Performance evaluation of multiple satellite rainfall products for Dhidhessa River Basin (DRB), Ethiopia

Gizachew Kabite Wedajo\textsuperscript{a,b,*}, Misgana Kebede Muleta\textsuperscript{c}, Berhan Gessesse Awoke\textsuperscript{b,d}

\textsuperscript{a}Department of Earth Sciences, Wollega University, P.O.Box 395, Nekemte, Ethiopia
\textsuperscript{b}Department of Remote Sensing, Entoto Observatory Research Center, Ethiopia Space Science Technology Institute, P.O.Box 33679, Addis Ababa, Ethiopia
\textsuperscript{c}Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, California, 93407
\textsuperscript{d}Department of Geography and Environmental Studies, Kotebe Metropolitan University, Addis Ababa, Ethiopia

*Correspondence: Email: Kabiteg@gmail.com; Mobile: 0911951952
Abstract

Precipitation is a crucial driver of hydrological processes. Ironically, reliable characterization of its spatiotemporal variability is challenging. Ground-based rainfall measurement using rain gauges can be more accurate. However, installing a dense gauging network to capture rainfall variability can be impractical. Satellite-based rainfall estimates (SREs) can be good alternatives, especially for data-scarce basins like in Ethiopia. However, SREs rainfall is plagued with uncertainties arising from many sources. The objective of this study was to evaluate the performance of the latest versions of several SRE products (i.e., CHIRPS2, IMERG6, TAMSAT3 and 3B42/3) for the Dhidhessa River Basin (DRB). Both statistical and hydrologic modelling approaches were used for the performance evaluation. The Soil and Water Analysis Tool (SWAT) was used for hydrological simulations. The results showed that whereas all four SREs products are promising to estimate and detect rainfall for the DRB, the CHIRPS2 dataset performed the best at annual, seasonal and monthly timescales. The hydrologic simulation based evaluation showed that SWAT’s calibration results are sensitive to the rainfall dataset. The hydrologic response of the basin is found to be dominated by the subsurface processes, primarily by the groundwater flux. Overall, the study showed that both CHIRPS2 and IMERG6 products can be reliable rainfall data sources for hydrologic analysis of the Dhidhessa River Basin.

Keywords: Satellite-based rainfall estimates; Dhidhessa River Basin; Performance evaluation; Statistical evaluation; Hydrological modelling performance.
1. Introduction

Precipitation is an important hydrological component (Behrangi et al., 2011; Meng et al., 2014). Accurate representation of its spatiotemporal variability is crucial to improve hydrological modelling (Grusson et al., 2017). Ironically, precipitation is one of the most challenging hydrometeorological data to be accurately represented (Yong et al., 2014). Climatic and topographic conditions are the primary factors that affect the accuracy of rainfall measurements.

Rainfall is measured either using ground-based (i.e., rain gauge and radar) or satellite sensors, where all measurement methods exhibit limitations (Thiemig et al., 2013). Ground-based rainfall measurements are direct and generally accurate near the sensor location. However, they are either of poor density to represent spatial and temporal variability of precipitation, or may not even exist (e.g., radars), especially in developing countries (Behrangi et al., 2011). Ground-based rainfall measurement techniques provide point and incomplete measurements (Kidd et al., 2012; Maggioni et al., 2016). It may also be infeasible to install and maintain dense ground-based gauging stations in remote areas like mountains, deserts, forests and large water bodies (Dinku et al., 2018; Tapiador et al., 2012). However, satellite-based rainfall estimates (SREs) provide high-resolution precipitation data including in areas where ground-based rainfall measurements are impractical, sparse, or non-existent (Stisen and Sandholt, 2010).

Consequently, high-resolution precipitation products have been developed over the last three decades. These products include Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007), the Precipitation Estimation from Remote Sensing Information Using Artificial Neuron Networks (PERSIANN; Sorooshian et al., 2000), Climate Prediction Center (CPC) morphing algorithm (CMORPH) (Joyce et al., 2004), African Rainfall Climatology (ARC) (Xie and Arkin 1995), Tropical Applications of Meteorology using SATellite (TAMSAT) (Maidment et al., 2017) and the Climate Hazards Group InfraredRed Precipitation with Stations (CHIRPS) (Funk et al., 2015). The consistency, spatial coverage, accuracy and spatiotemporal resolution of SREs have improved overtime (Behrangi et al., 2011).

As indirect rainfall estimation techniques, SREs products possess uncertainties resulting from errors in measurement, sampling, retrieval algorithm, and bias correction.
processes (Dinku et al., 2010; Gebremichael et al., 2014; Tong et al., 2014). Local topography and climatic conditions can also affect the accuracy of SREs estimation (Bitew and Gebremichael, 2011). Hence, SREs products should be carefully evaluated before using the products for any application. Statistical and hydrological modelling are two common methods for evaluating SREs. The statistical evaluation method examines the intrinsic precipitation data quality including its spatiotemporal characteristics via pairwise comparison of the SREs products and ground observations. Scale mismatches between SREs products and ground-based measurements is a typical drawback. The hydrological modelling method evaluates the performance of a SREs product for a specific application such as streamflow predictive ability at watershed scale (Su et al., 2017). The two methods complement each other where the statistical method provides information on data quality while the hydrological model technique assesses the usefulness of the data for hydrological applications (Thiemig et al., 2013). However, most studies used only statistical evaluation methods (e.g., Dinku et al., 2018; Ayehu et al., 2018).

Studies have recommended SREs products for data scarce basins (Behrangi et al., 2011; Bitew and Gebremichael, 2011; Thiemig et al., 2013). However, there is no consensus regarding “best” SREs product for different climatic regions. Nesbitt et al. (2008) found that CMORPH and PERSIANN produced higher rainfall rates compared to TRMM for the mountain ranges of Mexico. Dinku et al. (2008) reported better performance of the TRMM and CMORPH products in Ethiopia and Zimbabwe whereas PERSSINN outperformed TRMM in South America according to de Goncalves et al. (2006). Interestingly, the performance of SREs products seems to differ even within a basin. For the Blue Nile basin in Ethiopia, for example, CMORPH overestimated precipitation for the lowland areas but underestimated for the highlands (Bitew and Gebremichael, 2011; Habib et al., 2012; Gebremichael et al., 2014). The discrepancy in the findings of these studies shows the performance of SREs varies with region, topography, season, and climatic conditions of the study area (Kidd and Huffman, 2011; Seyyedi et al., 2015; Nguyen et al., 2018; Dinku et al., 2018). As such, many studies have recommended SREs evaluation at a local scale to verify its performance for specific applications (Hu et al., 2014; Toté et al., 2015; Kimani et al., 2017; Ayehu et al., 2018).

Studies have examined the performance of SREs in Ethiopia (Haile et al., 2013; Worqlul et al., 2014; Ayehu et al., 2018; Dinku et al., 2018). However, majority of these studies used the statistical method to evaluate SREs, and no study has been completed for the
Dhidhessa River Basin (DRB). With only 0.32 rain gauges per 1000 km$^2$, the DRB meets the World Meteorological Organization (WMO) data-scarce basin classification (WMO, 1994). The objective of this study was to evaluate the performance of various SREs products in terms of characterizing the spatiotemporal distribution of rainfall in the DRB. The study could assist with the planning and management of existing and planned water resources projects in the basin.

SREs are continuously updated to minimize bias and uncertainty. Evaluating and validating improved products for various climatic regions would be valuable (Kimani et al., 2017). Recently improved SREs products include Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis version 7 (hereafter referred to as 3B43 for monthly and 3B42 for daily products), Climate Hazards Group InfraRed Precipitation with Stations version 2 (CHIRPS2), Tropical Applications of Meteorology using SATellite version 3 (TAMSAT3) and Integrated Multi-satellitE Retrievals for GPM version 6B (IMERG6).

Studies have reported improvements of these new versions compared to their predecessors. However, to the best of our knowledge, the rainfall detection and hydrological simulation capability of these SREs datasets were not evaluated for the basins in Ethiopian including the DRB. This study examined the latest SREs products in terms of their rainfall detection and estimation skills, and improving hydrological prediction for DRB, a medium-sized river basin with scarce gauging data. As such, the objectives of this study were: 1) to evaluate the intrinsic rainfall data quality and detection skills of multiple SREs products (i.e., 3B42/3, CHIRPS2, TAMSAT3, and IMERG6), and 2) to examine hydrologic prediction performances of SREs for the DRB. The Soil and Water Assessment Tool (SWAT), a physically based distributed model that has performed well in humid tropical regions like Ethiopia, was used for the hydrologic simulation.

2. Methods and Materials

2.1. Descriptions of the study area

The Dhidhessa River drains to the Blue Nile River (Figure 1). It is one of the largest and most important River basins in Ethiopia in terms of its physiography and hydrology (Yohannes, 2008). Located between 7°42′43″N to 10°2′55″N latitude and 35°31′23″E to 37°7′60″E longitude, the River basin exhibits highly variable topography that ranges from 619 m to 3213 m above mean sea level (a.m.s.l). The Dhidhessa River starts from the Sigmo
mountain ranges and travels 494 km before it joins the Blue Nile River around the Wanbara and Yaso districts. The outlet considered for this study is the confluence of the Dhidhessa River and the Blue Nile River which covers a total drainage area of 28,175 km². The River basin has many perennial tributaries (Figure 1).

Temperature and precipitation in the Dhidhessa River basin exhibit substantial spatial and seasonal variability. The mean maximum and minimum daily air temperatures in the River basin range from 20-33 °C and 6-19 °C, respectively. The long-term mean annual rainfall ranges from 1200 mm to 2200 mm in the River basin. Soils in the River basin are generally deep and have high organic content implying they have high infiltration potential. The dominant soil type is Acrisols while Cambisols and Nitisols are common (OWWDSE, 2014). Igneous, sedimentary and metamorphic rocks are but igneous rock, particularly basalt, is dominant in the basin (GSE, 2000). Forest, shrubland, grassland, and agriculture are the dominant land cover types in the basin (Kabite et al., 2020). Major crops include perennial and cash crops like coffee, Mango, and Avocado (OWWDSE, 2014).

Figure 1. Location map of Dhidhessa River basin with ground stations (USGS, 1998).
2.2. Data sources and descriptions

For this study, we used different spatial and temporal datasets such as Digital Elevation Model (DEM), climate, streamflow, soil and land cover from different sources (Table 1).

Table 1. Data description and sources.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Data acquisition</th>
<th>Resolution</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRTM DEM</td>
<td>1998</td>
<td>30 * 30 m</td>
<td>USGS</td>
</tr>
<tr>
<td>3B42/3</td>
<td>1998-recent present</td>
<td>0.25 (~25 km)</td>
<td>NASA &amp; JAXA</td>
</tr>
<tr>
<td>CHIRPS2</td>
<td>1980-present</td>
<td>0.05’ (~5 km)</td>
<td>USGS &amp; Climate Hazard Group</td>
</tr>
<tr>
<td>TAMSAT3</td>
<td>1980-present</td>
<td>0.0375 (~4 km)</td>
<td>Reading University</td>
</tr>
<tr>
<td>IMERG6</td>
<td>2001-present</td>
<td>0.1 (~10 km)</td>
<td>NASA &amp; JAXA</td>
</tr>
<tr>
<td>Streamflow data</td>
<td>1980-2014</td>
<td>Daily</td>
<td>EMoWI</td>
</tr>
<tr>
<td>Meteorological data</td>
<td>1980-014</td>
<td>Daily</td>
<td>NMA</td>
</tr>
<tr>
<td>Land cover</td>
<td>2001</td>
<td>30*30 m</td>
<td>Kabite et al. (2020)</td>
</tr>
<tr>
<td>Soil map</td>
<td>2013/14</td>
<td>variable</td>
<td>EMoWI, FAO &amp; OWWDSE</td>
</tr>
</tbody>
</table>

Shuttle Radar Thematic Mapper (SRTM) derived Digital Elevation Model (DEM) of 30*30 m spatial resolution was obtained from the United States Geological Survey (USGS). It is one of the input data for SWAT model from which topographic and drainage parameters (e.g., drainage pattern, slope and watershed boundary) were derived. Soil map of 1:250,000 scale were obtained from source described in Table 1. Soil physical properties required for SWAT model were derived from the soil map. Supervised image classification was used to prepare land cover map of 2001. Together with land cover and soil maps, DEM was used to create Hydrologic Response Units (HRUs).

Rainfall data for nine stations within the River basin and for three nearby stations (Figure 1), from 2001 to 2014 were obtained from the National Meteorological Agency (NMA) of Ethiopia. The rainfall data was used to evaluate the SREs using the statistical and hydrological modelling evaluation methods. In addition, Enhanced National Climate Time-series Service (ENACTS) gridded (4 m *4 m) minimum and maximum air temperature data was obtained from the National Meteorological Agency (NMA) of Ethiopia. Daily streamflow data from 2001 to 2014 was obtained for a station near the town of Arjo (Figure 1) from Ethiopian Ministry of Water, Irrigation and Energy (EMoWI).
The hydrometeorological stations used for this study were selected due to their long-term records and better data quality. The observed streamflow was used to calibrate and validate SWAT model. Land use map for 2001 and soil map were obtained from Kabite et al. (2020) and Ethiopian Ministry of Water, Irrigation and Energy (EMoWI), respectively.

**2.2.1. Satellite rainfall products**

The Satellite Rainfall Estimates (SREs) considered in this study include 3B42/3, TAMSAT3, CHIRPS2 and IMERG6. These datasets were selected because of several reasons including that they: i) have relatively high spatial resolution, ii) are gauge-adjusted products, iii) are the latest products and have been found to perform well by recent studies, and iv) were not compared for the basins in Ethiopia particularly IMERG6.

The TMPA provides rainfall products for area covering 50°N-50°S for the period of 1998 to present at 0.25°*0.25° and 3h spatial and temporal resolution, respectively. The 3h rainfall product is aggregated to daily (3B42) and monthly (3B43) gauge-adjusted post real time precipitation. The performance of the 3B42v7 is superior compared to its predecessor (i.e., 3B42v6) and the real time TMPA product (3B42RT) (Yong et al., 2014). The 3B43 was used in this study for the statistical evaluation while the 3B42 was used for the hydrological performance evaluation. The detail description is given by Huffman et al. (2007).

TAMSAT3 algorithm estimates precipitation in an indirect method using cloud-index method, which compares the cold cloud duration (CCD) with predetermined temperature threshold. The CCD is the length of time that a satellite pixel is colder than a given temperature threshold. The algorithm calibrates the CCD using parameters that vary seasonally and spatially but constant from year to year. This makes interannual variations in rainfall to depend only on the satellite observation. The dataset covers the whole Africa at ~4 km and 5-day (pentadal) resolutions for the period of 1983 to present. The original 5-day temporal resolution is disaggregated to daily time-step using daily CCD from which monthly data are derived. TAMSAT3 algorithm are improved compared to its processor (i.e., TAMSAT2). The detail is described in Maidment et al. (2017).

The Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) is a quasi-global precipitation product with ~5km (0.05°) spatial resolution and is available at daily, pentadal (5-day) and monthly timescales. The CHIRPS precipitation data is available from
1981 to present. It is gauge-adjusted dataset, which is calculated using weighted bias ratios rather than using absolute station values, which minimizes the heterogeneity of the dataset (Din
dku et al., 2018). The latest version of CHIRPS that uses more station data (i.e., CHIRPS version 2) was used in this study. Detail description of CHIRIPS2 is given in Funk et al. (2015).

The Global Precipitation Measurement (GPM) is the successor of TMPA with better rainfall detection capability. GPM provides precipitation measurements at 0.1° and half-hourly spatial and temporal resolution. Integrated Multi-satellite Retrievals for GPM (IMERG) is one of the GPM estimated from all constellation microwave sensors, IR-based observations from
geosynchronous satellites, and monthly gauge precipitation data. The IMERG products includes Early Run (near real-time with a latency of 6h), Late Run (reprocessed near real-time with a latency of 18 h) and Final Run (gauge-adjusted with a latency of four months). The IMERG Final Run product provides more accurate precipitation information compared to the near-real time products as it is gauge-adjusted. The latest release of GPM IMERG Final Run version 6B (IMERG6) was used for this study. The detail is given by Huffman et al. (2014).

In this study, the performances of 3B42/3, TAMSAT3, CHIRPS2 and IMERG6 rainfall products were evaluated statistically and hydrologically. All the SREs considered in this study are gauge-corrected, and thus bias correction may not be required. Thus, rain gauge stations that were used for validating any of the SREs datasets were excluded for fair comparison. Table 1 summarizes details of the data types used for this study.

2.3. Methodology

Satellite rainfall estimates offer several advantages compared to the conventional methods but can also be prone to multiple errors. Rainfall detection capability of the products can be affected by local climate and topography (Xue et al., 2013; Meng et al., 2014). Therefore, performance of SREs should be examined for a particular area before using the products for any application (Hu et al., 2014; Toté et al., 2015; Kimani et al., 2017).

The two common SREs performance evaluation methods are statistical (i.e., ground-truthing) and hydrological modelling performance (Behrangi et al., 2011; Bitew and Gebremichael, 2011; Thiemig et al., 2013, Abera et al., 2016; Jiang et al., 2017), and were used in this study. The methods complement each other and their combined application is recommended for more reliable SREs evaluation. The statistical evaluation method involves
pairwise comparison of SREs and the rain gauge and/or radar precipitation products. The method provides insight into the intrinsic data quality whereas the modelling approach assesses the usefulness of the data for a desired application (Thiemig et al., 2013). Statistical evaluation was performed for all the SREs products considered in this study (i.e., 3B43, CHIRPS2, TAMSAT3 and IMERG6) to examine their rainfall detection skills. Numerical and categorical validation indices were used to evaluate performance of the products. In addition, the SREs product and gauge data sets were independently used as forcing to calibrate and verify SWAT model. Accordingly, hydrological prediction performance of the rainfall products was evaluated graphically and using statistical indices.

2.3.1. Statistical evaluation of satellite rainfall estimates

Statistical SREs evaluation method was conducted at monthly, seasonal and annual timescales for the overlapping period of all the rainfall data sources (i.e., 2001-2014). A daily comparison was excluded from this study due to weak performance reported in previous studies (Ayehu et al., 2018; Zhao et al., 2017; Li et al., 2018). This is attributed to the measurement time mismatch between ground and satellite rainfall products.

Two approaches are commonly used for the statistical evaluation method. The first approach is pixel-to-pixel pair-wise comparisons of the spatially interpolated gauge-based and satellite-based data. The second approach is point-to-pixel pair-wise comparison where satellite rainfall estimates are extracted for each gauge location and the satellite-gauge data pairs are generated and compared. The second approach was used for this study. This is because the 12 rainfall stations considered in this study are unevenly distributed through the basin to accurately represent spatial variability of rainfall in the DRB as required for the first approach. As a result, we chose to extract gauge-satellite rainfall pair values at each rain gauge location instead of interpolating the gauge measurements into gridded products.

Accordingly, 168 and 2016 pair data points were extracted for annual and monthly analysis, respectively, and were evaluated using numerical validation indices such as Pearson correlation coefficient ($r$), bias ratio (BIAS), Nash-Sutcliffe efficiency ($E$) and Root Mean Square Error ($RMSE$). The Pearson correlation coefficient ($r$) evaluates how well the estimates correspond to the observed values; BIAS reflects how the satellite rainfall estimate over- or under-estimate the rain gauge observations; $E$ shows how well the estimate predicted the
observed time series. On the other hand, RMSE measures the average magnitude of the estimate errors. The summary of performance indices are presented in Table 2.

<table>
<thead>
<tr>
<th>Indices</th>
<th>Mathematical expression</th>
<th>Description</th>
<th>Perfect score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation</td>
<td>( r = \frac{\sum (R_g - \bar{R}) (R_s - \bar{R}_s)}{\sqrt{\sum (R_g - \bar{R})^2 \sum (R_s - \bar{R}_s)^2}} )</td>
<td>( R_g ) is gauge rainfall observation; ( R_s ) satellite rainfall estimates; ( \bar{R}_s ) is average satellite rainfall observation; ( \bar{R}_g ) is average gauge rainfall estimates</td>
<td>1</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>( RMSE = \frac{\sum (R_s - \bar{R}_s)^2}{n} )</td>
<td>( n ) is the number of data pairs; the value ranges from 0 to ( \infty )</td>
<td>0</td>
</tr>
<tr>
<td>Bias ratio (BIAS)</td>
<td>( BIAS = \frac{\sum R_s}{\sum \bar{R}_s} )</td>
<td>A value above (below) 1 indicates an aggregate satellite overestimation (underestimation) of the ground precipitation amounts.</td>
<td>1</td>
</tr>
<tr>
<td>Relative bias (RB)</td>
<td>( RB = \frac{\sum (R_s - \bar{R}_s)}{\sum \bar{R}_s} \times 100 )</td>
<td>Describes the systematic bias of the SREs; positive values indicate overestimation while negative values indicate underestimation of precipitation amounts.</td>
<td>0</td>
</tr>
<tr>
<td>Mean Error (ME)</td>
<td>( ME = \frac{1}{n} \sum_{i=1}^{n} (R_s - \bar{R}_s) )</td>
<td>Describes the average errors of the SREs relative to the observed rainfall data.</td>
<td>0</td>
</tr>
<tr>
<td>Nash-Sutcliffe of efficiency coefficient (E)</td>
<td>( E = 1 - \frac{\sum (R_s - \bar{R}_s)^2}{\sum (R_s - \bar{R}_g)^2} )</td>
<td>The value ranges from (-\infty) to 1; (0 \leq E \leq 1) acceptable while (E \leq 0) is unacceptable</td>
<td>1</td>
</tr>
<tr>
<td>Probability of Detection</td>
<td>( POD = H/(H + M) )</td>
<td>( H ) is the number of hits; ( M ) is the number of misses</td>
<td>1</td>
</tr>
<tr>
<td>False alarm ratio</td>
<td>( FAR = F/(H + F) )</td>
<td>( F ) is the number of false alarms</td>
<td>0</td>
</tr>
<tr>
<td>Critical success index</td>
<td>( CSI = H/(H + M + F) )</td>
<td>Describe the overall skill of the satellite products relative to gauge observation.</td>
<td>1</td>
</tr>
<tr>
<td>Percent bias (%)</td>
<td>( PBIAS = \frac{\sum (Q_o - Q_s)}{\sum Q_o} \times 100 )</td>
<td>( Q_o ) is observed discharge; ( Q_s ) is simulated discharge for the available pairs of data where &lt; ( \pm 15% ) is very good</td>
<td>0</td>
</tr>
<tr>
<td>Coefficient of determination ( r^2 )</td>
<td>( R^2 = \frac{\sum_{i=1}^{n} (O_i - \bar{O})(S_i - \bar{S})^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2 \sum_{i=1}^{n} (S_i - \bar{S})^2} )</td>
<td>( O_i ) &amp; ( \bar{O} ) is observed &amp; average streamflow, respectively; ( S_i ) &amp; ( \bar{S} ) is simulated and average, respectively.</td>
<td>1</td>
</tr>
<tr>
<td>Nash-Sutcliffe coefficient of efficiency</td>
<td>( NSE = \frac{\sum (Q_o - \bar{Q}_o)^2 - \sum (Q_o - \bar{Q}_o))^2}{\sum (Q_o - \bar{Q}_o)^2} )</td>
<td>( \bar{Q}_o ) is mean value of the observed discharge for the entire time under consideration</td>
<td>1</td>
</tr>
</tbody>
</table>
In addition, categorical validation indices such as probability of detection (POD), false alarm ratio (FAR) and critical success index (CSI) were also used for this study. The POD score represents the fraction of gauge observations detected correctly by the satellite while the FAR shows portion of events identified by the satellite but not confirmed by gauge observations. The CSI combines different aspects of the POD and FAR, describing the overall skill of the satellite products in estimating rainfall.

In general, SREs with $r>0.7$ and relative bias (RB) within 10% can be considered as reliable precipitation measurement sources (Brown, 2006; Condom et al., 2011). However, attention should be given to certain indices depending on the application of the product (Toté et al., 2015). For flood forecasting purpose, for example, underestimation of rainfall should be avoided (i.e., mean error ($ME$)>0 and high POD are desirable). In contrast, for drought monitoring, overestimation must be avoided (i.e., $ME<0$ and low FAR is preferred) (Dembélé and Zwart, 2016).

### 2.3.2. SWAT model setup

Soil and Water Assessment Tool (SWAT) is a semi-distributed, deterministic and continuous simulation watershed model that simulates many water quality and quality fluxes (Arnold et al., 2012). It is a physically based and computationally efficient model that has been widely used for various hydrological and/or environmental application in different regions of the world (Gassman et al., 2014). Furthermore, the capability of SWAT model to be easily linked with calibration, sensitivity analysis and uncertainty analysis tools (e.g., SWAT-CUP) made it more preferable.

SWAT model follows a two-level discretization scheme: i) sub-basin creation based on topographic data and ii) Hydrological Response Unit (HRU) creation by further discretizing the sub-basin based on land use and soil type. HRU is a basic computational unit assumed to be homogeneous in hydrologic response. Hydrological processes are first simulated at the HRU level and then routed at the sub-basin level (Neitsch et al., 2009). The SWAT model estimates surface runoff using the modified USDA Soil Conservation Service (SCS) curve number method. In this study, a minimum threshold area of 400 km$^2$ were used for determining the number of sub-basins and 5% threshold for the soil, slope, and land use were used for the HRU definition. Accordingly, 13 sub basins and 350 HRUs are created for the Arjo gauging station as outlet.
2.3.3. SWAT model calibration and validation

Hydrologic modelling performance evaluation technique is commonly performed by either calibrating the hydrologic model with gauge rainfall data and then validating with SREs, (i.e., static parameters) or calibrating and validating the model independently with each rainfall products (i.e., dynamic parameters) and then compare accuracies of the streamflows predicted using the capacity of the rainfall products. The latter is preferred for watersheds such as the DRB where gauging stations are sparse and unevenly distributed. Moreover, studies have reported that independently calibrating the hydrologic model with SREs and gauge data improves performance of the hydrological model (Zeweldi et al., 2011; Vernimmen et al., 2012; Lakew et al., 2017).

Calibration, validation and sensitivity analysis of SWAT was done using the SWAT-CUP software. The Sequential uncertainty fitting (SUFI-2) implemented in SWAT-CUP was used in this study (Abbaspour et al., 2007). SUFI-2 provides more reasonable and balanced predictions than the generalized likelihood uncertainty estimation (GLUE) and the parameter solution (ParaSol) methods (Zhou et al., 2014; Wu and Chen et al., 2019) offered by the tool. It also estimates parameter uncertainty attributed to input data, and model parameter and structure as total uncertainty (Abbaspour, 2015). The total uncertainty in the model prediction is commonly measured by P-factor and R-factor. P-factor represents the percentage of observed data enveloped by the 95 percent prediction uncertainty (95PPU) simulated by the model. The R-factor represents the ratio of the average width of the 95PPU band to the standard deviation of observed data. For realistic model prediction, $P$-factor $\geq 0.7$ and $R$-factor $\leq 1.5$ is desirable (Abbaspour et al., 2007, Arnold et al., 2012).

The first steps in SWAT model calibration and validation process is determining the most sensitive parameters for a given watershed. For this study, 19 parameters were identified based on the recommendations of previous studies (Roth et al., 2018; Lemann et al., 2019). Global sensitivity analysis was performed on the 19 parameters from which 11 parameters were found sensitive for the DRB, and were used for calibration, verification, and uncertainty analysis. The hydrologic simulations were performed for the 2001 to 2014 period. Two years of spin-up (warm-up) period (i.e., 2001 and 2002), and 6 years of calibration period (2003 to 2008), and 6 years of verification periods (2009 to 2014) were used. Graphical and statistical measures were used to evaluate prediction capability of the rainfall datasets. Accordingly, the
performance of model forced by each rainfall datasets was tested using the most widely used statistical indices (i.e., $R^2$, NSE and PBIAS), in addition to the $P$-factor and $R$-factor.

3. Results

3.1. Statistical evaluation

Figure 2 compares mean annual spatial rainfall distributions of the DRB. Average annual rainfall of the study area for the 2001 to 2014 period was 1682.09 mm/year (1150 to 2127 mm/year), 1698.59 mm/year (1432 to 1837 mm/year), 1699.06 mm/year (1092 to 2414 mm/year) and 1680.28 mm/year (1342 to 1721 mm/year) according to the CHIRPS2, IMERG6, TAMSAT3 and 3B43 products, respectively. For reference, mean annual rainfall for the DRB is 1650 mm/year based on the rain gauge data, which is within 1.8% to 3% of the estimates provided by the products. However, total annual rainfall range estimates were substantially different among the products. The decreasing rainfall trend from the southern (highlands) to the northern (lowlands) part of the basin were captured by all products. In particular, TAMSAT3 and CHIRPS2 captured the rainfall variability in better detail, perhaps due to their high spatial resolution. On the other hand, resolution of the 3B43 rainfall product seems too coarse to satisfactorily represent spatial variability of rainfall in the basin.

Figure 2. Spatial mean annual rainfall distribution of the four SREs for DRB (2001 to 2014)

Figures 3 to 5 show results of numerical statistical evaluation indices calculated from rainfall from the rain gauges and from the SREs products. More specifically, Figures 3 and 4
show correlation coefficients for the annual and monthly timescales, respectively. The results show that all four SREs products produced rainfall that correlate better to the ground based rainfall observations at monthly timescale than at annual time scales. The values of statistical evaluation indices for all products are summarized in Table 3. The results show that the CHIRPS2 performed better for the DRB with relatively higher $r$ and $E$, and lower $BIAS$, $ME$ and $RMSE$ for annual and monthly timescales, respectively.

Table 3. Statistical evaluation indices of all SREs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHIRPS2</td>
<td>0.78</td>
<td>0.92</td>
<td>1.01</td>
<td>1.01</td>
<td>25.94</td>
<td>2.70</td>
<td>214.36</td>
<td>50.48</td>
<td>0.51</td>
<td>0.84</td>
</tr>
<tr>
<td>3B43</td>
<td>0.48</td>
<td>0.87</td>
<td>1.02</td>
<td>1.02</td>
<td>30.58</td>
<td>2.55</td>
<td>306.34</td>
<td>62.05</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>IMERG6</td>
<td>0.52</td>
<td>0.90</td>
<td>1.03</td>
<td>1.03</td>
<td>48.87</td>
<td>4.07</td>
<td>299.55</td>
<td>56.95</td>
<td>0.39</td>
<td>0.80</td>
</tr>
<tr>
<td>TAMSAT3</td>
<td>0.62</td>
<td>0.89</td>
<td>1.03</td>
<td>1.03</td>
<td>51.46</td>
<td>2.67</td>
<td>274.00</td>
<td>61.28</td>
<td>0.77</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Figure 3. Correlation coefficient of the four SREs at annual timescale over DRB.

Figures 3 to 5 and Table 3 show that generally, CHIRPS2 performed better than the other three products for the DRB. Correlation coefficients for both monthly and annual timescales as well as all the indices presented in Figure 5 favor CHIRPS2 indicating its superior performance. Relative performance of the other three SREs is inconsistent as it varies with the goodness-of-fit criteria. The 3B43 product, for example, performed worse based on Figure 3.
and 4 (i.e., correlation coefficients for annual and monthly timescales) and RMSE and $E$ (Figure 5), but performed better than the other two SREs based on $BIAS$ and $ME$.

Figure 4. Monthly correlation coefficient of the four SREs for the DRB.

Figure 5. Statistical indices of the four SREs for DRB at annual and monthly time scales.
Categorical analysis result (Figure 6) shows that all the SREs considered in this study have high rainfall detection capability for the DRB. The POD and CSI values are close to 1 for all products, and FAR values are near 0, which shows that the SREs products have good rainfall event detection and estimation skills. However, TAMSAT3 exhibited relatively less rainfall detection skill, which could be attributed to the sensitivity of TAMSAT3 to topographic effects.

![Figure 6. Categorical indices of the four SREs for the DRB.](https://doi.org/10.5194/amt-2020-355)

Figure 7 shows seasonal SREs performance evaluation results. The Figure generally shows that performance of the SREs varied from season to season and among the rainfall products. For example, CHIRPS2 is superior in detecting and estimating rainfall events for the DRB for all months (seasons). The rainfall detection and estimating capability of CHIRPS2 is better for rainy season compared to the dry season. Likewise, the rainfall detection capability of TAMSAT3 is stronger for the rainy season (May to November) but weaker for the dry season (December to April). Compared to the other SREs products, TAMSAT3 generally poorly correlated for all months (seasons), and its BIAS was the highest for rainy season but the lowest for the dry season.
3.2. Hydrological modelling performance evaluation

The centroid of each sub basins were used as gauging locations, and used for extracting rainfall for all the SREs rainfall datasets. Thus, each sub basins are represented by a separate and dense gauges unlike that of the measured rainfall representation. The performance of the rainfall products were evaluated using SWAT-CUP at monthly time steps.

Table 4 shows details of the calibrated parameters including their ranges, best fit values, sensitivity ranks when different rainfall datasets are used as inputs for the DRB. The table shows that ranges and the best fit values vary from rainfall data source to another. This indicates that accurate characterization of rainfall variability is very critical for reliable hydrological predictions. This finding is consistent with studies that reported that different precipitation datasets influence model performance, parameter estimation and uncertainty in streamflow.
predictions (Sirisena et al., 2018; Goshime et al., 2019). Relative sensitivity of the parameters also varied between the rainfall datasets. In general, $GWQMN.gw$, $ALPHA.BF.gw$, $GW.DELAY.gw$, $RCHRG_DP.gw$, and $CN2.mgt$ are top five sensitive parameters. This seems to indicate that groundwater processes dominate streamflow in the DRB. This could be attributed to the dominantly deep and permeable soil, vegetated land surface and dominant tertiary basaltic rocks in the DRB (Conway, 2000; Kabite and Gessesse, 2018).

Table 4. Initial parameter ranges, fit values, and sensitivity ranks for rainfall data sources.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial values</th>
<th>Gauge</th>
<th>CHIRPS2</th>
<th>IMERG6</th>
<th>3B42</th>
<th>TAMSAT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{GWQMN.gw}$</td>
<td>0 to 5000</td>
<td>4936.02</td>
<td>1</td>
<td>201.64</td>
<td>3</td>
<td>3379.76</td>
</tr>
<tr>
<td>$v_{ALPHA.BF.gw}$</td>
<td>0 to 1</td>
<td>0.00</td>
<td>2</td>
<td>0.45</td>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>$v_{GW.DELAY.gw}$</td>
<td>0 to 500</td>
<td>339.10</td>
<td>3</td>
<td>29.02</td>
<td>5</td>
<td>34.76</td>
</tr>
<tr>
<td>$v_{RCHRG_DP.gw}$</td>
<td>0 to 1</td>
<td>0.02</td>
<td>4</td>
<td>0.44</td>
<td>7</td>
<td>0.04</td>
</tr>
<tr>
<td>$r_{CN2.mgt}$</td>
<td>-0.25 to 0</td>
<td>310.12</td>
<td>5</td>
<td>-0.25</td>
<td>11</td>
<td>-0.17</td>
</tr>
<tr>
<td>$r_{SOL.K.sol}$</td>
<td>0 to 2000</td>
<td>260.96</td>
<td>6</td>
<td>1086.63</td>
<td>9</td>
<td>391.90</td>
</tr>
<tr>
<td>$v_{CH.N2.rte}$</td>
<td>-0.01 to 0.3</td>
<td>0.74</td>
<td>7</td>
<td>0.02</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>$CH.K2.rte$</td>
<td>-0.01 to 500</td>
<td>310.12</td>
<td>8</td>
<td>354.51</td>
<td>2</td>
<td>426.08</td>
</tr>
<tr>
<td>$v_{GW.REVAP.gw}$</td>
<td>0.02 to 0.2</td>
<td>0.00</td>
<td>9</td>
<td>0.15</td>
<td>8</td>
<td>0.20</td>
</tr>
<tr>
<td>$r_{SOL.AWC.sol}$</td>
<td>-0.5 to 0.5</td>
<td>-0.01</td>
<td>10</td>
<td>-0.49</td>
<td>6</td>
<td>-0.19</td>
</tr>
<tr>
<td>$v_{REVAPMNN.gw}$</td>
<td>0 to 500</td>
<td>170.26</td>
<td>11</td>
<td>14.52</td>
<td>10</td>
<td>381.84</td>
</tr>
</tbody>
</table>

Figure 8 compares the observed and the predicted streamflows for the calibration (2003 to 2008) and verification (2009 to 2014) periods for all five rainfall datasets. Goodness of the streamflow predictions is also summarized in Table 5. The result shows that streamflow is overestimated for all rainfall products, including the gauge rainfall. This could be due to the uncertainty of SREs for the extreme rainfall events at daily scale (Jiang et al., 2017). The overestimated streamflows could also be attributed to overestimation of rainfall by the SREs as described in the previous sections. Generally, the indices provided in Table 4 indicate that the streamflow predictions are satisfactory (Moriasi et al., 2017) for CHIRPS2, IMERG6, and the guaged rainfall but not for TAMSAT3 and 3B42.
Figure 8. Graphical calibration and validation of streamflow at monthly scale.

Table 5. Calibration and validation results for the different rainfall products.

<table>
<thead>
<tr>
<th>Rainfall products</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSE</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Gauge</td>
<td>0.55</td>
<td>0.54</td>
</tr>
<tr>
<td>CHIRPS2</td>
<td>0.69</td>
<td>0.7</td>
</tr>
<tr>
<td>IMERG6</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td>TAMSAT3</td>
<td>0.43</td>
<td>0.46</td>
</tr>
<tr>
<td>3B42</td>
<td>0.48</td>
<td>0.51</td>
</tr>
</tbody>
</table>
4. Discussion

The statistical SREs evaluation result showed that all the rainfall products captured the spatiotemporal rainfall variability of the DRB except the 3B43. Poor performance of 3B43 in capturing basin’s rainfall variability is in agreement with findings of two previous studies done for other basins in Ethiopia [Dinku et al., 2008; Worqlul et al., 2014]. The reason could be attributed to the fact that gauge adjustment for 3B43 product did not use adequate gauge data from Ethiopian highlands due to lack of data [Haile et al., 2013]. However, Gebremicael et al. (2019) reported better performance of 3B43 for the Tekeze-Atibara basin, which is located in the northern mountainous area of Ethiopia.

Better correlation of SREs with observed rainfall was observed at monthly than at annual timescales for all products. This is consistent with studies that reported the performance of SREs improved with increased time aggregation that peaks at monthly timescale (Dembélé and Zwart, 2016; Katsanos et al., 2016; Zhao et al., 2017; Ayehu et al., 2018; Li et al., 2018; Guermazi et al., 2019). The weak agreement of SREs with observed data at annual timescale shows that the SREs considered in this study generally did not capture the interannual rainfall variability. In this regards, particularly the 3B43 product failed to capture annual rainfall variability compared to the other three SREs. Overall, all four SREs products overestimated rainfall for the DRB by 10% for CHIRPS2 to 30% for IMERG6 and TAMSAT3 (Figure 5). This finding is consistent with studies that reported overestimation of IMERG6 and 3B43 products for the alpine and gorge regions of China (Chen et al., 2019). However, Gebremicael et al. (2019) reported underestimation of rainfall by CHIRPS2 for the Tekeze-Atbara basin, which is a mountainous and arid basin in northern Ethiopia. Ayehu et al. (2018) also reported slight underestimation of rainfall by CHIRPS2 for the upper Blue Nile Basin. The discrepancy between our finding and the previous studies done for the basins in Ethiopia may be due to differences in watershed characteristics.

Generally, this study showed that the SREs products considered in this study exhibited satisfactory rainfall detection and estimation capability for the DRB. The products are applicable for flood forecasting applications for the DRB (Toté et al., 2015). CHIRPS2 performed better than the other three SREs for annual, seasonal, and monthly timescales in detecting and estimating rainfall for the basin. The superiority of CHIRPS2 was also reported by previous studies for different parts of world (Katsanos et al., 2016; Dembélé and Zwart, 2016) including basins in Ethiopia (Bayissa et al., 2017; Ayehu et al., 2018; Dinku et al., 2018;
Gebremicael et al., 2019). For example, Dinku et al. (2018) reported better rainfall estimation capability of CHIRPS2 for East Africa compared to African Rainfall Climatology version 2 (ARC2) and TAMSAT3 products. Ayehu et al. (2018) reported better performance of CHIRPS2 for the Blue Nile Basin compared to ARC2 and TAMSAT3. Better performance of CHIRPS2 has been attributed to the capability of the algorithm to integrate satellite, gauge and reanalysis products and its high spatial and temporal resolution (Funk et al., 2015). On the contrary, generally, the 3B43 rainfall product performed poorly for the DRB for all timescales. This could be due to its course spatial resolution and lack of gauge-adjustment for highlands of Ethiopia (Haile et al., 2013). The IMERG6 showed better rainfall detection and estimation capability for the study area than the 3B43 product, which is consistent with findings of previous studies (Huffman et al., 2015; Zhang et al., 2018; Zhang et al., 2019). Better performance of IMERG6 is attributed to the inclusion of dual and high-frequency channels, which improve light and solid precipitation detection capability (Huffman et al., 2015).

Hydrologic simulation performance evaluation result of SREs showed that accurate characterization of rainfall variability is very critical for reliable hydrological predictions. This finding is consistent with studies that reported that different precipitation datasets influence model performance, parameter estimation and uncertainty in streamflow predictions (Sirisena et al., 2018; Goshime et al., 2019). Overestimation of streamflow for all SREs products could be attributed to uncertainty of SREs for extreme rainfall events at daily scale (Zhao et al., 2017). The overestimated streamflows could also be attributed to overestimation of rainfalls by the SREs as described in the previous sections.

Overall, this study showed that CHIRPS2 and IMERG6 predicted streamflow better than the guage rainfall and other two SREs products for the DRB. Superior hydrological performance of SREs products compared to gauge rainfall data were also reported by many other studies (Grusson et a., 2017; Bitew and Gebremichael, 2011; Goshime et al., 2019; Xian et al., 2019; Li et al., 2018; Belete et al., 2020). For example, Bitew and Gebremichael (2011) reported that satellite-based rainfall predicted streamflow better than gauge rainfall for complex high-elevation basin in Ethiopia. Likewise, a bias-corrected CHIRP rainfall dataset resulted in better streamflow prediction than a gauge rainfall dataset for Ziway watershed in Ethiopia (Goshime et al., 2019).

The relatively poor performance of gauge rainfall compared to the CHIRPS2 and IMERG6 shows that the existing rainfall gauges do not represent spatiotemporal variability of
rainfall in the DRB. The rain gauges are sparse, spatially uneven, and incomplete records for
the DRB. As previously mentioned, rain gauge density for the DRB is 0.32 per 1000 km²,
which is much lower than the World Meteorological Organization (WMO) recommendation
of one gauge per 100-250 km² for mountainous areas of tropical regions such as the DRB
(WMO, 1994). Overall, this rainfall products performed satisfactorily in terms of detecting
and estimating rainfall study showed that CHIRPS2 and IMERG6 rainfall products
performed satisfactorily in terms of detecting and estimating rainfall as well as predicting
streamflow for the DRB.

5. Conclusions

Satellite rainfall estimates are alternative rainfall data sources for hydrological and
climate studies for data scarce regions like Ethiopia. However, SREs contain uncertainties
attributed to errors in measurement, sampling, retrieval algorithm and bias correction
processes. Moreover, the accuracy of rainfall estimation algorithm is influenced by topography
and climatic conditions of a given area. Therefore, SREs products should be evaluated locally
before they are used for any application. In this study, we examined the intrinsic data quality
and hydrological simulation performance of CHIRPS2, IMERG6, 3B42/3 and TAMSAT3
rainfall datasets for the DRB. The statistical evaluation results generally revealed that all four
SREs products showed promising rainfall estimation and detection capability for the DRB.
Particularly, all SREs captured the south-north declining rainfall patterns of the study area.
This could be due to the fact that all the SREs products were gauge adjusted and that they are
the latest versions. However, all the SREs datasets overestimated rainfall for DRB. Correlation
coefficients of all SREs were strong for the monthly timescales than for the annual timescales,
which shows that all rainfall products failed to capture interannual rainfall variability.

The quantitative statistical indices showed that CHIRPS2 performed the best in
estimating and detecting rainfall events for the DRB at monthly as well as annual timescales.
This is likely due to the fact that CHIRPS2 was created by merging satellite, reanalysis and
gauge datasets at high spatial resolution. In the contrary, 3B43 performed poorly for the basin.

The hydrological modelling based performance evaluation showed that ranges, best fit
values, and relative sensitivities of SWAT’s calibration parameters varied with the rainfall
datasets. Overall, groundwater flow related parameters such as GWQMN.gw, ALPHA_BF.gw,
GW_DELAY.gw and RCHRG_DP.gw were found more sensitive for all rainfall products. This showed that subsurface processes dominate hydrologic response of the DRB.

The hydrological simulation performance results also showed that all the rainfall products, including the observed rainfall, overestimated streamflow especially the high flows, which could be attributed to the uncertainty of SREs rainfall to predict at shorter timescale (e.g., daily) and event rainfalls. The study showed CHIRPS2 and IMERG6 predicted streamflow for the basin satisfactorily, and even outperformed performance of the gauge rainfall. The relatively poor performance of the gauge rainfalls can be attributed to the fact that the gauges are too sparse to accurately characterize rainfall variability in the basin. Overall, CHIRS2 and IMERG6 products seem to perform better for the Dhidhessa River basin to detect rainfall events, to estimate rainfall quantity, and to improve streamflow predictions.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Supplementary Materials:** Provided upon request.

**Author contributions**
Gizachew Kabite: Conceptualization, Data collection, analysis and interpretation, writing—original draft preparation.
Misgana K. Muleta and Berhan Gessesse: Writing—review and editing. All authors have read and agreed to the published version of the manuscript:

**Conflicts of Interest**

The authors declare no conflict of interest.

**Acknowledgments**

We are grateful to the Ethiopian Space Science and Technology Institute for providing partial financial support for this research. We are also thankful to the developers of CHIRPS2, IMERG6, TAMSAT3 and 3B42 datasets and for providing the data free of charge. The National Meteorological Agency of Ethiopia and the Ethiopian Ministry of Water, Irrigation and Energy are also acknowledged for providing climate and streamflow data, respectively.
**References**


Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S.H.: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). In Algorithm Theoretical Basis Document (ATBD); NASA/GSFC: Greenbelt, MD, USA, 2014.


Nesbitt, S.W., Gochis, D.J., and Lang, T.J.: The diurnal cycle of clouds and precipitation along the Sierra Madre Occidental observed during NAME-2004: Implications for warm season precipitation estimation in complex terrain, J. Hydrometeor. 9, 728-743, 2008.


