Validation and Intercomparison of Satellite-Based Rainfall Products over Africa with TAHMO In Situ Rainfall Observations

DENIS MACHARIA,a,b KATIE FANKHAUSER,a JOHN S. SELKER,c JASON C. NEFF,d AND EVAN A. THOMASa

a Mortenson Center in Global Engineering, University of Colorado Boulder, Boulder, Colorado
b Regional Centre for Mapping of Resources for Development, Nairobi, Kenya
c Department of Biological and Ecological Engineering, Oregon State University, Corvallis, Oregon
d Sustainability Innovation Lab at Colorado, Department of Environmental Studies, University of Colorado Boulder, Boulder, Colorado

(Manuscript received 18 August 2021, in final form 18 April 2022)

ABSTRACT: Increasingly, satellite-derived rainfall data are used for climate research and action in Africa. In this study, we use 6 years of rain gauge data from 596 stations operated by the Trans-African Hydrometeorological Observatory (TAHMO) to validate three gauge-calibrated satellite rainfall products—CHIRPS, TAMSAT, and GSMaP— and one satellite-only rainfall product, GSMaP. Validations are stratified to evaluate performance across the continent and in East Africa, southern Africa, and West Africa at daily, pentadal, and monthly time scales. For daily mean rainfall over Africa, CHIRPS has the highest bias at 15.5% (0.5 mm) whereas GSMaP wGauge has the lowest bias at 0.02 mm (0.7%). We find higher daily rainfall event detection scores in the GSMaP products than in CHIRPS or TAMSAT. Generally, for every two rainfall events predicted by CHIRPS and TAMSAT, the GSMaP products predict three or more events. The highest mean monthly biases are produced by CHIRPS in East Africa (29%; wet bias of 26.3 mm), TAMSAT in southern Africa (13%; dry bias of 10.4 mm), and GSMaP in West Africa (23%; wet bias of 19.6 mm). Considerable biases in seasonal rainfall are observed in all subregions for every satellite product. There is an increase of 0.6–1.3 mm in satellite rainfall RMSE for a 1-km increase in elevation revealing the influence of elevation on rainfall estimation by satellite models. Overall, satellite-derived rainfall products have notable errors, while GSMaP products produce comparable or better results at multiple time scales relative to CHIRPS and TAMSAT.

KEYWORDS: Atmosphere; Africa; In situ atmospheric observations; Satellite observations; Error analysis; Statistics

1. Introduction

The livelihoods of millions of people in sub-Saharan Africa are dependent on rain-fed agriculture, which is vulnerable to extreme weather events and climate change. Gridded, multiyear, high-quality weather data are essential in understanding exposure to climate shocks and developing strategies to mitigate adverse impacts on livelihoods. Until recently, climate research and action in the region has relied on sparse, unreliable, and often inaccessible observations from weather stations which are simple to operate but expensive to maintain (Dinku 2019). Most in situ observations are managed by National Hydrometeorological Services (NHMS) which have data sharing regulations that restrict open access to available records. Moreover, the number of observations by NHMS has dramatically declined in recent years, thus compounding the climate monitoring challenge (van de Giesen et al. 2014).

a. Availability of climate data for research in Africa

The lack of adequate and reliable long-term in situ weather observations in Africa has led to an increased reliance on satellite-based products for hydrometeorological applications.

The advantages of satellite data are accessibility, consistency, and high temporal and spatial resolutions over large areas. Satellite data are available at time scales that are useful for trend analysis (Ayehu et al. 2018; Le Coz and van de Giesen 2020). However, the need for ground observations persists because satellite products require calibration and bias correction using in situ measurements. As a result, the importance of sources of in situ observations used to complement observations provided by NHMS has grown in the region. The unmet need for more accurate, reliable, and accessible weather data; advances in developing low-cost monitoring technologies; and increased satellite data availability have spurred this growth (Thomson et al. 2011). Citizen-science programs also contribute important weather data for climate research. Several of these programs are volunteer-led networks and examples include the Trans-African Hydrometeorological Observatory (TAHMO) (van de Giesen et al. 2014) and the Global Learning and Observations to benefit the Environment Program (GLOBE) (Butler and MacGregor 2003). Despite the introduction of new in situ data sources, the spatial coverage of ground stations remains inadequate for many hydrometeorological applications, making satellite-rainfall products (SRPs) a critical tool in weather monitoring.

b. Satellite data–driven hydrometeorological applications

A number of hydrometeorological applications in Africa rely on satellite rainfall data. These applications include drought-risk monitoring (Thomas et al. 2020; Funk et al. 2021a), runoff and streamflow modeling (McNally et al. 2017).
monitoring of pastoral water resources (Senay et al. 2013), trend analysis of seasonal rainfall (Lyon and DeWitt 2012; Liebmann et al. 2014), assessing extreme rainfall events (Ongoma et al. 2018; Harrison et al. 2019), and mapping the vulnerability of natural and social systems to climate change (de Sherbinin et al. 2019; Macharia et al. 2020). While they are widely used, SRPs have persistent, systematic errors across space and time (Derin et al. 2016) contributed by rainfall retrieval methods, sensor type, and sampling frequency (Prigent 2010). There are three main types of satellite sensor methods for rainfall estimation from space: Passive microwave sensors (PMW) from low-Earth-orbiting satellites, thermal infrared observations from geostationary satellites (Geo-IR), and active microwave sensor observations. Methods have also been developed that combine PMW, Geo-IR, and active microwave sensor measurements (Ayehu et al. 2018; Kidd et al. 2003; Di Tomaso et al. 2009; Huffman et al. 2010; Funk et al. 2021b). PMW-based observations rely on satellite sensors that can penetrate clouds and are considered a more direct measurement of rainfall because these sensors are sensitive to precipitation-sized particles in the atmosphere (Sun et al. 2018). Geo-IR sensors measure top of cloud brightness temperatures to extrapolate rainfall amounts (Ayehu et al. 2018; Young et al. 2014; Prigent 2010; Funk et al. 2021b). Active microwave sensors measure the size and number of rain or snow drops at multiple vertical layers in the cloud (Battaglia et al. 2020).

Gridded satellite rainfall products that are produced from multimethod approaches (Stampoulis and Anagnostou 2012), and rainfall products that combine PMW, Geo-IR, satellite radar, and ground rainfall measurements are now available in Africa (Funk et al. 2021b). A number of these products also use rain gauge measurements to correct systematic errors in satellite-based rainfall retrieval models (Le Coz and van de Giesen 2020). Gridded rainfall products that have high spatial and temporal resolution are available due to the merging of multisource satellite data and the integration of rain gauge measurements. Relevant integrated SRPs in Africa include Climate Hazards Infrared Precipitation with Stations (CHIRPS; Funk et al. 2015c), the Tropical Applications of Meteorology Using Satellite data (TAMSAT; Maidment et al. 2017), African Rainfall Climatology version 2 (ARC2; Novella and Thiau 2013), Global Precipitation Climatology Project Satellite and Gauge (GPCP-SG; Huffman et al. 1997; Huffman and Bolvin 2013), Rainfall Estimate version 2 (RFE2; Xie and Arkin 1996), Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG; Huffman et al. 2020), Multi-Source Weighted-Ensemble Precipitation (MSWEP; Beck et al. 2019), the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2010), and Climate Hazards Center IMERG with Stations (CHIMES; Funk et al. 2021b), among others. While the development of algorithms that combine multiple satellite and in situ gauge data has helped overcome some of the reliability and accuracy challenges of climate data in Africa (Funk et al. 2015b; Hijmans et al. 2005; Nerini et al. 2015), satellite-based rainfall products still contain errors that require continuous evaluation and corrections as improved rainfall retrieval models become available.

c. Performance evaluation of satellite-based rainfall products

Numerous performance evaluations have been conducted in Africa at different spatial scales with varying results. Most of these evaluations relied on rain gauge observations from national meteorological agencies which are declining in availability (van de Giesen et al. 2014; Dinku 2019). For regional-level evaluations, the reliance on gauge observations from national meteorological agencies may have significant differences in quality which contributes to errors in the evaluations. Conversely, TAHMO gauge observations are uniquely consistent across countries. The use of uniform rain gauge instruments and data quality control process promotes consistency across the region. Therefore, TAHMO data may be well suited to rainfall validation studies at regional and subregional scales.

Some examples of past evaluations include recent studies by Sagat et al. (2020), Dinku et al. (2018), Ayehu et al. (2018), Dezfuli et al. (2017), Toté et al. (2015), Kimani et al. (2017), Atiah et al. (2020a), Derin et al. (2016), Dembéle and Zwart (2016), Cattani et al. (2016), Maidment et al. (2013), Atiah et al. (2020b), Young et al. (2014), Abebe et al. (2020), Sahlou et al. (2017), Bayissa et al. (2017), Le Coz and van de Giesen (2020), Lu et al. (2020), Boluwe (2020), and Funk et al. (2021b). These evaluations concluded that the accuracy of SRPs is generally influenced by factors such as the local climate, topography, and rainfall seasonality. Others noted that errors in satellite rainfall estimation reduce with temporal aggregation (Dinku et al. 2018; Toté et al. 2015; Kimani et al. 2017; Ayehu et al. 2018; Atiah et al. 2020a). Evaluations done by Dezfuli et al. (2017) and Toté et al. (2015) highlighted the limited ability of SRPs to reproduce low- and high-intensity rainfall and diminished accuracy in coastal and mountainous regions. The most recent evaluation of CHIMES by Funk et al. (2021b) showed good accuracy at pentadal (5 days) and monthly time scales when compared with high-quality gridded station data, with potential for further improvement with additional gauge data.

The use of TAHMO gauge data in Africa for validation of satellite data is limited. Two studies that have used these data were small in scope in terms of the number of stations used, the time period covered, or both. Dezfuli et al. (2017) used data from three TAHMO stations in Kenya and Ghana covering the rainy seasons in 2015 to validate IMERG in Africa while Boluwe (2020) evaluated SRPs in Ghana and Uganda in 2018 using 77 and 40 stations, respectively. The authors in both studies indicated that their evaluations were inadequate due to the short period considered, and due to the fact that TAHMO was a relatively new network of gauge observations.

Our study expands on the work by Dezfuli et al. (2017) and Boluwe (2020) by validating four SRPs in Africa: CHIRPS, TAMSAT, the Global Satellite Mapping of Precipitation moving vector with Kalman filter (GSMaP), and the gauge-calibrated GSMaP version (GSMaP_wGauge). CHIRPS, TAMSAT and GSMaP_wGauge are all gauge-calibrated products whereas GSMaP is a satellite-only product. None of the gauge-calibrated products use observations from TAHMO in their rainfall calibration models, ensuring our validation study
is independent. CHIRPS and TAMSAT are based on precipitation estimates from quasi-global geostationary infrared satellites and both integrate precipitation climatologies and in situ rain gauge measurements for calibration. GSMaP and GSMaP_wGauge integrate diverse sources of rainfall estimates that include PMW, Geo-IR, satellite radar, and gauge data. CHIRPS and TAMSAT have been used widely for drought assessments in Africa. However, the GSMaP products are relatively new, their performance in Africa is relatively understudied, and their use is limited despite their potential usage for high temporal resolution applications. Further, with TAHMO, our analysis uses the highest density of gauge observations from a single source in the region to evaluate whether these satellite-rainfall products are appropriate hydrometeorological products for climate research in Africa.

2. Materials and methods

a. Study area

This cross-validation study was conducted over Africa and three subregions (Fig. 1). These subregions are broadly defined as East Africa (9°S–25°N, 20°–50°E), southern Africa (9°–35°S, 20°–55.5°E), and West Africa (20°W–20°E, 1°–20°S). The three subregions have diverse climate regimes which are influenced by mesoscale processes and complex topography. Rainfall in East Africa and southern Africa is strongly teleconnected to ocean conditions in the Pacific (Endris et al. 2013; Ogallo 1988; Williams and Funk 2011; Otieno and Anyah 2013). Giannini et al. (2003) revealed the influence of the Atlantic Ocean on rainfall patterns in West Africa. Rainfall in East Africa and southern Africa varies in space and time. The main factors influencing this variability are complex large-scale factors including topography, lakes, the Indian Ocean, and the seasonal dynamics of the intertropical convergence zone (ITCZ) (Dinku et al. 2018; Nicholson 2017). East Africa has the lowest and highest elevation points in Africa (Dinku et al. 2018). Topography has been shown to influence the flow of southwesterly monsoon winds and the Somali jet both of which influence rainfall in this region. Lake Victoria, conversely, creates mesoscale circulation systems which result in a nocturnal and an afternoon rainfall regime in the western and eastern half of the lake, respectively (Nicholson 2017). Rainfall distribution and intensity in East Africa and southern Africa varies from month to month following the
movement of the ITCZ. Equatorial regions (4°S–7°N) are dominated by boreal spring and autumn rainfall seasons otherwise known as the long rainy (March–May) and short rainy (October–December) seasons, respectively (Nicholson 2017). Much of the rainfall in the north (>7°N) is generally concentrated in the boreal summer (June–October) and in the boreal winter (November–April) in the southern African region (<4°S).

West Africa has an equally variable rainfall regime. The region can broadly be divided into three climatic zones: the Guinea coast (4°–8°N), the Savannah region (8°–11°N), and the Sahel (11°–16°N) (Abiodun et al. 2012). The Guinea coast has a subhumid climate. Annual rainfall in this zone ranges between 1250 and 1500 mm. The Savannah zone is a semi-arid zone with annual rainfall between 750 and 1250 mm. The Sahel zone is characterized by a short rainy season between June and September. This season is distinctively unimodal.

Sahel zone is characterized by a short rainy season between June and September. This season is distinctively unimodal with an average annual rainfall of about 750 mm. In addition to the mesoscale convective systems that influence rainfall in West Africa, topography plays an important role in its distribution (Akinsanola et al. 2017; Guilloteau et al. 2016).

b. Data sources

Station in situ rainfall data were from 596 quality-controlled gauge stations operated by TAHMO. TAHMO's goal is to improve hydrometeorological station networks across sub-Saharan Africa. The number of stations had increased from 57 in 2015 to 596 in 2020. The four SRPs used in this study have several properties that make them ideal for a number of hydrometeorological applications in Africa: (i) they have relatively high spatial (<10 km) and temporal (daily to monthly) resolution, (ii) they have a long time series record (>30 years for CHIRPS and TAMSAT, and ~20 years for GSMaP), and (iii) they are open source and widely available in the region.

1) IN SITU RAINFALL DATA

TAHMO's network of stations leverages current sensor and communication technology through a public-private partnership approach to increase the number of stations reporting meteorological information in real time. The stations are based on the ATMOS 41 model manufactured by the METER Group (https://www.metergroup.com/en/meter-environment/products/atmos-41-weather-station) and include sensors for rainfall, temperature, atmospheric pressure, relative humidity, solar radiation, wind speed, and wind direction. The rain gauge uses an electric drip counter, and several stations have redundant ECRN-100 tipping-bucket rainfall sensors for cross validation (ZeMicheal and Dietterich 2020). The rainfall data have a resolution of 0.0017 mm and an accuracy of ±5% of measurement for rainfall between 0 and 50 mm h⁻¹ (https://tahmo.org/technical-information). TAHMO stations have a 5-min measurement interval of meteorological variables (Dezfui et al. 2017).

The data are subjected to quality control processes that include cross calibration and comparison with nearby sensors, and qualitative comparison with satellite observations (van de Giesen et al. 2014). First, the datalogger on the instrument confirms whether the value it receives from the rainfall sensor is reasonable. If the result indicates potentially corrupted data, the system flags the data point as “erroneous.” Second, the range of rainfall data received from the stations is checked. A flag erroneous is applied if the value is below 0 mm or exceeding 100 mm in the 5-min interval. If the value exceeds 40 mm in the 5-min interval, then a “doubtful” flag is applied to the data for subsequent review. This threshold is based on empirical data and close to the specified maximum rainfall rate for the Atmos-41 drip counter.

Every week, a data quality control manager reviews an automated report consisting of the counts of flagged data and determines whether further action is warranted. In most cases, this step resolves many instances of overestimation of rainfall, but it may not be adequate to diagnose and resolve cases of underreporting. The quality control process also includes producing analytics at the daily level. This model was developed specifically for automated quality control of the TAHMO data and is part of a system known as SensorDX (ZeMicheal and Dietterich 2020). The SensorDX system uses novel nonparametric anomaly detection algorithms to analyze sensor data. Detected anomalies (e.g., high scores from broken sensors) from the SensorDX model are reported to TAHMO.

Data quality control is also supplemented by comparisons with satellite measurements, namely, comparing TAHMO data with CHIRPS and GSMaP estimates for monthly total rainfall. Station data that are less than 50% of the minimum or greater than 200% of the maximum totals estimated by CHIRPS and GSMaP are flagged for further scrutiny. TAHMO estimates that about 75% of detected anomalies are resolved by cleaning the stations. Users of TAHMO data are provided with quality-controlled data and the flagged records when they extract station data directly from the data portal.

We extracted TAHMO data at 5-min interval for all available stations and aggregated the data to daily, pentadal, and monthly time scales for the period 2015–20. TAHMO stations are located in Kenya, Uganda, Rwanda, Ethiopia, Tanzania, the Democratic Republic of Congo, Madagascar, Mozambique, Zambia, Zimbabwe, Lesotho, Malawi, South Africa, Ghana, Mali, Burkina Faso, Chad, Benin, Togo, Senegal, Nigeria, Cameroon, and Cote D’Ivoire. Regionally, there are 245 gauge stations in East Africa, 84 in southern Africa, and 267 in West Africa.

2) SATELLITE-BASED RAINFALL DATA

CHIRPS is a quasi-global (50°S–50°N) gauge-calibrated gridded precipitation product available at daily, pentadal, and monthly time scales. CHIRPS is developed from the satellite-only Climate Hazards Group Infrared Precipitation (CHIRP) product (Funk et al. 2015a) and the incorporation of gauge data from stations around the world. CHIRPS is produced by the University of California, Santa Barbara’s Climate Hazards Centre (CHC). Satellite-derived values are based on Geo-IR observations (Funk et al. 2021b) generated using local regression with the Tropical Rainfall Measuring Mission multisatel-
Briefly, the CHIRPS algorithm relies on (i) precipitation estimates from infrared cold cloud duration (CCD) observations; (ii) the Climate Hazards Group Precipitation Climatology (CHPclim; Funk et al. 2015b) that incorporates publicly available gauge data, satellite observations, elevation, longitude, and latitude; and (iii) blending gauge station data from several sources including the Global Telecommunication System (GTS), Global Summary of the Day (GSOD), Somalia Water and Land Information Management (SWALIM) project, regional and national meteorological services, and other volunteer networks to create a temporally consistent, gauge-calibrated dataset (CHIRPS) at 0.05° spatial resolution.

The CHIRPS bias correction procedure relies on correlations between CHIRP pixel precipitation and nearby stations. These correlations and bias ratios that are calculated from the five nearest stations to a CHIRP pixel are combined into a single bias correction factor by a weighted average technique. Correlation coefficients are squared to generate the weights. The bias-correction factors are multiplied by CHIRP values to create a corrected CHIRP value. The corrected and uncorrected CHIRP values are combined using the square of the correlation between CHIRP and “true” rainfall values and the ratio between the corrected and uncorrected values from the estimated correlation at the nearest station. A detailed description of the CHIRPS algorithm and data can be found in Funk et al. (2015c). CHIRPS data span the period from 1981 to present with a latency of roughly three weeks. We downloaded daily, pentadal and monthly version 2.0 products from the CHC site (https://data.chc.ucsb.edu/products/CHIRPS-2.0/) for the period overlapping the TAHOE gauge observations.

The TAMSAT rainfall product is produced at the University of Reading in the United Kingdom (Dinku et al. 2018). It is a gauge-calibrated rainfall product for Africa available at 0.0375° spatial resolution. It relies on methodology similar to that of CHIRPS and uses the CCD to derive cloud-top temperatures taken from Meteosat infrared radiances. TAMSAT uses gauge observations from publicly available GTS data and data provided by national meteorological services to develop a bias-corrected calibration climatology dataset. This climatology dataset is merged with additional gauge observations to produce daily, pentadal and monthly gridded precipitation data for the continent from 1983 to present.

The TAMSAT calibration process involves distinguishing rainy from nonrainy regions using daily CCD totals derived at a range of thresholds between –30° and -60°C. These CCD totals are summed to dekadal or pentadal time scales. Contingency tables are generated for every threshold and include comparisons between CCD which are greater than 0 at the pixel and true rainfall from collocated rain gauges. Calibration parameters based on linear regressions between CCD totals for a given temperature threshold and historical rainfall accumulations are generated and rainfall is estimated as a function of CCD and the calibration coefficients. TAMSAT version 3.0 uses a spatially and temporally varying bias correction approach for calibration to reflect the geographical and temporal variability in average rainfall. Detailed descriptions of the TAMSAT algorithm and data can be found in Tarnavsky et al. (2014). We used TAMSAT v3.1 product downloaded from the online data portal (http://www.tamsat.org.uk/data-subset/index.html) for the overlapping period of the TAHOE gauge observations.

GSMaP is a blended PMW-GeoIR rainfall product (Kubota et al. 2007, 2020; Aonashi 2009) provided primarily at a spatial resolution of 0.1° and hourly temporal resolution. It utilizes the greatest number of satellite inputs and is based on a technique that generates estimates by temporally interpolating PMW retrievals using a PMW-IR blending algorithm with a two-way morphing technique from Geo-IR images (Joyce et al. 2004) and a Kalman filter (Ushio et al. 2009). The moving vector derived from two successive IR images is used to propagate the rainy area from MW radiometry (MWR). This approach is similar to the Climate Prediction Centre (CPC) morphing technique (CMORPH) precipitation product which uses two IR images at 30-min intervals to calculate a moving vector of the cloud. Rainfall rates are estimated through propagation of the MWR estimates with the moving vector (Ushio et al. 2009).

The GSMaP algorithm focuses on the relationship between rainfall strength and IR brightness temperatures, and later, the Kalman filter is applied to the rainfall rates after propagation. This technique results in improved IR-based rain rates. Satellite data inputs in the GSMaP algorithm include MWR sensors [Special Sensor Microwave Imager (SSMI), TRMM Microwave Imager (TMI), and the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E)]. GSMaP also integrates Geo-IR sensors data from the National Oceanic and Atmospheric Administration (NOAA) acquired from Geo-IR radiometers aboard geostationary satellites (Janowiak et al. 2001).

GSMaP wGauge is the gauge-calibrated version of GSMaP. The GSMaP wGauge algorithm uses gauge analysis from the CPC unified gauge database of daily precipitation (Mega et al. 2019). Generally, the bias correction model optimizes the estimation of the rainfall rate in GSMaP wGauge at a given pixel. The first step in the model assumes that the rainfall rate for a given time period is the same as the time period 1 h prior. This assumption could potentially propagate yield errors as the actual rainfall rate increases or decreases with time. The distribution of the rate of rainfall is calculated from ground-based observations while the distribution of model error is calculated from radar data due to the unavailability of ground-based rainfall data for many regions. Rainfall rates generated from this step are applied globally in the GSMaP wGauge product.

The second step in the model assumes a linear relationship between GSMaP-MVK and GSMaP wGauge rainfall rates. This assumption is based on a correlation analysis between GSMaP-MVK and CPC gauge data, which showed the assumption to be statistically correct. In the third step, the system model optimizes the GSMaP wGauge rainfall rate in a given pixel by maximizing the probability density function of the GSMaP wGauge estimation. A detailed description of the GSMaP wGauge bias-correction algorithm can be found in Mega et al. (2019). We downloaded GSMaP and GSMaP wGauge hourly data from the JAXA website (https://sharaku.eorc.jaxa.jp/GSMaP) then aggregated to daily, pentadal, and monthly time scales for the same overlapping period as the TAHOE data.
Table 1. Description of the evaluation statistics used in this study. The terms $A$, $B$, and $C$ represent hits, false alarms, and misses, respectively; $S$, $G$, $G'$, and $N$ represent satellite-rainfall estimate, gauge rainfall measurements, mean of the gauge rainfall measurements, and the number of data pairs, respectively.

<table>
<thead>
<tr>
<th>Evaluation statistic</th>
<th>Formula</th>
<th>Unit</th>
<th>Best value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of detection</td>
<td>$POD = \frac{A}{A + C}$</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>False alarm ratio</td>
<td>$FAR = \frac{B}{A + B}$</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>$CC = \frac{\sum (G - G') (S - \bar{S})}{\sqrt{\sum (G - G')^2 \sum (S - \bar{S})^2}}$</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>$MAE = \frac{1}{N} \sum</td>
<td>S - G</td>
<td>$</td>
</tr>
<tr>
<td>Root-mean-square error</td>
<td>$RMSE = \sqrt{\frac{1}{N} \sum (S - G)^2}$</td>
<td>mm</td>
<td>0</td>
</tr>
<tr>
<td>Efficiency</td>
<td>$Eff = 1 - \frac{\sum (S - G)^2}{\sum (G - G')^2}$</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Multiplicative bias</td>
<td>$Bias = \frac{\sum S}{\sum G}$</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Critical success index</td>
<td>$CSI = \frac{A}{A + B + C}$</td>
<td>None</td>
<td>1</td>
</tr>
</tbody>
</table>

c. Evaluation approach

This section describes the approach used to compare each satellite rainfall product with the in situ data from TAHMO.

1) Spatial and temporal scale

Hydrometeorological applications that focus on flood and drought monitoring are done at various time scales. Hydrological analyses that assess floods tend to focus on shorter time scales, often at subdaily to dekadal (10 days), while those that focus on droughts tend to analyze longer time scales, such as months to seasons. Our validation was conducted at daily, pentadal (5 days), and monthly time scales over the designated study regions.

Validation of gridded satellite-rainfall products by gauge-based in situ observations relies on two main approaches: point-to-pixel and pixel-to-pixel comparisons. Point-to-pixel comparisons relate native rain gauge values to corresponding satellite pixel values directly (Cerón et al. 2020; Dembéle and Zwart 2016; Dinku et al. 2018). For pixel-to-pixel comparisons, gauge data are interpolated at a defined spatial resolution comparable to satellite rainfall data (Dinku et al. 2018). Each method has strengths and weaknesses. The pixel-to-pixel method smooths gauge values, which tends to decrease the accuracy of gauge data depending on the interpolation method and the underlying assumptions (Stampoulis and Anagnostou 2012; Nerini et al. 2015). We used the point-to-pixel approach in order to retain the original values of the gauge data. The pixel-to-pixel method would have smoothed extreme rainfall values that were of interest in rainfall detection skill statistics.

To prepare data for analysis, we converted TAHMO gauge locations into spatial point objects using latitude and longitude coordinates. We extracted grid values for each SRP on dates that intersected TAHMO observations at daily, pentadal, and monthly time scales. The observation period for this study was from April 2015 to December 2020. Erroneous and missing data in the TAHMO or SRP record were excluded from our analysis. CHIRPS data were downloaded for each of the three time scales separately since merging of CHIRP with station data is done at pentadal and monthly scales while daily data are created from the pentadal and monthly fields Dinku et al. (2018); Funk et al. (2015b). For the other SRPs, daily data were downloaded and aggregated to pentadal and monthly time scales.

2) Validation statistics

Well-studied and common validation statistics were used to evaluate the accuracy of the different satellite products against in situ rain gauge data (Dezfuli et al. 2017; Kimani et al. 2017; Dinku et al. 2018; Endris et al. 2013; Young et al. 2014; Le Coz and van de Giesen 2020; Toté et al. 2015). The definition and calculation of each statistic is given in Table 1.

The probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) were used to evaluate the accuracy of satellite products in detecting rainfall occurrence at the daily time scale. The POD assesses the ability of the satellite products to detect the occurrence of rainfall while the FAR is used to assess false rainfall detection. The CSI is the ratio of all events estimated and observed which are correctly detected. The best value for POD and CSI is 1, whereas 0 is the best score for FAR. These statistics rely on a known true observation. TAHMO data were used as the gold standard for all statistics. Rainfall events were defined as a
measurement 0.1 mm or greater for the gauge and satellite products.

The Pearson correlation coefficient (CC), multiplicative bias (bias), mean absolute error (MAE), root-mean-square error (RMSE), and the coefficient of efficiency (Eff) were used to evaluate the skill of the SRPs in estimating rainfall amounts. The CC measures the strength and direction of a linear relationship between two variables (i.e., satellite estimate and reference gauge measurement). Scores range from −1 to 1 with 1 indicating perfect agreement between the two. The bias statistic is a measure of how the average predicted value (satellite estimate) compares to a reference value (gage measurement). A value of one (1) is the most desirable. The MAE statistic quantifies the average absolute difference between the satellite product and the reference gauge measurement. A value of 0 indicates no difference. The RMSE statistic is interpreted in the same way as MAE but provides greater weight to outliers.

The Eff statistic demonstrates the skill of the satellite estimates relative to the mean of the reference gauge data. Eff scores vary from minus infinity to 1, where negative values indicate that the gauge data are a better estimate than the satellite product and a value of 1 means a perfect match between the reference gauge data and the satellite estimate (Toté et al. 2015). The POD, FAR, CSI, and Eff are used widely in validation studies (Ayehu et al. 2018; Boluwade 2020; Dezfuli et al. 2017; Toté et al. 2015; Dinku et al. 2018; Cattani et al. 2016). In most of Africa, gauge stations are predominantly found in densely populated agricultural areas where it is relatively easy to consistently monitor and carry out gauge station maintenance (Dinku et al. 2018). As a result, validation results are representative of the performance of SRPs in those areas and may not be generalized over larger areas that are not covered by gauge stations. TAHMO stations are mostly found in these high population density areas albeit further away from most ground stations managed by other observational networks such as the NHMS.

3. Results

a. Summary statistics for the gauge and satellite rainfall data

Table 2 shows the overall descriptive statistics for TAHMO gauge observations and the SRPs at the daily time scale over Africa. CHIRPS, TAMSAT, and GSMaP had higher mean daily rates than TAHMO while GSMaP_wGauge exhibited comparable mean daily rainfall to TAHMO gauge stations at 2.9 mm day$^{-1}$. Overall comparisons between TAHMO gauge observations and the four SRPs showed that over Africa, CHIRPS, TAMSAT, and GSMaP overestimated mean daily rainfall with CHIRPS having the largest overestimation at 0.5 mm day$^{-1}$ or a wet bias of 15.5%. TAMSAT and GSMaP overestimated the daily mean rainfall by 0.3 and 0.1 mm day$^{-1}$, respectively. GSMaP_wGauge had a small underestimation of 0.02 mm day$^{-1}$. The performance of GSMaP_wGauge suggests that this rainfall product has the best potential to accurately predict daily rainfall amounts. The variability as measured by the coefficient of variation (CV) and variance was larger for TAHMO gauge observations than all the SRPs. CHIRPS and TAMSAT had the least CV and the lowest daily rainfall range (0–330.3 and 0–117.9 mm day$^{-1}$, respectively) while TAHMO and GSMaP had the highest ranges (0–629.3 and 0–617.9 mm day$^{-1}$). The 25th and 50th percentile daily rainfall was 0 for all SRPs while the 75th percentile rainfall was 0.7 mm for TAHMO, greater than 3 mm for CHIRPS and TAMSAT, 0.5 mm for CHIRPS, TAMSAT and GSMaP overestimated mean daily rainfall with CHIRPS having the largest overestimation at 0.5 mm day$^{-1}$ or a wet bias of 15.5%. TAMSAT and GSMaP overestimated the daily mean rainfall by 0.3 and 0.1 mm day$^{-1}$, respectively. GSMaP_wGauge had a small underestimation of 0.02 mm day$^{-1}$. The performance of GSMaP_wGauge suggests that this rainfall product has the best potential to accurately predict daily rainfall amounts. The variability as measured by the coefficient of variation (CV) and variance was larger for TAHMO gauge observations than all the SRPs. CHIRPS and TAMSAT had the least CV and the lowest daily rainfall range (0–330.3 and 0–117.9 mm day$^{-1}$, respectively) while TAHMO and GSMaP had the highest ranges (0–629.3 and 0–617.9 mm day$^{-1}$). The 25th and 50th percentile daily rainfall was 0 for all SRPs while the 75th percentile rainfall was 0.7 mm for TAHMO, greater than 3 mm for CHIRPS and TAMSAT, 1.4 mm for GSMaP and 2.3 mm for GSMaP_wGauge. This means that 25% of the time, satellite rainfall estimates were 1.4 mm or greater than the upper quartile measured by the TAHMO gauges.

b. Validation at the daily time scale

The accuracy of the SRPs in detecting daily rainfall was assessed using the POD, Eff, CSI, and FAR statistics. Results in Table 3 showed that over Africa, GSMaP_wGauge had the most accurate rainfall detection scores (POD) followed by GSMaP, TAMSAT and CHIRPS. Of the total number of observations over Africa (441 768), 46% of days recorded rainfall events according to the TAHMO gauge stations. Comparatively, CHIRPS, TAMSAT, and GSMaP predicted fewer events than TAHMO (29%, 34%, and 44%, respectively) whereas GSMaP_wGauge predicted slightly more events (48%) than TAHMO. The number of rainfall events predicted by the satellite products were fewer for CHIRPS, TAMSAT, and GSMaP than the events recorded by the gauge observations by a factor of 0.63, 0.75, and 0.95 times, respectively, and greater for GSMaP_wGauge by a factor of 1.1 times. GSMaP_wGauge had better POD scores in East Africa, southern Africa, and West Africa compared to the other SRPs, further demonstrating its preferential rainfall detection performance even at a subregional level. In most cases, GSMaP_wGauge had also the best Eff and CSI scores over Africa and in the subregions. GSMaP_wGauge showed adequate detection performance

<table>
<thead>
<tr>
<th></th>
<th>TAHMO</th>
<th>CHIRPS</th>
<th>TAMSAT</th>
<th>GSMaP</th>
<th>GSMaP_wGauge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>2.9</td>
<td>3.35</td>
<td>3.21</td>
<td>3.14</td>
<td>2.88</td>
</tr>
<tr>
<td>Stdev (mm)</td>
<td>9.9</td>
<td>8.4</td>
<td>6.7</td>
<td>9.5</td>
<td>7.2</td>
</tr>
<tr>
<td>Variance (mm$^2$)</td>
<td>97.9</td>
<td>70.5</td>
<td>45.3</td>
<td>89.4</td>
<td>51.8</td>
</tr>
<tr>
<td>CV (%)</td>
<td>341.3</td>
<td>250.9</td>
<td>210</td>
<td>301.1</td>
<td></td>
</tr>
<tr>
<td>75th percentile (mm)</td>
<td>0.7</td>
<td>3.2</td>
<td>3.8</td>
<td>1.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Maximum (mm)</td>
<td>629.3</td>
<td>330.3</td>
<td>117.9</td>
<td>617.9</td>
<td>374.2</td>
</tr>
</tbody>
</table>
over Africa (0.11), East Africa (0.07), and southern Africa (0.18). GSMaP and GSMaP_wGauge had similar Eff performance in southern Africa (0.18). TAMSAT and CHIRPS had better Eff scores than the GSMaP products in West Africa (0.09 and 0.07, respectively) whereas the GSMaP products had negative scores that implied the mean of the gauge data was a better predictor of daily rainfall in West Africa than the mean of the GSMaP products.

Quantitative validation results in Table 4 showed that all SRPs had low linear correlation with TAHMO gauge observations. Generally, the CC values of the SRPs were less than 0.5 over Africa and in the three subregions with the exception of GSMaP in southern Africa which had a CC of 0.51. CHIRPS had low CC values in the four regions but performed slightly better than TAMSAT. GSMaP products had the highest CC values in the four regions. GSMaP_wGauge had CC scores of 0.40, 0.46, and 0.44 in East Africa, West Africa, and Africa, respectively, whereas GSMaP had a correlation of 0.51 over southern Africa. In terms of volumetric errors over Africa (0.11), East Africa (0.07), and Southern Africa (0.18), GSMaP had the best CC but also the highest RMSE in West Africa. The most likely explanation of this result is that the GSMaP product had more outliers relative to the gauge observations compared to the other SRPs, and this could have influenced the RMSE which is sensitive to large errors.

Biases in mean daily rainfall were observed in all the SRPs in the three subregions. Mean daily rainfall from the gauge observations in East Africa, Southern Africa, and West Africa was 3.04, 2.5, and 2.89 mm, respectively. CHIRPS, TAMSAT, and GSMaP overestimated the mean daily rainfall (bias scores > 1) in West Africa and underestimated it in Southern Africa (bias scores < 1). For the purpose of interpreting rainfall bias, wet bias means the satellite products predicted more rainfall (overestimation) than the gauge stations whereas a dry bias means that the satellite products predicted lower rainfall (underestimation) than the gauge stations. All SRPs but GSMaP underestimated mean daily rainfall in East Africa. GSMaP_wGauge underestimated mean daily rainfall in West Africa. With respect to rainfall amounts, these biases revealed that in East Africa, CHIRPS had the largest wet bias at 0.8 mm (28%) compared to the gauge observations. TAMSAT and GSMaP_wGauge had wet biases of 0.4 mm (11%) and 0.2 mm (7%), respectively, whereas GSMaP had a dry bias of 0.03 mm (2%). In Southern Africa, all satellite products underestimated daily rainfall, by 0.4 mm (16%) for TAMSAT, 0.3 mm (15%) for both GSMaP products, and 0.04 mm (1.4%) for CHIRPS. Wet biases of 0.2 mm (6%), 0.3 mm (11%), and 0.7 mm (23%) were observed in West Africa for CHIRPS, TAMSAT, and GSMaP, respectively, whereas GSMaP_wGauge had a dry bias of 0.1 mm (3%).

Table 4. Quantitative validation results for daily time scale over Africa and the three subregions. CC, MAE, and RMSE represent correlation coefficient, mean absolute error, and root-mean-square error, respectively. Values represent validation scores between the satellite rainfall product and TAHMO gauge observations. Bold values show the best score for the respective statistic; N is the total number of observations for the respective region.
The variability in correlations between the satellite products and TAHMO in situ data at the level of gauge stations is shown in Fig. 2. Higher correlations were observed in West Africa and southern Africa than in East Africa. Over Africa, correlations ranged between 0.02 and 0.81, between −0.01 and 0.8, between 0.05 and 0.9, and between −0.01 and 0.88 for CHIRPS, TAMSAT, GSMaP, and GSMaP_wGauge, respectively. Very low correlation values could be a result of serious failures by either satellite products or rain gauges to correctly measure rainfall in some locations. We investigated in situ data by flagging gauge stations that could have contributed to the very low correlations. Of the 17 TAHMO gauge stations identified as having lower correlations (Fig. 3), three gauge stations (TA00698, TA00709, and TA00052) had negative correlations with at least one satellite product. The elevation of these stations was 1241, 1206, and 305 m, respectively. The TAMSAT product had low negative correlations with TA00698 and TA00052 whereas the correlations between the stations and the other SRPs were moderate to high. We also found that the correlation between TA00052 and GSMaP_wGauge was negative but the correlations with CHIRPS, TAMSAT, and GSMaP were positive. In general, four of the flagged stations were located at low elevation (<700 m), four were at moderate elevation (1000–1500 m) and six were at high elevation (>1500 m). These results indicate that some of the satellite products had large errors in rainfall prediction at those gauge station locations and could further highlight the propensity for errors by satellite rainfall products in low and high elevations.

Correlations between the satellite rainfall products are shown in Fig. 4. The highest correlation was between GSMaP_wGauge and GSMaP (CC = 0.63) compared to correlation values for any other pair of SRPs. This high correlation can be explained by the similarities between the two products. The only difference between the two products is the integration of gauge observations in the former which had a clear impact on the performance of the GSMaP_wGauge product. Other high correlations were observed between CHIRPS and TAMSAT (CC = 0.58), and CHIRPS and GSMaP_wGauge (CC = 0.53). CHIRPS and TAMSAT algorithms contain similarities and this likely explains the relatively high CC value. Both algorithms are largely based on Geo-IR observations and some of their bias correction gauge observations come from similar sources such as the GTS and data from NHMS across Africa.
TAMSAT for all elevation categories. Simulations and higher rainfall detection scores than CHIRPS and GSMaP had a near-identical distribution with the TAHMO in situ CDF for low-intensity rainfall events over Africa and in the three subregions. CHIRPS and TAMSAT CDFs are similar for low and high rainfall, with particularly higher overestimation of low rainfall events compared to both GSMaP products. Further, GSMaP had a near-identical distribution with the TAHMO in situ CDF in East Africa and southern Africa but the pattern in West Africa was similar to the pattern observed over Africa.

c. Validation at the pentadal time scale

Results at the pentadal time scale are shown in Table 5. The pentadal time scale has a pattern which is notably similar to the daily time scale, although there were some improvements in CC values over Africa and the three subregions. CHIRPS and TAMSAT had the largest gain in performance from the change in temporal resolution. Overall results showed that CHIRPS and GSMaP_wGauge products had nearly similar CC values in Africa (0.52 and 0.53) and in East Africa (0.53 and 0.52). Overall, this was a gain of 44% and 41%, respectively, over Africa relative to the daily CC values, and 51% and 41%, respectively, in East Africa. In southern Africa, the two GSMaP products had higher CC values at 0.61 for GSMaP and 0.60 for GSMaP_wGauge representing an overall gain of 20% and 25%. In West Africa, all the products had similar CC values of ~0.5 but CHIRPS and TAMSAT showed the largest gain of 31% and 35%, followed by GSMaP_wGauge which had a gain of 11%. Another noteworthy change was in MAE and RMSE statistics where CHIRPS had lower values than TAMSAT and GSMaP over Africa, the lowest MAE and RMSE values in East Africa, and second lowest MAE and RMSE values in West Africa. Comparatively, CHIRPS had the highest RMSE by rank over Africa, East Africa, and southern Africa. We did not find significant differences in the mean bias patterns at the pentadal time scale compared to the daily time scale. The cumulative distribution function plots (Fig. 7) had similar patterns to the plots for daily time scales.

d. Validation at the monthly time scale

The SRPs demonstrated similar performance in all study regions at the monthly time scale. CC values over Africa were not distinct by SRP although GSMaP_wGauge and CHIRPS

![Figure 3. Correlation coefficients of 17 TAHMO stations that had low mean correlation values with one or more SRPs over Africa.](image-url)
had slightly higher values over Africa and East Africa (Table 6). 
CHIRPS and TAMSAT had similar CC values in West Africa, 
but only a 1.5% improvement over GSMaP_wGauge. There 
were very small differences in CC values for all SRPs in southern 
Africa. Correlations at the monthly time scale improved from 
those at the daily time scale. Monthly CC values were moderate 
for all regions, except in southern Africa where improvements 
were notably large. The largest improvement in correlation from 
daily to monthly rainfall estimates was in southern Africa for 
TAMSAT (94%) and CHIRPS (92%). The overall gain in CC 
values over Africa was higher for TAMSAT than CHIRPS 
(88% versus 83%). The GSMaP products were as correlated 
with gauge data as CHIRPS and TAMSAT for monthly rainfall 
but had smaller relative gains in correlation because CCs at the 
daily time scale were initially higher.

MAE and RMSE values over Africa were high for all satel-
lite products (49.2–57.9 mm and 82.06–96.32 mm, respectively) 
as well as in each subregion. GSMaP_wGauge exhibited the 
lowest MAE and RMSE values over Africa and East Africa, 
and lowest MAE and bias in West Africa. However, TAMSAT

**FIG. 4.** Comparisons between TAHMO gauge observations and satellite rainfall products for daily time scale over Africa and the three 
subregions. The lower and upper panels show scatterplots and correlation coefficients (r) between pairs of data products, respectively. 
The diagonal panel shows density distributions for each data product. All correlations were statistically significant (p < 0.001).
had the lowest RMSE value in West Africa. GSMaP demonstrated the highest MAE and RMSE values over Africa, East Africa, and West Africa whereas TAMSAT produced the highest values for the same metrics in southern Africa.

As with the daily and pentadal time scales, the CC values between CHIRPS and TAMSAT, and between CHIRPS and GSMaP_wGauge were higher than the correlations of any other pairs of SRPs over Africa, southern Africa, and West Africa (Fig. 9).

Figure 10 shows the annual rainfall cycle described by each SRP and the TAHMO in situ gauge observations across settings. All SRPs reproduced trends in annual rainfall in each region well. In East Africa, all satellite products overestimated the magnitude of April rainfall, the peak of the long rainy season [March–May (MAM)]. TAMSAT and CHIRPS overestimated average rainfall during this month the most [74% (115.5 mm) and 58% (90.9 mm), respectively]. The monthly mean, seasonal, and annual rainfall statistics for the three subregions are shown in Table 7. In East Africa, the MAM season total rainfall measured by TAHMO gauge stations was 392.7 mm. The satellite rainfall biases for this season were 171.8 and 149.2 mm for CHIRPS and TAMSAT, and 86.4 and 76.5 mm for GSMaP and GSMaP_wGauge, respectively. All SRPs except GSMaP overestimated total rainfall during the short rainy season [October–December (OND)] and the amount during the peak in November. CHIRPS and TAMSAT reported rainfall amounts 89.1 and 13.0 mm greater than rainfall recorded at TAHMO gauge stations while GSMaP underestimated rainfall by 64.6 mm. Rainfall estimated by GSMaP_wGauge was only 9.1 mm above the mean total OND rainfall measured by the TAHMO gauge stations. CHIRPS showed the largest wet bias of 47.0 mm in November whereas GSMaP had a dry bias of 23.1 mm for the same month. The long dry season (June–September) rainfall was also generally overestimated by CHIRPS and TAMSAT. Differences in estimation of annual rainfall revealed that CHIRPS, TAMSAT, and GSMaP_wGauge overestimated total annual rainfall by 310 mm (28.8%), 175 mm (16.2%), and 66 mm (6.1%), respectively, and GSMaP underestimated total annual rainfall in this subregion by 26 mm (2.4%).

Southern Africa has a long wet season that begins in October/November and, during this time there was less monthly variability between the satellite estimates and TAHMO. The SRPs had similar estimates of monthly rainfall and small wet and dry biases. Seasonal bias across satellite products was similar to observed annual bias due the fact that the rainy season [November–April (NDJFMA)] represents over 85% of total annual rainfall. Annual rainfall estimated by CHIRPS was 0.1% (1.1 mm) above the total rainfall recorded by the gauge stations while TAMSAT, GSMaP, and GSMaP_wGauge predicted lower rainfall by 127 mm (13.6%), 104 mm (11.2%), and 107 mm (11.5%), respectively.

In West Africa, trends in annual rainfall are influenced by the two main subregions where TAHMO stations are located: the Guinea coast and the Sahel. The rainfall season begins around April/May and ends in October/November with most
rainfall falling during the monsoon season (June–September). The distribution of monthly rainfall by CHIRPS, TAMSAT, and GSMaP_wGauge replicated the TAHMO observations well. All SRPs captured rainfall peaks in June and September. However, GSMaP demonstrated a large wet bias between March and June, which was reduced for the remaining months and similar to the other SRPs. The bias in GSMaP between March and June ranged from 10% to 80% of mean monthly rainfall measured at the gauge stations. The satellite products overestimated rainfall for the main rainy season [June–September (JJAS)] with TAMSAT having the highest wet bias of 104.4 mm followed by CHIRPS (98.4 mm), GSMaP (54.0 mm), and GSMaP_wGauge (31.3 mm). In addition, the satellite products overestimated total annual rainfall and wet biases were 185 mm (17.8%), 220 mm (21.1%), 335 mm (32.1%), and 49 mm (4.7%) for CHIRPS, TAMSAT, GSMaP, and GSMaP_wGauge, respectively. Similar observations can be drawn from the CDF plots in Fig. 11.

e. Investigating the relationship between TAHMO and other gauge observations in Kenya

Validation results can differ remarkably depending on the quality of the gauge observations used as reference data. Although the aim of this validation study was not to compare TAHMO to NHMS data, we reviewed the agreement between TAHMO and Kenya Meteorological Department (KMD) gauge observations for demonstration. Data from 11 rainfall stations operated by the KMD and collocated with a TAHMO station (i.e., KMD stations that were within 5 km of a TAHMO station) were collected at the daily time scale. Correlation between data sources was assessed using the CC statistic (Table 8). The highest correlation between a KMD station

<table>
<thead>
<tr>
<th>Table 5. As in Table 4, but for pentad time scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa (N = 87622)</td>
</tr>
<tr>
<td>CC</td>
</tr>
<tr>
<td>CHIRPS</td>
</tr>
<tr>
<td>TAMSAT</td>
</tr>
<tr>
<td>GSMaP</td>
</tr>
<tr>
<td>GSMaP_wGauge</td>
</tr>
</tbody>
</table>

FIG. 6. Comparison of cumulative distribution functions of the TAHMO gauge observations and satellite rainfall products for the daily time scale over Africa and the three subregions. The dotted vertical line is the 20-mm cutoff used to categorize low (<20 mm) and high (>20 mm) rainfall events.
and a collocated TAHMO station was 0.71. Correlation values for the other collocated pairs were low to moderate. Correlation initially appeared to decrease by distance, up to 2.36 km, but was higher for stations 2.36 km apart and increased among stations at a distance between 3.28 and 4.76 km.

4. Discussion

a. Overall characteristics of the mean daily rainfall

We evaluated four satellite rainfall products—three that are gauge-calibrated (CHIRPS, TAMSAT, and GSMaP_wGauge) and one satellite-only product (GSMaP)—using 6 years of independent TAHMO in situ gauge observations over Africa and East Africa, southern Africa, and West Africa subregions. We evaluated rainfall detection and volumes at daily, pentadal, and monthly time scales. Although the three gauge-calibrated products integrate data from other gauge stations in Africa, TAHMO stations are an independent data source for validation. Mean daily rainfall measured in situ was 2.9 mm. Comparisons between the TAHMO gauge observations and the rainfall satellite products over Africa showed that CHIRPS had the largest wet bias followed by TAMSAT and GSMaP whereas GSMaP_wGauge had a small dry bias.

At the subregional level, every satellite-derived product had biases. Mean daily rainfall measured by the gauge stations in East Africa was overestimated by CHIRPS by 28%, which was greater than the bias in any other product by a factor of 2 or more times. Mean daily rainfall in southern Africa was underestimated by all SRPs, with TAMSAT having the largest dry bias closely followed by the GSMaP products. The bias in CHIRPS was less than that in TAMSAT by a factor of 10 times. In West Africa, GSMaP overestimated mean daily rainfall by 23%, a larger bias than in any other SRP by a factor of 2–7. The relatively small biases in the GSMaP products compared to CHIRPS and TAMSAT across Africa and in two of the three subregions shows that the GSMaP algorithm performs well at predicting daily rainfall amounts in the region.

b. Errors in rainfall event detection

In terms of overall daily rainfall event detection, our results indicate that GSMaP products have better scores for most of the accuracy evaluation statistics. These results were calculated from all observations including both dry and wet seasons. The POD scores for the GSMaP products are higher than the scores for CHIRPS and TAMSAT by a factor of 1.5 or more in all regions. Generally, for every rainfall event detected by TAHMO gauge stations, GSMaP_wGauge predicted 1.1 events which was more events than any other satellite product. Critically, accurate rainfall detection is important during the growing season because the productivity of many livelihoods in Africa depend on the occurrence of rainfall. In East Africa, for every rainfall event detected by TAHMO gauge stations in the long rains season (MAM), GSMaP predicted the same number of events while CHIRPS and TAMSAT predicted fewer events by a factor of 0.8 and 0.7, respectively, and GSMaP_wGauge predicted more events by a factor of 1.1.
GSMaP_wGauge predicted the same number of rainfall events as the TAHMO gauge stations in the short rains season (OND) whereas CHIRPS, TAMSAT, and GSMaP predicted fewer events.

In southern Africa, for every rainfall event detected by TAHMO in the rainy season (NDJFMA), GSMaP and GSMaP_wGauge predicted the same number of events and both CHIRPS and TAMSAT predicted fewer events by a factor of 0.7 times. In West Africa, TAMSAT predicted the same number of events as the TAHMO stations in the rainy season (JJAS) confirming the results from Maidment et al. (2017), CHIRPS predicted fewer events by a factor of 0.8, and GSMaP and GSMaP_wGauge predicted more events by a factor of 1.1 times each. Studies evaluating rainfall detection capability for GSMaP products over Africa are lacking but in studies in other regions, e.g., Tian et al. (2010), GSMaP was shown to have good rainfall detection at high temporal scales over the continental United States. These results indicate that for applications that require good rainfall event detection capability, the GSMaP products may be most applicable in East Africa and southern Africa for the regional rainy seasons while TAMSAT may be a better choice for use in West Africa.

**FIG. 8.** As in Fig. 3, but for the pentadal time scale.
In terms of the ratio of falsely identified rainfall events (i.e., FAR scores), over Africa, CHIRPS performed slightly better than the other satellite products, although only by 3.6% relative to TAMSAT and 7.0% to the GSMaP products. A similar observation was made in East Africa. The FAR scores for TAMSAT were better than scores for the other SRPs in southern Africa and West Africa. These results mirror findings by Maidment et al. (2017), where they compared TAMSAT and CHIRPS with ground observations over countries in the three subregions. Their results showed FAR values that were within the ranges of the values that we found for the same products. While CHIRPS and TAMSAT had lower FAR scores, we found that the two products had lower rainfall detection scores (i.e., POD) compared to the GSMaP products. These POD results differ with Maidment et al. (2017) that showed high POD scores for TAMSAT over southern Africa and West Africa (0.76–0.9). Further investigation of our data revealed that the lower FAR but poorer POD scores for CHIRPS and TAMSAT was because the two products predicted fewer rainfall events in total than the GSMaP products, which likely explains their improved FAR scores. The FAR statistic quantifies the fraction of predicted rainfall events that are true nonevents. It is expected that a rainfall product with low rainfall detection capability will produce fewer rainfall events overall resulting in a higher ratio of false positive events (Camici et al. 2020).

All SRPs had low efficiency scores over Africa and the three subregions. Models that attain efficiency scores between 0 and 1 are considered to have acceptable performance while negative scores indicate poor performance (Moriasi et al. 2007). GSMaP_wGauge showed low but positive scores over Africa, in East Africa, and in southern Africa. TAMSAT and CHIRPS demonstrated acceptable performance in West Africa. Although few studies have validated GSMaP-based rainfall products over Africa, Derin et al. (2016) also found that GSMaP_wGauge had good rainfall detection capability over the Blue Nile basin in East Africa, supporting our observation that this product can be used for rainfall detection in East Africa.

c. Errors in estimating rainfall amount

Each SRP reproduced expected trends in annual rainfall cycles over the study regions. In terms of volumetric errors, CHIRPS and TAMSAT showed consistently inferior performance to the GSMaP products at the daily time scale in all the study regions. The only exception was southern Africa, where CHIRPS had the best bias score compared to the other products. The similarity in the CHIRPS and TAMSAT algorithms and data sources likely explains their comparable performance across multiple statistics and regions, although we noted that TAMSAT performed marginally better than CHIRPS at the daily and pentadal time scales. This result was consistent with findings by Dinku et al. (2018) in East Africa and Boluwade (2020) in West Africa. The three gauge-calibrated products showed lower variability in estimated rainfall quantity than the satellite-only GSMaP product and the TAHMO gauge observations. Moreover, the satellite-only GSMaP product and the TAHMO gauges reported similar maximum daily rainfall amounts. Quality control of the TAHMO data may have missed high values that should otherwise have been classified as erroneous and flagged for exclusion, and this should be a consideration for further investigation. We also noted the impact of gauge corrections when comparing GSMaP and GSMaP_wGauge. Between these two, gauge corrections reduced the maximum daily rainfall amount by 39% and reduced the CV by 17%. While correction by in situ rainfall gauges may reduce errors in satellite products, it could also reduce the variability (Lu et al. 2020) in observed rainfall quantities as indicated by the lowered CV, variance, and standard deviation of the gauge-calibrated products.

d. The impact of temporal aggregation on rainfall errors

We found an increase in correlation between the satellite products and TAHMO gauge observations when aggregating monthly versus daily. Temporal aggregation seemed to have the most impact for CHIRPS and TAMSAT whose correlation coefficients increased by 88% and 83%, respectively. The increases were smaller for GSMaP (34%) and GSMaP_wGauge (47%) because the correlation scores at the daily time scale were higher compared to CHIRPS and TAMSAT. The increase in correlations with temporal aggregation is a common characteristic. Temporal and spatial aggregation reduces random errors which makes the estimate from satellite products align more closely with a reference value from ground observations (Dembélé and Zwart 2016; Kimani et al. 2017; Gangopadhyay et al. 2004). Users of rainfall data at an individual pixel level (e.g., 5 km) can expect to have higher errors compared to users of data spatially aggregated at larger spatial scales (e.g., 50 km). Similarly, as we observed in our results, aggregating data over longer time periods has an effect on mean rainfall errors.

In this study, the SRP with the highest correlation with TAHMO gauge observations was CHIRPS at the monthly scale. Each SRP reproduced expected trends in annual rainfall cycles over the study regions. In terms of volumetric errors, CHIRPS and TAMSAT showed consistently inferior performance to the GSMaP products at the daily time scale in all the study regions. The only exception was southern Africa, where CHIRPS had the best bias score compared to the other products. The similarity in the CHIRPS and TAMSAT algorithms and data sources likely explains their comparable performance across multiple statistics and regions, although we noted that TAMSAT performed marginally better than CHIRPS at the daily and pentadal time scales. This result was consistent with findings by Dinku et al. (2018) in East Africa and Boluwade (2020) in West Africa. The three gauge-calibrated products showed lower variability in estimated rainfall quantity than the satellite-only GSMaP product and the TAHMO gauge observations. Moreover, the satellite-only GSMaP product and the TAHMO gauges reported similar maximum daily rainfall amounts. Quality control of the TAHMO data may have missed high values that should otherwise have been classified as erroneous and flagged for exclusion, and this should be a consideration for further investigation. We also noted the impact of gauge corrections when comparing GSMaP and GSMaP_wGauge. Between these two, gauge corrections reduced the maximum daily rainfall amount by 39% and reduced the CV by 17%. While correction by in situ rainfall gauges may reduce errors in satellite products, it could also reduce the variability (Lu et al. 2020) in observed rainfall quantities as indicated by the lowered CV, variance, and standard deviation of the gauge-calibrated products.

### Table 6. As in Table 4, but for monthly time scale.

<table>
<thead>
<tr>
<th>Region</th>
<th>CC</th>
<th>MAE</th>
<th>Bias</th>
<th>RMSE</th>
<th>CC</th>
<th>MAE</th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>0.66</td>
<td>50.92</td>
<td>1.18</td>
<td>85.85</td>
<td>0.64</td>
<td>58.45</td>
<td>1.29</td>
<td>94.56</td>
</tr>
<tr>
<td>East Africa</td>
<td>0.6</td>
<td>53.79</td>
<td>1.12</td>
<td>89.8</td>
<td>0.53</td>
<td>63.23</td>
<td>1.16</td>
<td>102.84</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>0.59</td>
<td>57.89</td>
<td>1.08</td>
<td>96.32</td>
<td>0.53</td>
<td>64.02</td>
<td>0.97</td>
<td>103.89</td>
</tr>
<tr>
<td>West Africa</td>
<td>0.65</td>
<td>49.16</td>
<td>1</td>
<td>82.06</td>
<td>0.61</td>
<td>54.5</td>
<td>1.06</td>
<td>86.14</td>
</tr>
</tbody>
</table>

In this study, the SRP with the highest correlation with TAHMO gauge observations was CHIRPS at the monthly scale. Each SRP reproduced expected trends in annual rainfall cycles over the study regions. In terms of volumetric errors, CHIRPS and TAMSAT showed consistently inferior performance to the GSMaP products at the daily time scale in all the study regions. The only exception was southern Africa, where CHIRPS had the best bias score compared to the other products. The similarity in the CHIRPS and TAMSAT algorithms and data sources likely explains their comparable performance across multiple statistics and regions, although we noted that TAMSAT performed marginally better than CHIRPS at the daily and pentadal time scales. This result was consistent with findings by Dinku et al. (2018) in East Africa and Boluwade (2020) in West Africa. The three gauge-calibrated products showed lower variability in estimated rainfall quantity than the satellite-only GSMaP product and the TAHMO gauge observations. Moreover, the satellite-only GSMaP product and the TAHMO gauges reported similar maximum daily rainfall amounts. Quality control of the TAHMO data may have missed high values that should otherwise have been classified as erroneous and flagged for exclusion, and this should be a consideration for further investigation. We also noted the impact of gauge corrections when comparing GSMaP and GSMaP_wGauge. Between these two, gauge corrections reduced the maximum daily rainfall amount by 39% and reduced the CV by 17%. While correction by in situ rainfall gauges may reduce errors in satellite products, it could also reduce the variability (Lu et al. 2020) in observed rainfall quantities as indicated by the lowered CV, variance, and standard deviation of the gauge-calibrated products.
time scale and GSMaP_wGauge at the daily time scale over Africa and all the three subregions. We also found that GSMaP_wGauge attained nearly the same correlation values as CHIRPS over Africa and East Africa. The satellite-only GSMaP product had lower correlations compared to the other gauge-calibrated SRPs in all regions except in southern Africa where it performed slightly better than TAMSAT and GSMaP_wGauge. This pattern was, however, not discernible in other volumetric error statistics. The MAE, RMSE, and bias scores were higher for longer time scales compared to shorter time scales. The RMSE increased by a factor of 8 or more for all SRPs between the daily and monthly time scales. Partly, this is expected because the quantity of total rainfall increases with longer observation periods, thereby increasing unit magnitude, but it could also be attributed to error propagation when daily rainfall data were aggregated to pentadal and monthly time scales. The GSMaP_wGauge product had the lowest relative RMSE scores over Africa, in the three subregions at the daily time scale, and in East Africa at pentadal and monthly time scales. The performance of CHIRPS in this study agrees with findings from studies that assessed the product at regional levels (Dinku et al. 2018). Specifically,
Dinku et al. (2018) showed that the mean absolute error for CHIRPS and TAMSAT in Ethiopia, Kenya, and Tanzania (East Africa subregion) doubled between the dekadal (10 days) and monthly time scales which further confirms error propagation through temporal aggregation.

The implication of an increase in the correlation between the SRPs and TAHMO gauge observations coupled with an increase in the RMSE with temporal aggregation is that satellite-derived rainfall estimation attains good correspondence with gauge observations in general, but the magnitude of the bias in the satellite products grows. The RMSE statistic is sensitive to error outliers (Entekhabi et al. 2010). We observed that mean rainfall bias in each satellite product increased when rainfall was accumulated to pentadal and monthly time scales. This observation is important when rainfall estimates are applied to other biophysical processes. For instance, soil moisture modeling in regions without other in situ data could rely on satellite-derived rainfall that captures the linear trend in actual on-the-ground rainfall but contains a high magnitude of error; error which would then be propagated in soil moisture outputs. Such outputs could incorrectly describe an important soil condition indicator that is relied on by many researchers and decision makers to assess drought risks.

Table 7. The amount of monthly, seasonal, and annual rainfall (mm) for all gauge stations and corresponding satellite grid pixels averaged over each subregion.

<table>
<thead>
<tr>
<th>Subregion</th>
<th>TAHMO</th>
<th>CHIRPS</th>
<th>TAMSAT</th>
<th>GSMaP</th>
<th>GSMaP_wGauge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>East Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly mean</td>
<td>90.6</td>
<td>116.9</td>
<td>105.0</td>
<td>88.1</td>
<td>96.1</td>
</tr>
<tr>
<td>Seasonal (MAM) total</td>
<td>392.7</td>
<td>564.5</td>
<td>541.9</td>
<td>479.1</td>
<td>469.2</td>
</tr>
<tr>
<td>Seasonal (OND) total</td>
<td>326.4</td>
<td>415.4</td>
<td>339.4</td>
<td>261.8</td>
<td>335.5</td>
</tr>
<tr>
<td>Annual total</td>
<td>1078.7</td>
<td>1388.9</td>
<td>1253.8</td>
<td>1052.6</td>
<td>1144.2</td>
</tr>
<tr>
<td><strong>Southern Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly mean</td>
<td>81.4</td>
<td>81.3</td>
<td>71.0</td>
<td>73.4</td>
<td>72.3</td>
</tr>
<tr>
<td>Seasonal (NDJFMA) total</td>
<td>831.8</td>
<td>860.4</td>
<td>773.2</td>
<td>783.2</td>
<td>757.5</td>
</tr>
<tr>
<td>Annual total</td>
<td>935.8</td>
<td>936.9</td>
<td>808.5</td>
<td>831.4</td>
<td>828.4</td>
</tr>
<tr>
<td><strong>West Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly mean</td>
<td>83.5</td>
<td>92.1</td>
<td>94.9</td>
<td>103.0</td>
<td>81.4</td>
</tr>
<tr>
<td>Seasonal (JJAS) total</td>
<td>548.6</td>
<td>647.0</td>
<td>653.0</td>
<td>602.6</td>
<td>579.9</td>
</tr>
<tr>
<td>Annual total</td>
<td>1042.3</td>
<td>1145.5</td>
<td>1180.3</td>
<td>1284.1</td>
<td>1016.0</td>
</tr>
</tbody>
</table>
The impact of satellite product errors on daily, monthly, and seasonal rainfall

CHIRPS and TAMSAT were more likely to overestimate low rainfall events (below 20 mm day$^{-1}$) and underestimate daily rainfall events that exceeded this amount compared to the GSMaP products. With respect to categorizing drought, CHIRPS and TAMSAT would show wetter conditions for low rainfall events and drier conditions for high rainfall events.

Both GSMaP products had low biases in mean monthly rainfall in East Africa whereas CHIRPS and TAMSAT showed the largest overestimation. This observation was also made in the MAM and OND seasons. CHIRPS predicted the same amount of mean daily and monthly rainfall as TAHMO gauge stations in southern Africa, but it had a small wet bias of 3.4% at the seasonal time scale. Dry biases were observed in all the other products for monthly, seasonal, and annual rainfall. TAMSAT and GSMaP had the largest mean monthly rainfall biases in West Africa whereas GSMaP_wGauge had a dry bias in the same subregion. The total amount of rainfall in the JIAS season in West Africa was overestimated by all the satellite products with GSMaP_wGauge showing the best match to the gauge rainfall.

To illustrate the magnitude of rainfall detection biases that could be detrimental to modeling and forecasting applications that are dependent on satellite rainfall, we compared the number of rain days predicted by the SRPs and the TAHMO gauge observations during the rainy seasons of East Africa. All SRPs but GSMaP_wGauge predicted much fewer rain days than TAHMO for March–May and October–December seasonal rainfall. TAHMO stations recorded a mean 38 and 42 rain days in the MAM and OND season, respectively, or at least 1 rain day every 3 days on average. CHIRPS and TAMSAT most severely underestimated the number of rain days with an average of 29 and 26 rain days in the MAM season, and 22 and 25 rain days in the OND season, respectively. The best performing SRP was GSMaP_wGauge, which predicted nearly the same number of rain days (43 days) as TAHMO in both seasons.

The impact of error in satellite-derived rainfall estimation can be demonstrated by two recent meteorological events in East Africa: the 2017 MAM and 2019 OND seasons. The 2019 OND seasonal rainfall was termed as one of the wettest
events in recent years by intensity and duration. Total rainfall was estimated to have exceeded more than double the climatological rainfall and continued into the usually dry month of January 2020 (Wainwright et al. 2020). Averaged over all TAHMO stations in East Africa in the OND season, we found that TAHMO recorded 52 days with rain while CHIRPS, TAMSAT, GSMaP, and GSMaP_wGauge predicted 28, 36, 48, and 53 rain days, respectively. From these results, CHIRPS and TAMSAT predicted just over half the number of rain days recorded by the TAHMO stations while the GSMaP products reported a higher number of days more comparable with the ground reference data. Total rainfall biases in the satellite products compounds detection errors. In this example, CHIRPS grossly underestimated the number of rain days but where it detected rainfall, the intensity was overestimated. Thus, many models or applications that rely on CHIRPS would have been informed by an exaggerated amount of rainfall. In comparison, GSMaP_wGauge and TAMSAT underestimated the total rainfall for the same season by only 2 and 1 mm, respectively, meaning that they would have accurately predicted total seasonal rainfall. GSMaP predicted 23% or 119 mm less total rainfall. In this context, GSMaP_wGauge is the most appropriate satellite-derived product since it predicted both the number of rain days and total rainfall over the season with little error.

The 2017 MAM drought was one of the driest seasons since the 2010/11 drought in the Horn of Africa (Han et al. 2022). TAHMO stations recorded average rainfall of 250 mm, which was overestimated by the satellite products. CHIRPS predicted 155 mm more rainfall than the gauge stations whereas both TAMSAT and GSMaP predicted 39 mm more rainfall and GSMaP_wGauge predicted 74 mm more rainfall. Consequently, these satellite products would have estimated a wetter 2017 MAM season than the gauge stations. In fact, CHIRPS estimates would have suggested rainfall amounts indicative of a near normal season, thus failing to capture the occurrence and severity of this drought.

These examples highlight the impact of satellite model errors when predicting rainfall accumulation and distribution, and the importance of reducing errors through better prediction and bias correction algorithms. Such improvements could mitigate the propagation of rainfall errors into weather and climate applications, such as drought and flood forecasting. Studies have shown that the total number of rain days, intensity, and distribution are important factors in water balance models. Even where total accumulations agree between a satellite rainfall product and reference in situ dataset, timing and duration of rainfall events is critical for factors that affect water budgets (Sheffield et al. 2004). Rainfall is an important model forcing parameter for many hydrometeorological applications and input data with large biases should be avoided. The significance of satellite-derived rainfall estimates in gauge-scarce environments like Africa makes the choice between products particularly critical because many models and decisions are reliant on the data.

f. The influence of elevation on satellite rainfall errors

In agreement with Dinku et al. (2011), our results demonstrated that elevation influences the accuracy of satellite-rainfall products. The mean elevation at TAHMO stations was 1317 m in East Africa, 979 m in southern Africa, and 248 m in West Africa. Errors were larger at higher elevations compared to lower elevations at the daily time scale for all products, but especially for CHIRPS and TAMSAT. On average, a 1-km increase in elevation was associated with an increase of 1.3, 1.1, and 0.6 mm in the RMSE for CHIRPS, TAMSAT, and GSMaP_wGauge, respectively, and a decrease of 0.03 mm in RMSE for GSMaP. In most cases, GSMaP_wGauge had superior rainfall detection skill and correlation with in situ gauge data at different elevations than any other satellite-derived product. GSMaP also performed well, revealing the similarity between the two products. GSMaP_wGauge rainfall detection was particularly better than that of the other products at elevations greater than 1000 m. For example, at elevations greater than 2000 m, the POD score of GSMaP_wGauge was 141% and 75% higher than that of CHIRPS and TAMSAT, respectively. Most stations that had elevations that were greater than 2000 m were found in East Africa which could explain the CHIRPS and TAMSAT performance for this elevation range.

The performance of the GSMaP_wGauge product could be attributed to several intrinsic advantages. First, this rainfall product integrates data from multiple sources including passive microwave sensors, infrared sensors, radar sensors and gauge observations which produce reasonably accurate rainfall estimates (Sarmiento et al. 2021). Radar sensors are sensitive to cloud droplets and small ice crystals (Battaglia et al. 2020), which improves rainfall detection. Second, the GSMaP_wGauge algorithm integrates a rainfall classification scheme that is able to discriminate between orographic and nonorographic rainfall (Kubota et al. 2020). The ability to discriminate between the two processes is a challenge for satellite-rainfall estimation algorithms in regions that have heterogeneous terrain. Many satellite-rainfall products show large errors over higher elevation and mountainous regions characterized by these orographic rainfall regimes. Overall, for the period covered by our analysis, the GSMaP products would be ideal for applications that require high rainfall detection skills at the daily time scale in the three subregions such as flood risk analysis, soil moisture estimation, and erosion modeling among others. Relatively large errors in CHIRPS and TAMSAT in areas with elevation greater than 1000 m at the daily time scale make them unsuitable for use where more accurate alternatives exist.

Some of our findings on the agreement between gauge measurements and SRPs differed from results reported by others over the same regions. Many of these studies used gauge data from national meteorological services (Novella and Thiaw 2013; Mashingia et al. 2014; Dinku et al. 2018; Gebremichael et al. 2014; Kimani et al. 2017; Cattani et al. 2016; Thiemig et al. 2012) and we believe that there are significant differences between these sources and data from TAHMO weather stations. These differences are most likely to produce converging results on shorter time scales that would become less prominent with increasing temporal aggregation. A comparison
between gauge data from the national meteorological department and TAHMO in Kenya supports this point (Table 8). Correlation between the closest pair of collocated stations was 0.71. This correlation decreased for longer separation distances, but did not, however, exhibit a consistent pattern. Difference in rainfall measurement methods likely contributed to disagreement between the two sources of gauge data. TAHMO operates automatic rain gauge stations whereas KMD relies on manual rain gauges. Manual gauges are prone to human errors while automatic gauges could experience sensor malfunction. Local environmental conditions at TAHMO and KMD instruments, such as winds, could also affect gauge readings. These factors may explain some of the differences we observed between our results and other studies. We recommend further research that compares TAHMO to an expanded dataset of gauge data from meteorological services the covers Africa and each subregion (e.g., Muita et al. 2021) before combining the two data sources in research and operational climate services.

5. Conclusions

In this study, we found systematic differences and errors in the satellite rainfall products that make them suitable to different applications at different spatial and temporal scales. The GSMaP products produced more accurate estimates of rainfall at the daily time scale and were comparable to CHIRPS and TAMSAT at the pentad and monthly time scales. Moreover, detection of daily rainfall events was high for the GSMaP products and low for CHIRPS and TAMSAT. Large biases in subregional monthly and seasonal rainfall quantities were observed in satellite-derived estimates compared to in situ TAHMO gauge measurements. CHIRPS produced the highest wet bias in East Africa, TAMSAT produced a large dry bias in southern Africa, and the GSMaP satellite-only product exhibited the highest wet bias in West Africa. Elevation influenced error in satellite rainfall estimation, with error increasing with elevation.

Although, correlation between the gauge observations and the satellite rainfall products improved with temporal aggregation, error in other validation statistics increased with data aggregation over coarser time scales. Higher correlation at coarser resolutions could benefit applications that require longer time scale (i.e., monthly to annual scale) to detect long-term trends over large geographical areas. However, as shown in our analysis, an increase in MAE and RMSE with temporal aggregation could cause concerns for other applications. Larger errors in CHIRPS and TAMSAT at the daily time scale compared to the GSMaP rainfall products may be a limiting factor in water-balance modeling that requires good rainfall detection capability at shorter time scales. In sparsely gauged regions like the drylands of Africa, evaporation rates are high and rainfall can be short-duration, high-intensity events that satellite-rainfall products may not accurately estimate. This limitation is especially critical for crop yield modeling, flood risk analysis, soil erosion estimation, and rangeland vegetation condition monitoring. The limitation extends to decisions that rely on information generated directly from satellite-based data products or from modeled outputs that rely on these data products. It is important to investigate the magnitude of errors in data products like soil moisture and runoff from land surface models before decisions are made from those products.

The validation of satellite rainfall products is important for several reasons. First, many hydrometeorological applications require high temporal and spatial resolution data not offered by gauge-only measurements. Characterization of errors in satellite data could benefit these applications by identifying products that perform better over diverse temporal and spatial scales and regions. From our study, rainfall estimation models that combined microwave, infrared, rainfall radar sensors, and gauge measurements produced lower errors than other models that combined fewer data types. Second, validation provides developers of satellite-based weather estimation with useful insights on the performance of rainfall models and potential areas for improvement. Such improvements would involve further refinement of estimates through bias-correction using more in situ measurements. Many arid regions in sub-Saharan Africa lack good coverage of in situ meteorological observations, despite supporting livelihoods that are most vulnerable to changing environmental conditions and climate risks. Consequently, satellite rainfall products may be the only available source of rainfall data in such regions. The introduction of low-cost monitoring networks by TAHMO provides a new source of quality gauge data that can contribute data for improving rainfall models. TAHMO plans to install 20,000 gauge stations in Africa, which will significantly expand the coverage of weather observations that will improve hydrometeorological applications and decision making in the region.

Acknowledgments. The authors acknowledge Nick van de Giesen, Emily C. Adams, Vikalp Mishra, Amy McNally, Kimberly Slinski, Robinson Mugo, and Ryan Langendorf for their invaluable contribution during the analysis and writing of this article. The work described in this publication was made possible through support provided by the National Aeronautics and Space Administration, under the terms of NASA Grant 18-SERVIR18_2-0003. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

Data availability statement. Data are available through the links provided in this manuscript or on request to the corresponding author (dmacharia@rcmrd.org).

REFERENCES


—, and Coauthors, 2021b: Introducing and evaluating the Climate Hazards Center IMERG with Stations (CHIMES):


