sufficient for predicting the temporal dynamics of $S_u$, but it remains useful in the multivariate calibration setting. Surprisingly, $E_a$ does not bring substantial information to the multivariate prediction of $S_u$. Contrastingly, Pomeon et al. (2018) obtained a slight improvement in $S_u$ simulation (+7%) when using absolute values of satellite $E_a$ in their multivariate calibration. The model performance in reproducing spatial patterns is measured with $ESP$, and its components (i.e., $r_s$, $\gamma$, and $\alpha$) are summarized in Figure 6a. The results for each month (Figure S41) and for the climatic zones (Figures S22–S25) are provided in the supporting information.

Figure 5. Climatology of (a) terrestrial water storage, (b) soil moisture, and (c) actual evaporation with the Pearson correlation coefficient ($r$) indicating the performance of all model calibration cases.
Consequently, satellites $S_u$ and $S_t$ are the most important variables for improving the spatial patterns of modeled $S_u$. Better simulation of $S_u$ in multivariate settings is also reported by Lopez et al. (2017) using $E_a + S_u$ calibration and by Pomeon et al. (2018) with $Q + E_a$ calibration.

Figure 6. Spatial statistics of model performance for (a) soil moisture and (b) actual evaporation. The best score is 1 for all the metrics. The number of elements per boxplot corresponds to the number of months in the calibration period ($n = 72$) or in the evaluation period ($n = 48$). The colors correspond to the model calibration cases.

Consequently, satellites $S_u$ and $S_t$ are the most important variables for improving the spatial patterns of modeled $S_u$. Better simulation of $S_u$ in multivariate settings is also reported by Lopez et al. (2017) using $E_a + S_u$ calibration and by Pomeon et al. (2018) with $Q + E_a$ calibration.

Figure 7. Long-term monthly average of (a) soil moisture and (b) actual evaporation for all model calibration cases over the simulation period (2003–2012). The reference map represents the satellite product (ESA CCI for $S_u$ and GLEAM for $E_a$). Masked pixels are gaps in satellite measurements or lake areas not modeled in mHM. The values are normalized for better emphasizing on patterns and using a unique color scale.
The long-term monthly average (2003–2012) of $S_{aq}$ is illustrated in Figure 7a. See section 5.5.2 for $S_{aq}$ comparison with $T_s$. Although the spatial patterns of modeled $S_{aq}$ in the multivariate cases are still different from the reference $S_{aq}$ (Figure 7a), they are better than the Q-only case.

### 5.4. Model Performance for Evaporation

The climatology of basin-averaged reference (i.e., GLEAM) actual evaporation ($E_a$) and different modeled $E_a$ are depicted in Figure 5c. The maps of climatology (Figures S15 and S16), the full monthly time series (Figure S26), and the climatological trends (Figure S27) per climatic zone are provided in the supporting information. All calibration cases give a good performance ($r > 0.91$), reproducing well $E_a$ seasonality during both the calibration and evaluation periods.

The best performance is obtained with case MV-$Q$ ($r = 0.99$). In general, all simulations tend to underestimate $E_a$. During the rising limb (February–August), corresponding to the increasing rainfall period, the highest deviation is given by case MV-$E_a$, although the overall performance is good ($r = 0.98$). Mismatches are more prominent during the recession limb (September–January), where all modeled $E_a$ decrease faster than the reference. The highest deviation is observed when only $Q$ data are used for model calibration ($r = 0.92$), followed by case MV-$S_t$ ($r = 0.96$). It can be inferred that the model case MV-$S_t$ is missing adequate information on the available water amount to be evaporated, which can be obtained from the satellite $S_t$ signal. During the recession period, little to no rainfall occurs in the basin, but a part of antecedent rainfalls is stored in reservoirs and lakes, which represent a major source of land evaporation. It can be argued that $Q$ alone is not sufficient for modeling $E_a$, while $S_t$ brings additional and useful information for simulating $E_a$, which supports our research hypothesis. The performance of case MV-$S_{aq}$ is similar to that of case MV ($r \approx 0.98$), meaning that $S_t$ is not critical for predicting the temporal dynamics of $E_a$. Moreover, satellite $E_a$ improves the modeled $E_a$ during water accumulation in the basin (i.e., February–August) and is no longer critical when the basin is not water limited (i.e., September–January). This result suggests that the model can mainly rely on GRACE-derived $S_t$ to reproduce $E_a$. Similar results on the good estimation of $E_a$ with GRACE-derived $S_t$ are found in literature (e.g., Bai, Liu, & Liu, 2018; Livneh & Lettenmaier, 2012; Rakovec et al., 2016). Pomeon et al. (2018) also obtained a higher model performance for $E_a$ in their multivariate setting (i.e., $Q + E_a$) with mHM in West Africa.

Figure 6b gives the spatial pattern efficiency of $E_a$ for all model calibration cases. The results for each month (Figure S41) and for the climatic zones (Figures S28–S31) are provided in the supporting information. In general, the performance decreases from the calibration period to the evaluation period, and the modeled $E_a$ with all model calibration cases has higher spatial pattern efficiency scores ($E_{SP} > 0.25$) compared to modeled $S_t$ ($E_{SP} < 0.1$). All multivariate calibration cases outperform the Q-only calibration, giving the lowest performance with median $E_{SP} = 0.28$. The Q-only calibration gives a good spatial correlation ($r_s = 0.8$) but overestimates the variability ($\gamma = 1.28$) and struggles to match the spatial location of grid cells ($\alpha = 0.39$) of $E_a$. The best spatial pattern matching is given by case MV with median $E_{SP} = 0.46$ ($r_s = 0.86$, $\gamma = 0.95$, and $\alpha = 0.53$). Removing $Q$ from the multivariate setting (i.e., case MV-Q) results in an underestimation of the spatial variability of $E_a$, with median $E_{SP} = 0.43$ ($r_s = 0.88$, $\gamma = 0.84$, and $\alpha = 0.52$). In contrast, the spatial variability of $E_a$ is overestimated for case MV-$E_a$ with median $E_{SP} = 0.45$ ($r_s = 0.85$, $\gamma = 1.13$, and $\alpha = 0.52$), while case MV-$S_{aq}$ yields a lower spatial location score with median $E_{SP} = 0.42$ ($r_s = 0.85$, $\gamma = 0.95$, and $\alpha = 0.50$). The spatial pattern performance of $E_a$ is more sensitive to the removal of $S_t$ as shown by case MV-$S_t$ with median $E_{SP} = 0.35$ ($r_s = 0.81$, $\gamma = 1.07$, and $\alpha = 0.49$). These results indicate that spatial patterns of $S_t$ can improve the spatial patterns of $E_a$ and $S_t$ is critical for reproducing both the temporal and spatial dynamics of $E_a$. Demirel et al. (2018) similarly reported better spatial pattern performance for $E_a$ when using a multivariate setting (i.e., $Q + E_a$) compared to that when using the Q-only calibration.

Figure 7b illustrates the long-term (2003–2012) monthly average of $E_a$. See section 5.5.2 for $E_a$ comparison with $T_s$. The southern region of the basin, with a subhumid climate, is where the multivariate calibration cases show more differences in spatial patterns compared to case Q. Besides the south-north differences, it is interesting to see strong differences in the west-east variability of the spatial pattern. As the southern part is subhumid ($E_a \geq 70\%$), small variations in $E_a$ are not well represented when the model is calibrated using
only $Q$ compared to those in the semiarid northern part. These findings are in agreement with the study of Rakovec et al. (2016), which revealed a more pronounced sensitivity in $E_a$ estimation in humid catchments in Europe through a multivariate calibration setting ($Q + S_i$). Similar results are obtained by Bai et al. (2016) when testing different $E_p$ formulas in China. Contrastingly, Bai, Liu, and Liu (2018) found that their multivariate calibration setting ($Q + S_i$) benefitted more to $E_a$ simulation in dry catchments than in wet catchments in China.

5.5. Parameter Transferability Across Spatial Scales
5.5.1. Streamflow Evaluation Across Spatial Scales
The model performance of streamflow in terms of scale transferability of the global parameters is given in Figure 8. The differences in model performance among calibration cases are conserved across spatial scales, with a median coefficient of variation of 1.6% for $E_{KG}$, 0.1% for $r$, and 6.4% for $\beta$.

5.5.2. Spatial Pattern Evaluation Across Spatial Scales
Long-term monthly maps of $S_u$ (Figure 9a) and $E_a$ (Figure 9b) are plotted along with $T_s$ maps at various spatial resolutions. Here only the coarsest and finest resolutions (i.e., 28 and 1.75 km) are shown, but the same figures with intermediate resolutions and the spatial pattern efficiency are given in the supporting information (Figures S34–S37 and S32–S33).

The patterns of $S_u$ is consistent with the patterns of $T_s$ because the expectation is that the higher the $T_s$, the lower the $S_u$ and vice versa (Figure 9a). For semiarid regions, $E_a$ largely depends on water availability (i.e., rainfall) and is dominant for open water storages. In the VRB, $E_a$ depicts an opposite pattern to $T_s$, which is shown in Figure 9b. The reproduction of both $S_u$ and $E_a$ in the multivariate calibration cases and across spatial scales show more plausible patterns with $T_s$, which are well preserved across scales with higher consistency, than their representation with case $Q$. The maps of the temporal correlation per grid cell are provided in the supporting information (Figures S35 and S37).

5.6. Benefit of Spatial Patterns and Data Types in Multivariate Calibration
5.6.1. Analysis of the Lake Volta Region
Evidence of the benefit of multivariate calibration with SRS data is exemplified in Figure 10 by zooming-in on the Volta Lake region in the southern part of the VRB (cf. Figure 1). Notwithstanding that mHM does not have a lake module, it is nicely noticeable that the model represents the heterogeneity in spatial patterns with the multivariate calibration cases being better than the $Q$-only calibration case. As it should be
expected from the $T_s$ patterns, case MV shows higher $S_u$ and $E_a$ over the Lake Volta but with lower $E_a$ in its surroundings. This improvement in spatial patterns is not observed with case Q, confirming the limitations of the $Q$-only calibration and emphasizing the importance of patterns of SRS data for model calibration. Moreover, it can be inferred from the results that $S_t$ is the most important variable for representing the lake area, while $Q$ is the less critical variable.

Figure 9. Long-term monthly average of land surface temperature compared to (a) soil moisture and (b) actual evaporation for all model calibration cases over the simulation period (2003–2012) at various spatial resolutions. The values are normalized for better emphasizing on patterns and using a unique color scale.

Figure 10. Comparison of the spatial patterns of (top row) soil moisture and (bottom row) actual evaporation with land surface temperature (first map from left) in the Volta Lake region. The $T_s$ map used as benchmark shows the Lake Volta depicted in dark blue with the lowest temperature in the region. The ability of the mHM model to highlight the lake area is assessed with the patterns of $S_u$ and $E_a$ for all model calibration cases.
5.6.2. Comparison of the Multivariate Calibration Cases to the Benchmark

Figure 11 gives the gain or loss in model performance with different multivariate calibration cases compared to the Q-only calibration case (i.e., the benchmark).

In general, the multivariate calibration cases show higher model predictive skill for many components of the hydrological system (i.e., Q, S_t, S_u, and E_a) when compared to the Q-only calibration. The decrease in the model performance for Q is usually counterbalanced with an increase in performance for S_t, S_u, and E_a, which might simply be the result of a solution to the artifacts caused by the Q-only calibration. These results reveal that the simulation of spatial patterns of S_u benefits most from the multivariate settings, followed by the simulation of E_a and S_t, while the decrease in Q performance varies widely depending on the multivariate calibration cases. A summary of the importance of each variable in predicting the others (i.e., Q, S_t, S_u, and E_a) is given in Table 4.

The most determinant variable for streamflow prediction in a multivariate setting is streamflow itself, similarly for S_u, while it is E_a for S_t and S_u for E_a. Zeng and Cai (2016) also found that S_t controls the temporal variability of E_a. Surprisingly, E_a is the less critical variable for S_t prediction. However, it is worth stressing that only the spatial patterns of satellite E_a is exploited here. Moreover, E_a calculation in the model setup might be a reason of the limited contribution of satellite E_a in S_u prediction. The E_a calculation (cf. section 4.2) is done with time-variant and gridded leaf area index data that imposes heterogeneity on modeled E_a (Birhanu et al., 2019). Consequently, additional contribution from the satellite E_a in S_u prediction is expectedly limited in case the leaf area index data are in agreement with the satellite E_a. Moreover, not explicitly weighting the components of the multivariate objective function might have led to implicit weighting, which led to the artifact that some variables are not very good predictors for themselves.

5.7. Summary and Outlook

This work is a follow-up on several recent studies on multiobjective calibration and spatial pattern improvement in hydrological modeling (e.g., Demirel et al., 2018; Koch et al., 2018; Nijzink et al., 2018; Stisen et al., 2018; Yassin et al., 2017; Zink et al., 2018). The proposed multivariate calibration approach is a step forward in improving the realism of hydrological model predictions (Baroni et al., 2019; Clark et al., 2015; Rakovec et al., 2016) because not only a reliable temporal dynamic in the modeling objective but also plausible spatial patterns of several hydrological processes simultaneously are sought for. A key element of our study is the assessment of the plausibility of spatial patterns of soil moisture and evaporation with independent data of land surface temperature not used during the model calibration. With respect to the obtained performances, it can be concluded that spatial patterns of satellite data are a highly relevant and robust feature that can be used in multivariate calibration to improve the overall representation of the hydrological system even with trade-offs among the variables, which thereby confirms our research hypothesis.

A rigorous comparison of the proposed bias-insensitive metric with other spatial pattern metrics is left for future work. Further investigations can focus on setting a threshold for the acceptability of the modeled spatial patterns, which was not required here as the goal was to check the increase or decrease of spatial pattern performance rather than determining whether the patterns are good or bad in an absolute sense, when switching between streamflow-only and multivariate calibration cases.

Our methodology lacks in situ data for model evaluation, except streamflow. However, in situ measurements of soil moisture, evaporation, and terrestrial water storage at a large scale are rather rare (Vereecken et al.,

---

**Table 4**

*Importance of Different Variables in Predicting Others in a Multivariate Calibration Setting*

<table>
<thead>
<tr>
<th>Predictands</th>
<th>Temporal dynamics</th>
<th>Spatial patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictors</td>
<td>Q</td>
<td>S_t</td>
</tr>
<tr>
<td>Q</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>S_t</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>E_a</td>
<td>++</td>
<td>++++</td>
</tr>
<tr>
<td>S_u</td>
<td>++</td>
<td>+</td>
</tr>
</tbody>
</table>

*Note.* The degree of importance is as follows: low (+), moderate (++), high (+++), and very high (++__).
2008; Zink et al., 2018) and are also subject to uncertainties due to the nonuniformity of the data collection in space (Stisen et al., 2011). As we focus on spatial pattern assessment in this study, satellite data remain the only possible option for our large study area in West Africa, where ground measurements are a luxury (Dembélé et al., 2019).

The presented multivariate calibration reveals trade-offs among the objective functions for streamflow and for satellite data. However, trade-offs cannot be avoided as they originate from errors in input data, model structure, and lack of knowledge of the hydrological system (Bergström et al., 2002; Gupta et al., 1998; Yassin et al., 2017). Moreover, it was a deliberate choice to equally weight the components of the multivariate objective function (equation (4)) because no prior knowledge on the importance of each variable was available, and it was an objective of this study to know their contributions in the calibration procedure. In such situation, the default choice is to weight them equally (Bergström et al., 2002; Stisen et al., 2018). Weights are sometimes assigned to objective function components by iterative optimization testing different weights, which is, however, computationally demanding. It is also possible to transform the components of the multivariate objective functions to solve differences in their magnitudes (Madsen, 2003; Zink et al., 2018), but the effects of such transformations on the calibration procedure are unknown and they are not required if the metrics are dimensionless or of the same order of magnitude (Bergström et al., 2002). For completeness, Pareto plots showing the absolute values and the trade-off among the used objective functions are provided in the supporting information (Figures S38 and S39).

The climatic inputs influence somehow the spatial variability of the hydrological processes due to the aridity gradient in the VRB. The detailed results are valid for the VRB, but they can be generalized to regions with similar hydroclimatic characteristics. However, the applicability of the proposed multivariate calibration framework is, in principle, universal, as long as a DHM is used and spatial data sets are available. Further research can explore the applicability of the presented multivariate calibration strategies in different hydroclimatic regions with different spatial data sources, and different DHMs to understand how the model structure interacts with the performance of different calibration strategies. Choosing an adequate hydrological model (Addor & Melsen, 2019) is key to any good experiment. The MPR scheme used in mMHM might have facilitated to some extent the reproduction of spatial patterns, but the MPR scheme can be similarly implemented with other models as demonstrated by previous studies (e.g., VIC and PCR-GLOBWB models; Mizukami et al., 2017; Samaniego et al., 2017). A sensitivity analysis to identify the model parameters that influence the representation of spatial patterns is a recommended outlook.

Future methodological developments could in particular focus on improved formulation of the multiobjective functions inspired by previous findings on the following topics: fitting of low flows and system signatures (Fowler, Peel, et al., 2018; Hrachowitz et al., 2014; Krause et al., 2005; Pushpalatha et al., 2012), gauge measurement weighting (Madsen, 2003), or subperiod calibration (Gharari et al., 2013). Additional key questions to address in this context include the model structural deficiencies (Gupta et al., 1998; Gupta et al., 2012) and the uncertainties of modeling data sets (i.e, input, calibration, and evaluation data), which can lead to erroneous model rejection (Beven, 2010, 2018, 2019b).

The above efforts in model improvement are particularly important for prediction in a changing environment (Fowler et al., 2018), and they can set avenues for prediction in ungauged basins solely from space.

6. Conclusion

This study presents a calibration approach using multiple data sources simultaneously, with the specificity of integrating only spatial patterns of satellite remote sensing data in the parameter estimation procedure. A bias-insensitive and multicomponent metric is proposed for spatial pattern matching. The study is carried out in the Volta River basin in West Africa. Results reveal the benefit of the multivariate calibration setting over the traditional calibration using only streamflow data. The main findings are as follows:

• Streamflow is a necessary variable, but alone it is not sufficient for reliably reproducing other hydrological fluxes and states.
• Spatial patterns of satellite data, without the absolute values, can be incorporated in the calibration procedure with bias-insensitive metrics.
• Multivariate calibration based on streamflow and satellite data can improve the overall representation of the hydrological system and thereby increase the model predictive skill.

• The reduction in streamflow performance in a multivariate setting is largely compensated by the gain in performance for other hydrological processes (i.e., terrestrial water storage, soil moisture, and evaporation).

We advocate for the adoption of multivariate calibration procedure focusing on spatial patterns in distributed hydrological models because it is a robust approach for addressing equifinality, reducing uncertainties, and enhancing the predictive skill of hydrological models in a changing environment.

**Appendix A: Equations**

**Euclidean Distance**

The Euclidean distance ($DE$) between two points $X$ and $Y$ of coordinates $(x_1, x_2, ..., x_n)$ and $(y_1, y_2, ..., y_n)$ in an $n$-dimensional space (Upton & Cook, 2014) is given by

$$DE = \sqrt{\sum_{i}^{n} (x_i - y_i)^2}.$$  \hfill (A1)

**Nash-Sutcliffe Efficiencies**

The Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) of streamflow ($ENS$) and the Nash-Sutcliffe efficiency of the logarithm of streamflow ($ENS_{log}$) are formulated as follows:

$$ENS = 1 - \frac{\sum_i^t (Q_{mod}(t) - Q_{obs}(t))^2}{\sum_i^t (Q_{obs}(t) - \bar{Q}_{obs})^2},$$  \hfill (A2)

and

$$ENS_{log} = 1 - \frac{\sum_i^t \left[\log(Q_{mod}(t)) - \log(Q_{obs}(t))\right]^2}{\sum_i^t \left[\log(Q_{obs}(t)) - \bar{\log(Q_{obs})}\right]^2},$$  \hfill (A3)

where $Q_{mod}$ and $Q_{obs}$ are the modeled and observed streamflow, respectively, and $t$ is the number of time steps. $ENS$ and $ENS_{log}$ range from $-\infty$ to their perfect score, that is, 1. These two metrics are used as skill scores to identify and discard parameter sets that provide implausible representations of the system and to subsequently better predict both high and low flows (Krause et al., 2005; Oudin et al., 2006; Pushpalatha et al., 2012; Santos et al., 2018).

**z-Scores and Root-Mean-square Error**

The standardized values (z-scores, $Z_X$) of $X_{mod}$ and $X_{obs}$ and the root-mean-square error ($ERMS_X$) of a modeled variable ($X_{mod}$) and an observed variable ($X_{obs}$) of $n$ elements, are defined as follows:

$$Z_X = \frac{X - \mu}{\sigma},$$  \hfill (A4)

and

$$ERMS_{X_{mod}, X_{obs}} = \sqrt{\frac{1}{n} \sum_{i}^{n} (X_{mod,i} - X_{obs,i})^2}.$$  \hfill (A5)

where $\mu$ and $\sigma$ are the mean and the standard deviation of a given variable $X$, respectively.

**Kling-Gupta efficiency**

The Kling-Gupta efficiency ($E_{KG}$) was introduced by Gupta et al. (2009) and modified by Kling et al. (2012) to avoid some limitations of $ENS$. $E_{KG}$ combines correlation, bias, and variability measures and is defined as follows:
where $r$ is the Pearson correlation coefficient, $\beta$ represents the bias term (i.e., ratio of means), and $y$ is the variability term (i.e., ratio of coefficients of variation).

Acknowledgments

The authors thank the providers of the data sets (see Table 2) used in this study. They thank the mHM team at CHS/UFZ (Germany), particularly Oldrich Rakovec and Pallav Kumar Shrestha, for guidance in setting up the model. They are grateful to the streamflow data providers and the National Hydrologic Services in the VBA member countries. The streamflow data sets were obtained from the Direction Générale des Ressources en Eau (DGRE) of Burkina Faso, the Hydrological Services Department (HSD) of Ghana, the Volta Basin Authority (VBA), and the Direction Générale de l'Assainissement (DGEA) of Togo through a contact at the Université d'Abomey Calavi (UAC). Data available from the abovementioned data providers can be obtained from them. The streamflow data sets can be accessed from the Global Runoff Data Centre (www.bafg.de/GSRC). The model outputs are available at the Zenodo website (https://doi.org/10.5281/zenodo.3531873). The first author thanks the Swiss Confederation for financial support through the Swiss Government Excellence Scholarship (2016/0531/Burkina Faso/OF) and the Swiss National Science Foundation (SNF, PP00P2_157611). The authors thank Luis Samaniego (Associate Editor), Juraj Parajka, Simon Stisen, and the anonymous reviewer for their constructive comments that contributed in improving the final manuscript.

References


Food and Agriculture Organization/GLOBAL Information and Early Warning System (GIEWS), Simulation of snow, temperature and crop situation 1998, 8 pp, Food and Agriculture Organization (FAO)/GLOBAL Information and Early Warning System (GIEWS), https://www.fao.org/docrep/004/x0505e/x0505e00.htm.


