SPECIAL Waves to Weather (W2W)



SIMON AGEET,<sup>a,d</sup> ANDREAS H. FINK,<sup>a</sup> MARLON MARANAN,<sup>a</sup> JEREMY E. DIEM,<sup>b</sup> JOEL HARTTER,<sup>c</sup> ANDREW L. SSALI,<sup>d</sup> AND PROSPER AYABAGABO<sup>e</sup>

> <sup>a</sup> Karlsruhe Institute of Technology, Karlsruhe, Germany <sup>b</sup> Department of Geosciences, Georgia State University, Atlanta, Georgia <sup>c</sup> Environmental Studies Program, University of Colorado Boulder, Boulder, Colorado <sup>d</sup> Uganda National Meteorological Authority, Kampala, Uganda <sup>c</sup> Rwanda Meteorological Service, Kigali, Rwanda

> > (Manuscript received 23 July 2021, in final form 16 November 2021)

ABSTRACT: Rain gauge data sparsity over Africa is known to impede the assessments of hydrometeorological risks and of the skill of numerical weather prediction models. Satellite rainfall estimates (SREs) have been used as surrogate fields for a long time and are continuously replaced by more advanced algorithms and new sensors. Using a unique daily rainfall dataset from 36 stations across equatorial East Africa for the period 2001–18, this study performs a multiscale evaluation of gauge-calibrated SREs, namely, IMERG, TMPA, CHIRPS, and MSWEP (v2.2 and v2.8). Skills were assessed from daily to annual time scales, for extreme daily precipitation, and for TMPA and IMERG near-real-time (NRT) products. Results show that 1) the SREs reproduce the annual rainfall pattern and seasonal rainfall cycle well, despite exhibiting biases of up to 9%; 2) IMERG is the best for shorter temporal scales while MSWEPv2.2 and CHIRPS perform best at the monthly and annual time steps, respectively; 3) the performance of all the SREs varies spatially, likely due to an inhomogeneous degree of gauge calibration, with the largest variation seen in MSWEPv2.2; 4) all the SREs miss between 79% (IMERG-NRT) and 98% (CHIRPS) of daily extreme rainfall events recorded by the rain gauges; 5) IMERG-NRT is the best regarding extreme event detection and accuracy; and 6) for return values of extreme rainfall, IMERG, and MSWEPv2.2 have the least errors while CHIRPS and MSWEPv2.8 cannot be recommended. The study also highlights improvements of IMERG over TMPA, the decline in performance of MSWEPv2.8 compared to MSWEPv2.2, and the potential of SREs for flood risk assessment over East Africa.

KEYWORDS: Complex terrain; Remote sensing; Satellite observations

#### 1. Introduction

Most people throughout tropical Africa are highly impacted by extreme weather events. For example, most of the 4.3 million disaster-related internal displacements of people in 2020 in sub-Saharan Africa were due to flooding (IDMC 2021). Unfortunately, these extremes are increasing in frequency and intensity in most parts of the world (Collins et al. 2019). Within tropical Africa, Uganda, for instance, has recorded an increase in the number of hydrometeorological hazards with floods posing the biggest risk to the population (World Bank 2020). Coupled with Uganda's rapidly rising population projected to be almost doubled to about 70.5 million by 2040 (UBOS 2020), the increase in the hydrometeorological hazards makes accurate and high-resolution rainfall data very vital, particularly for trend and extreme precipitation events analyses.

Unfortunately, in most African countries, data from rain gauges (RGs), often considered as reference data, are spatially and temporally sparse, temporally inconsistent, and in some cases of low quality (Kizza et al. 2009; Maidment et al. 2013; Diem et al. 2014; Monsieurs et al. 2018). In fact, the station network coverage across Africa has been shrinking over the years (Washington et al. 2006; Asadullah et al. 2008; Dinku 2019). Therefore, accurately analyzing and predicting rainfall trends and variability over the study area is challenging. Satellite rainfall estimates (SREs) have filled the spatiotemporal data gaps. In most cases, the SREs are derived from radiances in the visible, infrared (IR), and microwave spectra measured by satellites. Singularly or in combination, these retrievals provide high-resolution rainfall estimates, with a prime example being the Integrated Multisatellite Retrieval for Global Precipitation Measurement (GPM) (IMERG) V06B dataset (Huffman et al. 2020a). However, SREs are not without shortcomings. Despite having a wide field of view and high temporal availability, IR-based techniques rely on a

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-21-0145.s1.

<sup>&</sup>lt;sup>3</sup> Denotes content that is immediately available upon publication as open access.

Corresponding author: Simon Ageet, simon.ageet@kit.edu

cloud top temperature-precipitation relationship, which is often unsuitable for nonconvective precipitation (Kidd and Huffman 2011) or nonprecipitating cold cirrus shields (Young et al. 2014). Microwave-based satellites, in turn, have been found to struggle with retrieving rain from warm clouds (Dinku et al. 2010b; Monsieurs et al. 2018; Maranan et al. 2020) and to overestimate rainfall in places where convective rainfall dominates (Tian et al. 2009). However, passive microwave (PMW)-based approaches are considered to be physically more direct than IR-based estimates (Kidd and Huffman 2011; Bitew and Gebremichael 2011). For example, over land, the background microwave signal from Earth is scattered by precipitation-sized hydrometeors, especially rainfall-inducing ice crystals, leading to a decrease of the PMW brightness temperatures (Kidd and Huffman 2011). Additionally, the skill of SREs has been shown to be dependent on other factors. For example, Dinku et al. (2010b), Diem et al. (2014) and Monsieurs et al. (2018) noted that the SREs performed poorly over complex topography, namely, the Ethiopian Highlands, the Albertine and East African Rifts, and the Rwenzori Mountains. The skill of SREs also likely depends on the rainfall formation processes (McCollum et al. 2000; Maranan et al. 2020) and gauge calibration (Awange et al. 2016). Therefore, SREs should be meticulously validated in order to ascertain their suitability for a given area.

Several validation studies have been conducted in and around Uganda (e.g., Asadullah et al. 2008; Diem et al. 2014; Ashouri et al. 2015; Maidment et al. 2017; Dinku et al. 2018; Monsieurs et al. 2018; Camberlin et al. 2019). Most of these studies found the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) (Huffman et al. 2007) and Climate Hazards Group Infrared Precipitation with Stations version 2 (CHIRPS2) (Funk et al. 2015) to be among the best performing products. For instance, using 31-yr-averaged (1960-90) monthly rainfall data over Uganda, Asadullah et al. (2008) showed that TMPA 3B42 had the least error in mean annual rainfall and also captured the mean seasonal and climatological spatial pattern of rainfall better than Climate Prediction Center morphing technique (CMORPH) (Joyce et al. 2004), Tropical Applications of Meteorology using Satellite and Ground-Based Observations (TAMSAT) (Maidment et al. 2017), Rainfall Estimation Algorithm version 2 (RFE2) (Love et al. 2004), and Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) (Ashouri et al. 2015). Using gridded monthly RG data between 1960 and 2004, Kizza et al. (2012) also showed that TMPA correlates to the observed monthly rainfall variability at the coast of Lake Victoria better than PERSIANN. Using 10 years (2001-10) of daily rainfall data, Diem et al. (2014) showed that TMPA 3B42 outperformed Africa Rainfall Climatology version 2 (ARC2) (Novella and Thiaw 2013) and RFE2 over Western Uganda. Camberlin et al. (2019) evaluated seven satellite products over central Africa which included part of western Uganda and found that TMPA performed best at daily time scale while at interannual time scale, CHIRPS and TMPA performed best. CHIRPS has been found to particularly perform well at longer time scales. For example, Dinku et al. (2018) and Diem et al. (2019) both

showed CHIRPS performed better than ARC2 and TAM-SAT at dekadal to seasonal time scales over eastern Africa and western Uganda, respectively.

IMERG and the Multi-Source Weighted-Ensemble Precipitation (MSWEP) (Beck et al. 2019) are relatively new and only few validation studies have been done in the region for these datasets. The transition from TMPA to its successor IMERG has led to improvements in rainfall monitoring (Dezfuli et al. 2017a,b). Dezfuli et al. (2017a) validated IMERG and TMPA over West and East Africa for the respective rainy seasons of 2015. They showed that IMERG was closer to the mean of the RGs and better represents extreme events in both regions owing to its higher temporal resolution. However, in a follow-up study, Dezfuli et al. (2017b) showed that the gain/loss in performance of IMERG over TMPA are spatially heterogenous based on the patterns of statistical metrics, e.g., the Heidke skill score (HSS). The few studies evaluating MSWEPv2.2 have so far shown promising results (Satgé et al. 2019; Beck et al. 2019). With respect to the Kling-Gupta efficiency (KGE), Satgé et al. (2019) showed that MSWEPv2.2 performed best among 23 rainfall products over West Africa based on 3 years (2000-03) of daily rainfall data. Over contiguous United States, Beck et al. (2019) showed that MSWEPv2.2 had the best median KGE compared to four other rainfall products, namely, MSWEPv1, CMORPH, European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) (Dee et al. 2011), and Modern-Era Retrospective Analysis for Research and Application, version 2 (MERRA-2) (Reichle et al. 2017).

The aim of the study is to assess the skill of IMERG, TMPA, CHIRPS, and MSWEP for Uganda and bordering parts of Kenya, Tanzania, Rwanda, and South Sudan at seasonal, monthly, and daily time scales. Based on the earlier studies, we conclude that the performance of SREs is variable among products depending on the region and/or spatiotemporal scales. Most of the studies used monthly data, were not Uganda-wide, used few stations and, aside from Monsieurs et al. (2018), did not consider extremes rainfall events. We use a unique network of daily, quality-controlled RG data from 2001 to 2018 for validation. A large number of station data were not ingested into the Global Telecommunication System (GTS) and therefore not used to calibrate the SREs, thus they can serve as independent references. CHIRPS and TMPA were chosen because they are recommended for extreme event applications (Le Coz and van de Giesen 2020) and perform satisfactorily over the study area. As mentioned earlier, IMERG has been found to be superior to TMPA and belongs to the most recent precipitation products alongside with MSWEP. An additional motivation to include both TMPA and IMERG is to investigate the improvements gained with IMERG over TMPA. At the time of writing this manuscript, MSWEP transitioned from v2.2 into v2.8, which allowed a direct evaluation of the impact of version changes. We also include the near-real-time (NRT) products of IMERG and TMPA since they cater for application requiring short latency data, e.g., floods assessment.

The rest of paper is structured as follows: section 2 gives a description of the study area, as well as the SREs and RG datasets. The methods used are explained in section 3. The



FIG. 1. The study area with location and names of the rain gauge stations used in the study shown with blue circles. Data from stations with black cross in the center are not transmitted to the Global Telecommunication System (GTS). The major lakes, rivers, and mountains, viz., Rwenzori in the west and Elgon in the east, are also indicated. The color shading shows the elevation from Global Land One-km Base Elevation Project (GLOBE; Hastings et al. 1999). The map inset highlights the location of Uganda within Africa (red box).

results are presented and explained in section 4. Finally, in section 5, the major conclusions are summarized and put in perspective of other studies.

#### 2. Study area and datasets description

# a. Study area

The study focuses on Uganda and the surrounding areas (Fig. 1). The topography is heterogeneous with elevation ranging from 375 to 5109 m above mean sea level at Lake Turkana in the northeast and the peak of Mt. Rwenzori in the west, respectively. It is also composed of a number of large

inland lakes, mountains, parts of the Great East African Rift Valley, and rivers. Most parts of the region exhibit two distinct rainfall seasons [March–May (MAM) and September– November (SON)], locally referred to as the "long rain season" and "short rain season" with peaks in April and October, respectively (Nicholson 2017; Seregina et al. 2019). Over western Uganda, a transitional region between the eastern and western equatorial Africa, the MAM and SON seasons form the short and long rains, respectively (Diem et al. 2019). The northern part of the study area exhibits a unimodal rainfall peak in May–August (Phillips and McIntyre 2000; Seregina et al. 2019). As discussed in Nicholson (2017), the climate of the region is controlled by three classes of factors;

Dataset	Full name	Native spatial resolution	Temporal resolution	Latency	Temporal coverage	Reference	
TMPA 3B42 v7	Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis	0.25°	3-hourly	2 months	January 1998–present (2001–18)	Huffman et al. (2007); Huffman and Bolvin (2018)	
TMPA-RT 3B42RT v7	Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis	0.25°	3-hourly	8 h	March 2000–present (2001–2018)	Huffman et al. (2007); Huffman and Bolvin (2018)	
IMERG	Integrated Multisatellite Retrieval for GPM (Global Precipitation Measurement)	0.1° (0.25°)	Half-hourly	3.5 months	June 2000-present (2001-18)	Huffman et al. (2020b)	
IMERG-E	Integrated Multisatellite Retrieval for GPM (Global Precipitation Measurement)	0.1° (0.25°)	Half-hourly	4 h	June 2000-present (2001-18)	Huffman et al. (2020c)	
CHIRPS	Climate Hazard Infrared Precipitation with Stations	0.05°/0.25°	3-hourly	2 months	January 1981–present (2001–18)	Funk et al. (2015)	
MSWEP v2.2	Multi-Source Weighted- Ensemble Precipitation V2.2	0.1° (0.25°)	3-hourly	Irregular	January 1979–October 2017 (2001–16)	Beck et al. (2019)	
MSWEPv2.8	Multi-Source Weighted- Ensemble Precipitation V2.8	0.1° (0.25°)	3-hourly	Irregular	January 1979– December 2020 (2001–18)	Beck et al. 2019	

TABLE 1. Summary of datasets used in the study. The period and regridded resolution used in this study is given in parentheses in the "temporal coverage" and "native spatial resolution" columns, respectively.

regional forcing features (e.g., Walker circulation, El Niño–Southern Oscillation, Indian Ocean dipole, and Madden–Julian oscillation), regional circulation features (e.g., tropical easterly jet, low-level westerlies, and monsoons), and local geographic factors (e.g., Lake Victoria, East African highlands). These factors inevitably lead to variability in rainfall and coupled with the heterogenous topography present a challenge to satellite rainfall retrievals (e.g., Monsieurs et al. 2018; Le Coz and van de Giesen 2020).

# b. Datasets

# 1) SATELLITE PRODUCTS

We included eight satellite datasets in this validation study (Table 1). The temporal resolutions of the datasets vary from 30 min to daily, while the spatial resolutions vary from  $0.1^{\circ}$  to  $0.25^{\circ}$  latitude–longitude grid boxes. They are described in the following in more detail.

TMPA V07 (Huffman et al. 2007; Huffman and Bolvin 2018) is a multisatellite precipitation product with latencies of about 8 h and about 2 months for the "real time" and "final" runs, respectively. TMPA consists of PMW precipitation estimates, IR precipitation estimates, and the Global Precipitation Climatology Project (GPCP) monthly precipitation (Huffman et al. 1997). First, PMW estimates are calculated using the Goddard Profiling (GPROF) algorithm (Kummerow et al. 1996, 2001). Then all available PMW estimates are combined and completed with MW-calibrated IR data in case of

existing gaps. Finally, monthly multisatellite (MS) accumulations are combined with Global Precipitation Climatology Centre (GPCC) monthly gauge analysis to create a satellite–gauge (SG) monthly product (TMPA 3B43). A SG/MS ratio is then calculated for each 0.25° and used to calibrate each 3-hourly field of the month, producing the final 3-hourly product (TMPA 3B42).

IMERG V06B (Huffman et al. 2020a) is a MS precipitation product with latencies of about 4 h, 14 h, and 3.5 months for the Early, Late, and Final runs, respectively. IMERG builds on the TRMM legacy and uses similar data inputs as TMPA, with the GPM core observatory replacing the TRMM counterpart. The PMW estimates are calculated using a more recent version of the GPROF algorithm. A seasonal GPCP calibration is applied to the PMW estimates to yield 30-min  $0.1^{\circ} \times 0.1^{\circ}$  fields. These fields are then spatiotemporally morphed to fill the gaps for areas without PMW overpasses. This step results in the MS, half-hourly IMERG Early (Huffman et al. 2020b) and IMERG Late products (IMERG-E and IMERG-L henceforth). IMERG-E contains only forward morphing while IMERG-L has both forward and backward morphing. Monthly satellite-gauge estimates are created by summing the half-hourly estimates for the month and calibrating with GPCP monthly precipitation analysis. Finally, gauge calibration of the half-hourly estimates is exercised in a similar fashion to TMPA to create the Final version of IMERG (IMERG-F hereinafter; Huffman et al. 2020c).

CHIRPS v2 (CHIRPS hereinafter) is a quasi-global dataset providing daily, pentadal, and monthly estimates (Funk et al. 2015). CHIRPS is based on a global monthly precipitation climatology (CHPclim), thermal infrared (TIR) cold cloud duration (CCD), and daily and monthly RG data. TIR CCD data created at a constant CCD temperature threshold of 235 K are locally calibrated using TMPA 3B42 pentadal precipitation. The pentadal estimates are then multiplied by their corresponding CHPclim estimate to produce Climate Hazard Infrared Precipitation (CHIRP). CHIRP is merged with gauge data to produce CHIRPS on pentadal and monthly time steps. The pentadal CHIRPS is disaggregated into daily CHIRPS precipitation using daily CCD data.

MSWEP is a global precipitation dataset available from 1979 to October 2017 (MSWEPv2.2, Beck et al. 2019) and until December 2020 in a recent update (MSWEPv2.8; Beck et al. 2021a). The dataset is created through weighted merging of gauge-, satellite-, and reanalysis-based precipitation products. The merging weights are determined by assessing the performance of the input datasets individually against quality-controlled gauge precipitation. The weighted merging results into a 3-hourly reference precipitation dataset which is then calibrated using daily and monthly gauge precipitation to create MSWEPv2.2. Although the algorithm did not change substantially, changes in the production of MSWEPv2.8 include a reduction to two underlying datasets from the five used in MSWEPv2.2 and a reduction in the amount of gauge data used (Beck et al. 2021b).

#### 2) RAIN GAUGE DATA

Besides quality issues, not nearly enough data are ingested into the GTS in our study area. For this study, 24 of the 36 stations transmitted rainfall data to the GTS, and only 6 of these 24 stations had more than 50% of available daily data for the study period (Fig. 2a). Benefitting from collaborations, the Karlsruhe African Surface Station Database (KASS-D; Vogel et al. 2018) has many observations not reported to the GTS. Figure 2b shows the full station ensemble and its data availability in KASS-D within the study region. Hence, 12 non-GTS stations are available for the present analysis and even the GTS reporting stations have much more complete data in KASS-D.

The quality control of the gauge dataset was carried out following two steps. First, the timestamp was checked using a methodology similar to Beck et al. (2019) to ensure a consistent treatment of the timestamp across all datasets. The datasets were then analyzed for suspicious outliers and zero values using a combination of techniques, e.g., using the "letter-value plot" method in the python package "seaborn" and accumulation curves.

# 3. Methods

#### a. Point-to-pixel comparison

A point-to-pixel approach is performed by comparing point gauge data to the closest satellite pixel value (e.g., Monsieurs et al. 2018; Maranan et al. 2020). We recognize the

shortcomings of this approach given that gauges are highly localized and may not be representative of a gridbox coverage. Engel et al. (2017) and Monsieurs et al. (2018) demonstrate the shortcoming using grids with multiple RGs. We have not carried out a similar analysis given our dataset has no grid boxes with more than one RG. We instead retain the IMERG native resolution  $(0.1^{\circ})$  and compare it with the remapped 0.25° resolution to assess changes in the performance of SREs due to changes in resolution. While interpolation of irregular station data to the SRE grids is recommended for satellite validation studies (Maidment et al. 2013), we did not use this approach due to the relatively low density of rain gauges and the complex topography. All the SREs are regridded to the 0.25° resolution of TMPA using a first-order conservative remapping (Jones 1999) to allow for a fair comparison of the SREs. Since gauge data are missing at some stations for some days, we only used satellite data when gauge data were available at a given station and day.

# b. Validation metrics

This study employs some commonly used standard validation approaches (e.g., Ebert 2007; Wilks 2011). First, quantilequantile (QQ) plots were used to compare the distribution of rain rates. The ability of SREs to detect rainy days, defined here as days with rainfall total of more than 0.2 mm, was assessed based on the contingency table (Table 2) using dichotomous metrics, namely, probability of detection (POD), probability of false alarm (POFA), bias in detection (BID), and Heidke skill score (HSS). For the analysis of extremes, we applied a threshold (95th percentile of the subset of rainy days in the RGs) to differentiate between an "extreme" and "non-extreme" rainfall day. Thus, a hit occurs when both the RG and SRE have a rainfall amount greater than the threshold. The second group of metrics includes the Pearson's correlation coefficient (r), mean error (ME), percent bias (PB), mean absolute error (MAE), and root-mean-square error (RMSE) to assess the SREs' accuracy for rain rates using the subset of hits. Additionally, the Nash-Sutcliffe coefficient of efficiency E (Nash and Sutcliffe 1970; Legates and McCabe 1999) was used to assess the skill against climatology. The metrics are summarized in Table 3.

#### c. Spatiotemporal analysis

The validation was performed for individual stations to fully use the available gauge data and for daily, pentadal, dekadal, monthly, and annual aggregations. To reduce randomness in some of the analyses, we reorganized the stations into groups. We therefore used the nonhierarchical *K*-means clustering algorithm (Pedregosa et al. 2011) to partition the stations into groups. This approach aims at placing the stations in clusters such that the intracluster variance is minimized (Hartigan and Wong 1979). As noted by Cattani et al. (2016), *K*-means clustering is prone to subjectivity since the maximum number of clusters has to be determined a priori. We used the "elbow" method (Thorndike 1953; Zhang et al. 2016) which is based on the intracluster sum of squared differences to determine the optimum number of clusters. For each



FIG. 2. (a) Temporal availability of the gauge data used in the study with stations reporting to the GTS and (b) the full dataset as obtained from KASS-D for the period 2000–18. The GTS stations subset, identifiable by their five-digit only WMO station number, has only 24 stations compared to full list of stations obtained from KASS-D. The blue strips shading shows daily data availability.

TABLE 2. The 2  $\times$  2 contingency table for comparing the rainy days in gauge and satellite estimate. A day is considered as rainy if 0.2 mm of rainfall or more is recorded. Note that for when a threshold is set, a wet day occurs when an amount greater than the set threshold was recorded.

	Gauge $\geq 0.2 \text{ mm}$	Gauge < 0.2 mm
Satellite $\ge 0.2 \text{ mm}$	Hit (H)	False alarm (F)
Satellite $< 0.2 \text{ mm}$	Miss (M)	Correct rejection (R)

station, a 30-day running mean was applied over their respective Julian day of year-based climatology of rainfall. The resulting time series were then normalized with the maximum value of the 30-day running mean at the respective stations to ensure that stations with similar seasonal cycles are grouped together regardless of the amount of precipitation received. For each cluster, the mean monthly rainfall of the satellite products and RGs were computed by taking the average of the stations in that cluster. Additionally, to investigate the influence of gauge calibration, we subset the data into two groups, namely, GTS (stations with the highest percentage of data reported to the GTS) and non-GTS (stations whose data is not reported to the GTS). To allow for a comparison of stations subsets with different climatologies, we normalized the error metrics (ME, MAE, and RMSE) with the standard deviation of the concatenated RGs time series of the respective subsets.

# d. Extreme rainfall events analysis

Since hydrometeorological hazards are largely a result of extreme daily rainfall events, the ability of satellites to capture these events is analyzed. We define an "extreme event" as daily rainfall totals exceeding the 95th percentile on rainy days in the RG dataset. The extreme events thresholds were computed at individual stations and the corresponding values for the SREs were obtained from the closest pixels. We also analyzed SREs' suitability in simulating the return values of extreme event. Extreme value models are used to study the behavior of the tail of a distribution (Bommier 2014), enabling the simulation of extreme values and their return periods. Similar to Engel et al. (2017), we use the peak-overthreshold (POT) method to fit the generalized Pareto distribution (GPD; Lemos et al. 2020) to the RG and SREs daily rainfall. The GPD was fitted to the subsets of extreme events (i.e., >95th percentile) in the RG and SREs datasets. The extremes in the SREs were obtained in a similar way to the

TABLE 3. Summary of the validation metrics used; *H*, *F*, *M*, and *R* denote hits, false alarms, misses, and correct rejections, respectively. The  $x_i$ ,  $y_i$ ,  $\overline{x}$ , and  $\overline{y}$  terms represent the rainfall totals at the gauge for a given time, SRE rainfall totals for a given time, the mean rainfall totals at gauge, and the mean rainfall totals of SRE, respectively. The range and best possible score of each statistic is added in the first column in square brackets and bold numbers, respectively.

Statistic	Formula	Description
POD, [0, 1], <b>1</b>	$POD = \frac{H}{H+M}$	Ability of the satellite to correctly identify rainy days
POFA, [0, 1], <b>0</b>	$POFA = \frac{F}{H+F}$	Proportion of rainy days in the SREs that were not observed in RGs
BID, $[-\infty, +\infty]$ , <b>1</b>	$BID = \frac{H+F}{H+M}$	Assesses whether the SRE overestimates or underestimates rainy day frequency
HSS, [−∞, 1], <b>1</b>	$HSS = \frac{2(HR - FM)}{(H + M)(M + R) + (H + F)(F + R)}$	Assesses the skill of SREs products compared to random chance
r, [-1, 1], <b>1</b>	$r = \frac{\sum_{i}^{n} (x_i - \overline{x}) \sum_{i}^{n} (y_i - \overline{y})}{\sqrt{\sum_{i}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i}^{n} (y_i - \overline{y})^2}}$	Assesses the covariance of the gauge data with that of the SRE
ME (mm), $[-\infty, +\infty]$ , <b>0</b>	$\mathbf{ME} = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i),$	Measures the bias and its direction (underestimation or overestimation) by the SREs
PB (%), $[-\infty, +\infty]$ , <b>0</b>	$PB = 100 \frac{\frac{1}{n} \sum_{i}^{n} (y_i - x_i)}{\overline{x}}$	Assesses the tendency of the SREs to underestimate or overestimate rain rates relative to the mean of the RGs
MAE (mm), $[0, +\infty]$ , <b>0</b>	$MAE = \frac{1}{n} \sum_{i=1}^{n}  (y_i - x_i) $	Measures the bias of the SREs regardless of direction
RMSE (mm), $[0, +\infty]$ , <b>0</b>	$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n} (y_i - x_i)^2}$	Measures the bias of the SREs but assigning more weight to outliers
<i>E</i> , [−∞, 1], <b>1</b>	$E = 1 - \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$	Assesses the skill of SREs relative to climatology



FIG. 3. Mean annual rainfall totals for six satellite products (shaded) with mean annual rainfall totals of the RGs annotated in the circles. The white contours are isohyets of mean annual rainfall, labeled in black. The dashed black lines are country boundaries while the solid black lines demarcate inland lakes. The values on the top-left corners are the mean annual rainfall and the standard deviation taken over the entire domain for the respective satellite product. All the products are regridded to a 0.25° common grid.

TABLE 4. Scores for annual rainfall totals (2001–18, hence, total number of observations, N = 648) for all the satellite products, except for MSWEPv2.2 (2001–16, hence, total number of observations, N = 576). The best performing final product for each metric is given in bold font.

Statistic/product	IMERG-F_0.10°	IMERG-F_0.25°	IMERG-E	TMPA	TMPA-RT	CHIRPS	MSWEPv2.2	MSWEPv2.8
r	0.77	0.78	0.75	0.75	0.74	0.84	0.84	0.82
Ε	0.57	0.59	-0.36	0.52	0.23	0.70	0.65	0.63
$ME (mm yr^{-1})$	57	57	349	-60	-214	-10	-99	-91
PB (%)	5	5	28	-5	-19	-1	-8	-7
MAE (mm $yr^{-1}$ )	208	200	402	218	300	171	169	193
RMSE (mm yr <sup>-1</sup> )	294	287	520	309	392	243	264	272

RGs extremes. To make the stations and all the rainfall products comparable, we normalized the modeled return values with the RG-modeled return values at the stations, then averaged over all stations for each dataset. The normalized return values of the RG data were taken as the reference for evaluating the SREs. The performance of SREs was considered "satisfactory" if the return value modeled using the SRE fall within the standard deviation of the RG's return value.

#### 4. Results

While we present all the eight SREs products on the same figures and tables, the comparison between the early products (IMERG-E and TMPA-RT) is discussed separately, as is the comparison between the early and the respective final versions (IMERG-F and TMPA). Furthermore, a separate analysis of the influence of different spatial resolutions on the skill is conducted using IMERG-F at 0.25° and at its native resolution 0.1°.

#### a. Annual rainfall

Based on RGs, the study area receives an annual average of 1217 mm with high spatial variability. The wettest region is the western part and shorelines of Lake Victoria (Fig. 3). The Ugandan Ssese Islands are known to have more than 2000 mm  $yr^{-1}$  (Flohn and Fraedrich 1966) with a short-term measurement on the Tanzanian Nabuyongo Island in the middle of the lake indicating the possibility of 3000 mm  $yr^{-1}$  (Flohn and Burkhardt 1985). No rainfall data over the lake are available in the study period, only Bukoba at the western shore with 1924 mm  $yr^{-1}$  indicates the wetness of this area. Other wet regions are the Rwenzori and Elgon Mountains in western and eastern Uganda, respectively. While no station is available in the Rwenzori mountain ranges for this study, amounts of 2000-3000 mm yr<sup>-1</sup> are reported (Eggermont et al. 2009). Buginyanya at Mt. Elgon at an elevation of 1845 m receives an annual total of 2278 mm. From Mt. Elgon, a wetter region with annual totals of more than 1200 mm stretches northwestward across the Lake Kyoga region in central Uganda to northwestern Uganda west of the Albert Nile (Fig. 3; Basalirwa 1995). The northeastern and southwestern parts of the study area are located along the East African Rift Valley, an area locally known as the "cattle corridor," which is generally associated with drier semiarid conditions. Here, rain shadow effects are the likely causes of the low rainfall totals in low-elevation areas (e.g., Diem et al. 2014).

The satellite products generally reproduce the spatial pattern of annual rainfall, with CHIRPS showing the best performance for most of the scores (Fig. 3 and Table 4). The correlations between the RGs and satellites' mean annual rainfall totals are high, with CHIRPS and MSWEPv2.2 having the highest correlation of 0.84, followed by MSWEPv2.8 (0.82), IMERG-F (0.78), and then TMPA (0.75, Table 4). All the satellite products are skillful in depicting annual rainfall totals with all of them having an *E* value greater than 0.5. Although the patterns are relatively well reproduced, biases do exist. For example, when averaged over all stations, IMERG-F overestimate the annual rainfall by about 5%, while TMPA, CHIRPS, MSWEPv2.2, and MSWEPv2.8 all underestimate annual rainfall by about 5%, 1%, 8%, and 7%, respectively.

Two salient features in Fig. 3 are worthy of mentioning; the very high degree of wetness of the IMERG products over Lake Victoria with values over  $3600 \text{ mm yr}^{-1}$  (Figs. 3a,b), corroborating the findings of Nicholson et al. (2021), and the sharp horizontal discontinuity between the northern and southern parts of the lake in TMPA products (Figs. 3d,e). While IMERG remedies the discontinuity likely with an improved land seas mask for the gauge calibration (see Fig. S1 in the online supplemental material), the overestimation with respect to RGs and the other SREs might be related to a larger weight given to PMW retrievals in IMERG that are known to overestimate rainfall in areas with deep convection (Sungmin and Kirstetter 2018; Nicholson et al. 2021). IMERG-E and TMPA-RT also show quite high correlation values of 0.75 and 0.74, respectively, although the E value of IMERG-E suggests a worse performance than climatology. Additionally, IMERG-E overestimates annual rainfall by 28% (349 mm yr<sup>-1</sup>), while on the other hand TMPA-RT underestimates annual rainfall by 19% (214 mm yr<sup>-1</sup>). The large difference between IMERG-E and TMPA-RT is likely due to a very significant difference in the algorithm between IMERG and TMPA. These larger errors in the early products compared to their respective final products underscore the importance of gauge calibration. Based on the domain-averaged annual rainfall of the SREs, MSWEPv2.8 retrieves the least annual rainfall of 997 mm with a standard deviation of 310 mm. The mean annual rainfall is substantially less than that retrieved by MSWEPv2.2 (1125 mm). Moreover, the r



FIG. 4. (a) Result of the *K*-means clustering. Black-, green-, purple- and orange-colored stations indicate the Northern cluster (N-cluster), Central cluster (C-cluster), Southwestern cluster (SW-cluster), and Lake Victoria Basin cluster (LVB-cluster), respectively. Some stations may be relatively far from their cluster, e.g., the three SW-cluster stations located at the shores of Lake Victoria. Manual investigation showed the seasonal cycles of these stations match the SW cluster. (b)–(e) The mean seasonal cycles of rain gauges and satellite products in the four clusters for the period 2001–18 (IMERG-F\_0.1°, IMERG-F\_0.25°, IMERG-E, TMPA, TMPA-RT, CHIRPS, and MSWEPv2.8) and 2001–16 (MSWEPv2.2).

and *E* values show a decline in performance from MSWEPv2. 2 to MSWEPv2.8 (Table 4).

# b. Monthly rainfall

Based on the elbow method (cf. section 2c), the optimum number of clusters to be used in the K-means clustering was found to be four. The names of the clusters given in Fig. 4

were assigned based on the relative location of the majority of the stations forming the clusters [i.e., Northern (N-) cluster, Central (C-) cluster, Southwestern (SW-) cluster, and Lake Victoria Basin (LVB-) cluster]. The four clusters depict different seasonal cycles (Figs. 4b-e). The N-cluster depicts a unimodal rainfall regime ranging from April to October with two small peaks in May and August. These peaks are partly RMSE (mm month

49.6

Statistic/product	IMERG-F_0.10°	IMERG-F_0.25°	IMERG-E	TMPA	TMPA-RT	CHIRPS	MSWEPv2.2	MSWEPv2.8
r	0.79	0.80	0.74	0.77	0.72	0.79	0.81	0.79
Ε	0.63	0.64	0.24	0.58	0.45	0.62	0.65	0.62
ME (mm month <sup><math>-1</math></sup> )	4.8	4.8	30.3	-5.3	-20.4	-1.1	-8.8	-8.1
PB (%)	4	4	28	-5	-20	-1	-8	-8
MAE (mm month $^{-1}$ )	34.8	34.4	50.2	36.2	42.0	35.1	31.0	34.9

70.1

52.0

59.6

TABLE 5. As in Table 4, but for monthly rainfall (N = 7776 for all SREs, except MSWEPv2.2 with N = 6912).

48.6

associated with the migration of the tropical rain belt during its meridional transitions (Seregina et al. 2019). The remaining three clusters have bimodal rainfall regimes. In the C-clusters the wet seasons run from March to June and August to November, with peaks in April and October. In the SW- and LVB-cluster, the first wet season occur from March to May, peaking in April. The second rainy seasons are September-November and October-December, for the SWand LVB-cluster, respectively. The high overestimation of IMERG-E is clearly seen in all the clusters.

49.2

Generally, all the satellite products replicate the seasonal cycle well (Figs. 4b-e, Fig. S2 in the supplemental material, and Table 5). The agreement between the RGs and the satellite products is evident from the high correlation values of between 0.77 (TMPA) and 0.81 (MSWEPv2.2). All the satellite products perform better than the reference RG-based climatology (i.e., E > 0; 0.58–0.65). Considering the final products only, IMERG-F overestimates monthly rainfall by about 4%, while MSWEPv2.2, MSWEPv2.8, TMPA, and CHIRPS tend to underestimate rainfall by about 8%, 8%, 5%, and 1%, respectively. The results for this time scale also point to a decline in performance of MSWEPv2.8 compared to MSWEPv2.2. The NRT products also reproduce the seasonality of rainfall in the study areas, but the errors are larger compared to their respective final versions. IMERG-E (TMPA-RT) overestimates (underestimates) monthly rainfall

49.

47.3



FIG. 5. QQ plots of daily gauge rain rates against daily satellite rain rates in the four clusters. The solid diagonal line is the 1:1 line (ideal fit). Note that the scales are logarithmic and only hits are considered (days when both SRE and RG > 0.2 mm).

TABLE 6. Summary of skill score on daily, pentadal, and dekadal time steps. We do not calculate the POD, POFA, BID, and HSS for pentadal and dekadal accumulations as these metrics are less meaningful at these time scales compared to a daily time scale. Similar to Tables 4 and 5, the bold number denotes the best scores among the final products.

Time step	Satellite product	POD	POFA	BID	HSS	r	Ε	ME (mm day <sup>-1</sup> )	PB (%)	MAE (mm day <sup>-1</sup> )	RMSE (mm day <sup>-1</sup> )
Daily	IMERG-F_0.1°	0.88	0.40	1.47	0.47	0.41	-0.04	-1.6	-16	7.6	12.4
2	IMERG-F_0.25°	0.93	0.44	1.65	0.43	0.43	0.06	-1.7	-18	7.3	11.8
	IMERG-E	0.94	0.45	1.72	0.40	0.41	-0.05	-0.1	-1	8.0	12.5
	TMPA	0.76	0.36	1.19	0.47	0.35	-0.13	-2.3	-24	8.1	12.9
	TMPA-RT	0.75	0.39	1.19	0.47	0.35	-0.07	-3.3	-35	7.8	12.6
	CHIRPS	0.77	0.41	1.33	0.41	0.22	-0.14	-3.1	-33	8.2	13.0
	MSWEPv2.2	0.88	0.40	1.47	0.47	0.43	0.02	-2.6	-27	7.0	12.0
	MSWEPv2.8	0.97	0.50	1.97	0.29	0.37	0.07	-3.3	-35	7.0	11.8
Pentadal	IMERG-F_0.1°					0.58	0.25	0.6	1	14.1	20.6
	IMERG-F_0.25°					0.60	0.31	0.1	1	13.6	19.7
	IMERG-E					0.56	0.04	5.2	23	16.2	23.3
	TMPA					0.53	0.16	-2.0	-8	14.9	21.8
	TMPA-RT					0.52	0.15	-5.0	-22	14.6	21.9
	CHIRPS					0.50	0.19	-1.4	-6	14.8	21.3
	MSWEPv2.2					0.60	0.29	-2.4	-11	12.8	20.0
	MSWEPv2.8					0.58	0.31	-2.8	-12	13.3	19.3
Dekadal	IMERG-F_0.1°					0.68	0.42	1.2	3	19.6	27.8
	IMERG-F_0.25°					0.69	0.46	1.2	3	19.0	26.9
	IMERG-E					0.64	0.13	10.4	26	24.0	34.0
	TMPA					0.63	0.34	-2.5	-6	20.7	29.6
	TMPA-RT					0.61	0.30	-7.9	-20	21.0	30.7
	CHIRPS					0.63	0.38	-1.1	-3	20.4	28.9
	MSWEPv2.2					0.69	0.45	-3.6	-9	17.8	27.0
	MSWEPv2.8					0.67	0.45	-3.5	-9	18.8	27.1

by about 28% (20%). Overall, MSWEPv2.2 outperforms all the other products at this time scale.

## c. Daily, pentadal, and dekadal rainfall

All the SREs, except MSWEPv2.2 slightly overestimate the low rain rates whereas high rainfall rates are largely underestimated. This is evident from the QQ plots in Fig. 5. MSWEPv2.8 visibly underestimates the higher rainfall rates compared to MSWEP v2.2 but overestimates the low rain rates. Although MSWEPv2.2 is below the diagonal for almost all the percentiles, it has a better fit to the RG data. The distribution in MSWEPv2.8 is likely strongly influenced by ERA5 (see Fig. S3 in the supplemental material), given that their patterns are similar.

The ability of the final versions of the satellite products to correctly capture a rainy day as reported in the gauges is generally good. This is illustrated by high POD values ranging from 0.76 in TMPA to 0.97 in MSWEPv2.8 (Table 6). However, all the products are prone to false alarms and tend to overestimate rainfall frequency (seen in Figs. 6b,c and Table 6). This corroborates results in Fig. 5, especially the overestimation in occurrence frequency at the low rainfall rates. The false alarms negatively influence the HSS, e.g., MSWEPv2.8 with the highest POFA has the lowest HSS, and the reverse is true for TMPA. However, the skill of SREs in detecting rainy days is better than random chance for all the products (HSS > 0). The native resolution of IMERG (IMERG-F\_0.1°), as expected, has a higher HSS of 0.47 compared to the regridded

 $0.25^{\circ}$  version (IMERG-F\_0.25°) with a HSS of 0.43. This difference in skill is about 9% and may be attributed to the reduction in POFA.

Considering daily rain rates, the *r* values suggest a low to average agreement between the satellites and gauge data (0.22 in CHIRPS to 0.43 in IMERG-F and MSWEPv2.2), while the *E* values show the satellite products being comparable to (IMERG-F, MSWEPv2.2 and MSWEPv2.8) or worse than (TMPA and CHIRPS) climatology. All the products underestimate the rainfall rates by 1.6 mm (IMERG-F) to 3.3 mm (MSWEPv2.8). The fact that the underestimation is mostly in the higher rain rates is confirmed by the high RMSE values (on the order of 12 mm day<sup>-1</sup>) for all the products. Overall, IMERG-F emerges as the best product at a daily time scale followed by MSWEPv2.2 and TMPA, CHIRPS, and then MSWEPv2.8.

The performance of the satellite products significantly varies across the stations (Fig. 6) for a variety of reasons. One explanation is the varying weights applied in the gauge calibration of the SREs. The comparison of two subsets of stations: 1) GTS (7 stations whose data is reported to the GTS and have >49% availability of daily data, Fig. 2a), and 2) non-GTS (stations which do not report to the GTS, hence, were not used in the gauge calibration) confirmed this assumption. The QQ plots (Fig. S4 in the supplemental material) show that in all the SREs, the distribution of rain rates for the non-GTS station differs more substantially from that of the RGs compared to the GTS stations. On average,

FEBRUARY 2022



FIG. 6. Boxplots showing the performance of the satellite products on a daily time scale across all the stations. Each circle represents a station. The box is equivalent to the interquartile range (IQR) while orange horizontal line is the median (50th percentile). The lower and upper whiskers are equal to  $Q1 - 1.5 \times IQR$  and  $Q3 + 1.5 \times IQR$  (0.35th and 99.65th percentiles), respectively. Circles beyond the whiskers are outliers and constitute 0.7% of the data.

the performance of the SREs at the pixels closest to the GTS stations is better than that at the pixels closest to the non-GTS stations for all the metrics (Fig. 7). The standard deviation (light red shading in Fig. 7), calculated from an ensemble of 7-station combinations out of the 12 non-GTS stations, demonstrates that for most combinations of non-GTS stations, the performance of the GTS stations is superior. It is further argued here that the overall highest station-to-station

variability observed in most of the nondichotomous (i.e., r, E, ME, MAE, and RMSE) metrics for MSWEPv2.2 is most likely related to the strong weights of gauge observations in this product at grid points close to the GTS stations. MSWEPv2.2 showed the highest spatial variability for r, E, MAE, and RMSE while IMERG-E showed the highest variability for ME (Figs. 6e–l). The disparities in performance at the GTS stations and non-GTS stations were also largest in



FIG. 7. Metrics showing the performance of seven stations with the highest percentage of data reported to the GTS vs performance of station which never report their data to the GTS (non-GTS). The red curve is the ensemble mean of all the possible 7-station combinations out of the 12 non-GTS stations and the light red shading is the standard deviation of the ensemble. The black line shows the performance of all the stations. Note that the normalized version of ME, MAE, and RMSE are preferred here because the subsets of the stations being compared exhibit different rainfall climatologies.

MSWEPv2.2. For example, based on Fig. 7, there was a difference of 0.35, 0.5, and 0.21 for the r, E, and HSS, respectively, in MSWEPv2.2 compared to, for example, 0.11, 0.15, and 0.08 in IMERG-F for the same metrics. MSWEPv2.8 is also substantially different from its predecessor, with smaller differences of 0.12, 0.17, and 0.06, respectively. The difference between the two versions of MSWEP may partly be explained by the reduction in gauges used in the newer version due to stricter gauge inclusion criteria (Beck et al. 2021b).

The performance of the SREs in the different cluster varies (Fig. 8). The POD is highest in MWSEPv2.8 (Fig. 8a), but the skill is degraded by the high rate of false alarms (Fig. 8b). The correlation coefficient (Fig. 8d) is best in IMERG products, followed by the MSWEP products. However, as seen previously, the variability in performance is very large in MSWEPv2.2. The C- and SW-clusters exhibit correlation values of about 0.32 while that of the LVB-cluster is substantially higher at 0.65 (Fig. 8d). This superiority in skill at the LVB-

cluster is partly due to the fact that majority of stations in this cluster are GTS stations. Products perform worst in SW-cluster with the highest rate of false alarms and hence, lowest HSS for all the products (Figs. 8b,c).

As expected, aggregating daily rainfall totals to longer time scales improves the skill in all the SREs (Table 6 and Fig. 9), since any errors due to time mismatches are reduced. The r and E values of all the products improve considerably. For example, the correlation value in CHIRPS increases from 0.22 to 0.50 moving from daily to pentadal accumulations (Table 6). This is likely because CHIRPS is initially created at pentadal time scale and then disaggregated into daily data (Funk et al. 2015). Generally, the performance of the satellite products converges with longer time scale. The "Taylor score" (Taylor 2001), which combines the performance with respect to correlation and standard deviation, increases (i.e., improves) with increasing time scale (Fig. 9). IMERG products overestimate both pentadal and dekadal



FIG. 8. Statistics for the four clusters calculated by concatenating daily rainfall totals for all the stations in a cluster for the period 2001–18 (IMERG-F\_0.1°, IMERG-F\_0.25°, IMERG-E, TMPA, TMPA-RT, CHIRPS, and MSWEPv2.8) and 2001–16 (MSWEPv2.2). All the correlations are statistically significant at the 99% significance level.

rainfall while the other SREs underestimate rainfall at these time scales. All the products underestimate the variability and amount of rainfall at a daily time scale (all points below the RG standard deviation and the arrows point toward the center, Fig. 9). IMERG-F performs best at pentadal and dekadal time steps for most of the metrics, followed by MSWEPv2.2 and MSWEPv2.8, CHIRPS, and then TMPA.

# d. Extremes and return periods

Due to variations in the rainfall climatology as seen in sections 4a and 4b, extreme events are independently determined for individual stations. First, we considered the daily rainfall totals above the 95th percentile of all rainy days at a given station as "extreme." The distribution of extreme events differs significantly between the RGs and SREs for all the SREs (Fig. 10). All the points are below the diagonal meaning the SREs fail to capture extreme rain rates in all the clusters. The deviation is largest in MSWEPv2.8 and CHIRPS while IMERG-F and TMPA are the closest to the 1:1 line. IMERG-E exhibits a better distribution than IMERG-F while TMPA-RT is worse than TMPA. IMERG-F\_0.1° performs better than the IMERG-F\_0.25° as expected since coarse graining averages extremes out. Next, we checked the distribution of extreme events in the SREs (the 95th percentile of the individual SREs is taken as the thresholds) and the corresponding values in the RGs (Fig. S5 in the supplemental material). In this case, the points are above the diagonal for the weaker extremes suggesting that the SREs overestimate these extremes. However, at the stronger extremes, the plot moves below the diagonal pointing toward a tendency of the SREs to underestimate the stronger extremes. Similar results were shown by Monsieurs et al. (2018).

The performance of the SREs declines for extreme events for all the satellite products at all stations in comparison to

when the full dataset is used (cf. Fig. 6). The ability of the satellites to detect extreme events decreases to a POD < 0.21 for all final products (Table 7). The skill of the satellites is also reduced (lower HSS values compared to Fig. 6 and Table 6) due to the increased POFA in all the products. This is most pronounced in CHIRPS (Figs. 11b,d and Table 7). The satellites also underestimate the frequency of extreme rainfall events (BID < 1). All the products greatly underestimate the extreme rain rates (Fig. 11g) on the order of 26 mm day<sup>-1</sup> (53%) or more (Table 7). The skill is worse than climatology for all the products (Fig. 11f and Table 7). The satellite estimates are weakly (0.06-0.26) correlated with the RGs (Fig. 11e and Table 7). Overall, for the final version of the products, MSWEPv2.2 marginally emerges as the best product for extreme events, followed by IMERG, TMPA, MSWEPv2.8, and then CHIRPS. However, as seen previously, the performance of MSWEPv2.2 is highly variable suggesting the influence of the applied gauge weighting. Comparing the final products of IMERG and TMPA with their respective early versions, IMERG-E is better than IMERG while TMPA-RT is worse than TMPA for extreme events detection and accuracy. Indeed, considering all the SREs, IMERG-E emerges best overall for extreme events, having the best scores in all but one metric evaluating rain rates, the best extreme-events detection percentage (21%), and the lowest BID (Table 7). This suggests that that IMERG-E is best suited for extreme events analysis.

With regard to rare events, modeled with the POT method, the performance of SREs is spatially variable and the products generally underestimate the return values (Fig. 12 and Fig. S6 in the supplemental material). Note that the return values are dependent on length and quality of available data (results not shown). Therefore, we constrained the analysis to the period covered by all products (2001–18), except for MSWEPv2.2 (2001–16). Figure 12



FIG. 9. Taylor diagrams showing the correlation coefficient, standard deviation, centered RMSE (CRMSE), and the Taylor score between gauge data and the satellite estimates on (a) daily, (b) pentadal, (c) dekadal, and (d) monthly time scales. The red square marker on the horizontal axis is the reference (gauge) dataset, and therefore, the closer a satellite product to the reference, the better it is. The length of arrows represents the bias magnitude while the direction the sign of the bias with arrows pointing toward (away from) the reference means underestimation (overestimation). The bias shares the same coordinate system and scale as the CRMSE. The shading and blue contour indicate the Taylor score [combines correlation coefficient and standard deviation, and ranges from 0 (worst) to 1 (best)].

illustrates that SREs perform satisfactorily at some stations, e.g., in Kitale (Fig. 12a) for MSWEPv2.2 and poorly in other stations. Apart from nine stations in MSWEPv2.2, all SREs underestimate the return values at all other stations (see Fig. S6 in the supplemental material). The underestimation is most pronounced in CHIRPS, MSWEPv2.8, and TMPA-RT. Since this study has a strong focus on comparison between SREs, we converted the absolute return values at each station to relative return value deviations in order to make the return values comparable across the stations in different climates and datasets. Considering the mean values, all the products always underestimate the return values of extreme events except MSWEPv2.2 which overestimates after about 40 years Satellite estimate (mm/day)

Satellite estimate (mm/day)

(a) N-cluster

(c) SW-cluster



Gauge rainfall (mm/day) FIG. 10. QQ plots showing the distribution of extreme rainfall as recorded by the RGs and the corresponding values in the SREs.

(Fig. 13). The underestimation and overestimation of the return values at the shorter and longer periods, respectively, coupled with a large standard deviation especially at long return periods (Fig. 13h) does not allow for a conclusive statement about the general usefulness of MSWEPv2.2 for the analysis of extreme rainfall. The performance of IMERG is greatly affected by the coarse graining, since the native resolution performs best among all products, but the coarse-grained version is only better than TMPA-RT, MSWEPv2.8, and

CHIRPS. The Early product of IMERG also outperforms the Final product.

# 5. Discussion and conclusions

We analyzed the performance of four satellite products against daily RG data over the complex topography of EEA at multiple spatiotemporal aggregations for the period

TABLE 7. As in the section for daily accumulations in Table 5, but for daily extremes events (rainfall amount > 95th percentile in rainy days subset of RGs). Note that the POD, POFA, BID, and HSS are computed based on the extreme threshold for the individual stations (see Table S1 in the supplemental material), and this threshold is applied to the whole rainy-day (>0.2 mm) subset. The rest of the metrics consider the extremes in the RGs and the corresponding SREs values. The numbers in bold and italics font denote the best scores among the final versions and scores when an early product outperforms all products, respectively.

Satellite product	POD	POFA	BID	HSS	r	E	$\frac{ME}{(mm \ day^{-1})}$	PB (%)	$\begin{array}{c} \text{MAE} \\ (\text{mm day}^{-1}) \end{array}$	RMSE (mm day <sup>-1</sup> )
IMERG-F_0.1°	0.18	0.70	0.59	0.19	0.18	-4.28	-28.9	-58	30.8	35.6
IMERG-F_0.25°	0.15	0.68	0.43	0.18	0.19	-4.28	-29.8	-61	30.9	35.5
IMERG-E	0.21	0.72	0.76	0.21	0.16	-3.78	-26.5	-53	29.1	34.1
TMPA	0.14	0.75	0.56	0.15	0.14	-5.07	-32.4	-62	34.0	38.6
TMPA-RT	0.10	0.72	0.37	0.13	0.14	-5.50	-35.3	-68	36.0	40.4
CHIRPS	0.02	0.81	0.12	0.03	0.06	-6.30	-38.7	-77	38.8	42.5
MSWEPv2.2	0.19	0.56	0.42	0.24	0.26	-4.32	-29.5	-59	30.5	35.7
MSWEPv2.8	0.03	0.55	0.07	0.05	0.15	-5.61	-36.2	-74	36.3	39.8



FIG. 11. As in Fig. 7, but for extreme events. A hit occurs when both the RG and SRE record rainfall exceeding the 95th percentile of raindays subset of the RG data.

2001–18 (except MSWEPv2.2, 2001–16). Additionally, we compared the early versions of IMERG and TMPA, both against each other and with their respective final versions. The native resolution of IMERG was also validated to evaluate how resolution affects the performance of satellite products. We included the two versions of MSWEP in order to compare the change in performance of the latest version compared to its predecessor. Since RG data were missing for some periods, we used satellite data only at corresponding time steps with available RG data. Two sets of metrics that

assess rainfall occurrence and accuracy of the rainfall totals were used. The suitability of the products around extreme events was also tested using the same metrics. Additionally, using the POT method, we assessed if the SREs can be used to simulate return values of rare events for specified return periods. The main conclusions of this validation study are as follows:

1) Generally, all the products reproduce the annual and seasonal rainfall pattern, but the amounts are overestimated



FIG. 12. Return values based on the POT method using RG and the SREs for four randomly select stations in the study area. The gray shading is the confidence bounds (95% significance level) of the return values modeled using RG data.

in IMERG by about 5% and underestimated by TMPA, CHIRPS, and MSWEP by 5%–8% (annually) and 1%–9% (monthly).

- 2) IMERG outperforms the other products on daily, pentadal, and dekadal time scales followed by MSWEPv2.2, TMPA, and then CHIRPS. On monthly and annual time scales, MSWEPv2.2 and CHIRPS, respectively, perform best. MSWEPv2.8 performs worse than its predecessor, MSWEPv2.2 for almost all the spatiotemporal aggregations and for extremes.
- 3) The performance of the satellites is greatly reduced for southwestern part of Uganda, mainly due to the high false alarm rate, which degrades the skill.
- 4) The performance of the all the SREs is spatially variable for reasons mostly elusive to the user. However, MSWEPv2.2 is a salient example where a high gauge weight is capable of drastically improving the performance at grid points close to the station, but the skill drops considerably in places away from the RGs used in the calibration. Unless documented thoroughly, this can lead to serious nontransparency of an SRE's skill for the end-user.
- 5) Except for IMERG-E, all the products miss more than 80% of the daily extreme events reported by the RGs and severely underestimate extreme daily rainfall totals. The fact that IMERG-E outperforms all the other products

demonstrates the potential of NRT products for operational applications such as floods assessment, which need short latency products.

6) IMERG, IMERG-E, TMPA, and MSWEPv2.2 show closest agreement with the RGs for return values analysis while MSWEPv2.8, CHIRPS, and TMPA-RT should be avoided for this purpose. This is not surprising given that the SREs that are the least suitable for modeling extreme events underestimate extreme rainfall events most severely.

Good agreement of the products with RGs on a seasonal scale supports the findings in Camberlin et al. (2019), Diem et al. (2014), and Asadullah et al. (2008) with similar performance statistics and seasonal cycles which are largely controlled by the periodic migration of the tropical rain belt (Nicholson 2017; Seregina et al. 2019). Therefore, SREs capture the drivers of synoptic seasonal rainfall over the study area well. The fact that all the products used in the present study are gauge calibrated may have contributed to the good performance (Awange et al. 2016; Dinku et al. 2018). Additionally, at longer temporal accumulations, any errors due to time mismatches between the RG and SREs data are reduced, which, in turn, improve the scores.

The superior skill of IMERG compared to the other SREs, especially at lower temporal aggregations has been seen in



FIG. 13. (a) Modeled relative return values of extreme events for all the satellite estimates and (b)–(h) individual satellite product relative return value plus the standard deviation across stations (shaded). For each dataset, the values are obtained by taking the mean across all stations and time.

other studies. For instance, Dezfuli et al. (2017a) showed that IMERG was better than TMPA in West and East Africa. In general, the good performance of IMERG can be explained by better performance of PMW sensors (Kidd and Huffman 2011; Bitew and Gebremichael 2011) and improved spatiotemporal resolution (Dezfuli et al. 2017b). However, IMERG overestimates rainfall occurrence and intensity compared to both the RGs and other SREs, especially over Lake Victoria. This is attributed to the tendency of PMW sensor to overestimate rainfall in convectively active regions (Sungmin and Kirstetter 2018; Nicholson et al. 2021). The generally good performance of MSWEPv2.2 agrees with those of Lakew et al. (2020). However, MSWEPv2.8 has a weaker performance compared to MSWEPv2.2. This may stem from 1) the assignment of large weights to ERA5, which has a tendency to rain too often at low rain rates, and 2) the reduction in the number of RG data used in gauge calibration due to a more stringent selection criterion (Beck et al. 2021b). The superior performance of CHIRPS compared to the other SREs at larger temporal accumulations corroborates the findings of Dinku et al. (2018), Diem et al. (2019), and Camberlin et al. (2019) and may be due to 1) use of the gauge–satellite climatology, CHPclim which reduces mean biases, and 2) the fact that gauge calibration is done at longer temporal resolutions (i.e., pentadal and monthly).

SREs are known to show variable performances depending on different factors. Although other factors like the gauge calibration and topography (e.g., Diem et al. 2014; Monsieurs et al. 2018) play a significant role in the variability of the performance of SREs in the study area, we specifically found for MSWEPv2.2 that the weight assigned to RGs relative to the other data inputs during gauge calibration process had a significant influence on how the product performed at a given grid point. This result corroborates several studies (e.g., Awange et al. 2016; Camberlin et al. 2019) that also showed that gauge calibration impacts the performance of SREs.

Major difficulties in capturing rainfall by SREs in the domain of the SW-cluster have already been highlighted in Diem et al. (2014) and Monsieurs et al. (2018). The complex topography partly accounts for the performance. Many parts

Similar to studies by Monsieurs et al. (2018) over western Uganda and Thiemig et al. (2012) over African river basins, the performance for all the SREs declines for extreme events compared to the whole dataset. This may be partly due to the fact that PMW sensors may miss the short-lived intense rainfall events given their infrequent overpasses. Additionally, being gridded products, SREs contain spatial averages and hence, extremes may be smoothed out, which is not the case for the point measurements at RGs. The better detection rate and reduced error in IMERG-E around extreme events compared to the other SREs points to the fact that gauge calibration introduces some uncertainties in the final version of the SREs (Bitew and Gebremichael 2011). For most SREs, the calibration is initially done at a low temporal resolution, e.g., monthly for IMERG, before daily rainfall is rescaled accordingly. This in itself may be problematic for daily extremes. However, for the NRT product of TMPA, results showed overall poorer scores compared to the final version. The stronger underestimation of TMPA-RT was also found by Monsieurs et al. (2018) and this was attributed to the lack of monthly gauge calibration. Unsurprisingly, the SREs with highest errors at the extreme events (CHIRPS, TMPA-RT, and MSWEPv2.8) had the highest error compared to the RGs when applied for modeling the return period of extreme events.

Based on this study, care needs to be taken when using the SREs for a given application as performances varied substantially. This variation stems from various sources, e.g., gauge calibration, algorithms, data inputs and region considered. We are also aware that the performances could also have been influenced by the availability and quality of RG dataset. Also, given that the SREs ranking was different for individual metrics, it is difficult to categorically say which product is the best. The SREs showed great promise at the longer temporal aggregations given the high scores in most metrics. Challenges remain for shorter time scales, especially over heterogeneous topography, and extreme rainfall events. Despite these challenges, IMERG, TMPA, and MSWEPv2.2 are the most suitable products among the tested SREs for hydrometeorological disaster applications as they provide the crucial information for a data-sparse region like EEA. Validating the SREs with a denser gauge network may highlight the apparent regional difference in performance even further and also enable the quantification of the influence of the RGs on the results, if any.

Acknowledgments. Simon Ageet was supported by a DAAD PhD fellowship. The second author acknowledges support from the Transregional Collaborative Research Centre SFB/TRR 165 "Waves to Weather" (www.wavestoweather.de) funded by the German Research Foundation (DFG). We also

would like to thank Hylke Beck for an insightful discussion about the MSWEP algorithms. We also thank the two anonymous reviewers for their constructive comments that greatly helped to improve the manuscript.

Data availability statement. All the SREs used in this study are publicly available from the data creator's repositories; https://daac.gsfc.nasa.gov/datasets, https://data.chc.ucsb.edu/ products/CHIRPS-2.0/, https://platform.princetonclimate.com/ PCA\_Platform, http://www.gloh2o.org/mswep/ for TMPA and IMERG products, CHIRPS, MSWEPv2.2 and MSWEPv2.8, respectively. The RG data are available from the second author upon request.

# REFERENCES

- Asadullah, A., N. McIntyre, and M. Kigobe, 2008: Evaluation of five satellite products for estimation of rainfall over Uganda. *Hydrol. Sci. J.*, **53**, 1137–1150, https://doi.org/10.1623/hysj.53.6. 1137.
- Ashouri, H., and Coauthors, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1.
- Awange, J. L., and Coauthors, 2016: Uncertainties in remotely sensed precipitation data over Africa. *Int. J. Climatol.*, 36, 303–323, https://doi.org/10.1002/joc.4346.
- Basalirwa, C. P. K., 1995: Delineation of Uganda into climatological rainfall zones using the method of principal component analysis. *Int. J. Climatol.*, **15**, 1161–1177, https://doi.org/10. 1002/joc.3370151008.
- Beck, H. E., E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP v2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.
  - –, ––, ––, ––, ––, ––, and ––, 2021a: MSWEPv2.8 3-hourly  $0.1^{\circ} \times 0.1^{\circ}$  historical data. GloH2O, accessed 1 March 2021, http://www.gloh2o.org/mswep/.
  - —, —, —, —, —, , , , and —, 2021b: MSWEPv2.8 technical documentation. GloH2O, 8 pp., accessed 4 May 2021, http://www.gloh2o.org/mswep/.
- Bitew, M. M., and M. Gebremichael, 2011: Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. *Hydrol. Earth Syst. Sci.*, 15, 1147–1155, https://doi.org/10.5194/hess-15-1147-2011.
- Bommier, E., 2014: Peaks-over-threshold modelling of environmental data. U.U.D.M. Project Rep. 2014:33, 35 pp., https:// www.diva-portal.org/smash/get/diva2:760802/FULLTEXT01. pdf.
- Camberlin, P., and Coauthors, 2019: Evaluation of remotely sensed rainfall products over Central Africa. *Quart. J. Roy. Meteor. Soc.*, 145, 2115–2138, https://doi.org/10.1002/qj.3547.
- Cattani, E., A. Merino, and V. Levizzani, 2016: Evaluation of monthly satellite-derived precipitation products over East Africa. J. Hydrometeor., 17, 2555–2573, https://doi.org/10. 1175/JHM-D-15-0042.1.
- Collins, M., and Coauthors, 2019: Extremes, abrupt changes and managing risk. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner et al., Eds., IPCC, 589–655, https://www.ipcc.ch/site/assets/uploads/sites/3/ 2019/11/10\_SROCC\_Ch06\_FINAL.pdf.

- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/ 10.1002/qj.828.
- Dezfuli, A. K., C. M. Ichoku, K. I. Mohr, and G. J. Huffman, 2017a: Precipitation characteristics in West and East Africa from satellite and in situ observations. J. Hydrometeor., 18, 1799–1805, https://doi.org/10.1175/JHM-D-17-0068.1.
- —, —, G. J. Huffman, K. I. Mohr, J. S. Selker, N. van de Giesen, R. Hochreutener, and F. O. Annor, 2017b: Validation of IMERG precipitation in Africa. *J. Hydrometeor.*, 18, 2817–2825, https://doi.org/10.1175/JHM-D-17-0139.1.
- Diem, J. E., J. Hartter, S. J. Ryan, and M. W. Palace, 2014: Validation of satellite rainfall products for western Uganda. J. Hydrometeor., 15, 2030–2038, https://doi.org/10.1175/JHM-D-13-0193.1.
- —, B. L. Konecky, J. Salerno, and J. Hartter, 2019: Is equatorial Africa getting wetter or drier? Insights from an evaluation of long-term, satellite-based rainfall estimates for western Uganda. *Int. J. Climatol.*, **39**, 3334–3347, https://doi.org/10. 1002/joc.6023.
- Dinku, T., 2019: Challenges with availability and quality of climate data in Africa. *Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation*, Elsevier, 71–80, https://doi.org/10.1016/B978-0-12-815998-9.00007-5.
- —, K. Cressman, and S. J. Connor, 2010a: Evaluating detection skills of satellite rainfall estimates over desert locust recession regions. J. Appl. Meteor. Climatol., 49, 1322–1332, https://doi. org/10.1175/2010JAMC2281.1.
- —, S. J. Connor, and P. Ceccato, 2010b: Comparison of CMORPH and TRMM-3B42 over Mountainous Regions of Africa and South America. *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 193–204, https://doi.org/10.1007/978-90-481-2915-7\_11.
- —, C. Funk, P. Peterson, R. Maidment, T. Tadesse, H. Gadain, and P. Ceccato, 2018: Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quart. J. Roy. Meteor. Soc.*, 144, 292–312, https://doi.org/10.1002/gi.3244.
- Ebert, E. E., 2007: Methods for verifying satellite precipitation estimates. *Measuring Precipitation from Space*, V. Levizzani, P. Bauer, and F. J. Turk, Eds., Springer, 345–356, https://doi. org/10.1007/978-1-4020-5835-6\_27.
- Eggermont, H., K. Van Damme, and J. M. Russell, 2009: Rwenzori Mountains (Mountains of the Moon): Headwaters of the White Nile. *The Nile: Origin, Environments, Limnology and Human Use*, H. Dumont, Ed., Springer, 243–261.
- Engel, T., A. H. Fink, P. Knippertz, G. Pante, and J. Bliefernicht, 2017: Extreme precipitation in the West African cities of Dakar and Ouagadougou: Atmospheric dynamics and implications for flood risk assessments. J. Hydrometeor., 18, 2937– 2957, https://doi.org/10.1175/JHM-D-16-0218.1.
- Flohn, H., and K. Fraedrich, 1966: Tagesperiodische Zirkulation und Niederschlagsverteilung am Victoria-See (Ostafrika). *Meteor. Rundsch.*, 19, 157–165.
  - , and T. Burkhardt, 1985: Nile runoff at Aswan and Lake Victoria; an example of a discontinuous climate time series.
    Z. Gletschkd. Glazialgeol., 21, 125–130.
- Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations – A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10. 1038/sdata.2015.66.

- Hartigan, J. A., and M. A. Wong, 1979: A K-means clustering algorithm. J. Roy. Stat. Soc., 28, 100–108, https://doi.org/10. 2307/2346830.
- Hastings, D. A., and Coauthors, 1999: The Global Land One-Kilometer Base Elevation (GLOBE) digital elevation model, version 1.0. NOAA National Geophysical Data Centre, accessed 23 July 2021, https://www.ngdc.noaa.gov/mgg/topo/ globe.html.
- Huffman, G. J., and D. T. Bolvin, 2018: TRMM and other data precipitation set documentation. NASA TRMM Doc., 48 pp., https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/ 3B42\_3B43\_doc\_V7.pdf.
- —, and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20, https://doi.org/10.1175/1520-0477(1997) 078<0005:TGPCPG>2.0.CO;2.
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeor., 8, 38–55, https://doi.org/10.1175/JHM560.1.
- —, D. T. Bolvin, D. Bolvin, D. Braithwaite, K. Hsu, and R. Joyce, 2020a: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD), version 6.3, accessed 11 May 2020, 35 pp., https://gpm.nasa.gov/sites/default/files/2020-05/IMERG\_ATBD\_V06.3.pdf.
- —, E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and J. Tan, 2020b: GPM IMERG Final Precipitation L3 Half Hourly 0.1° x 0.1° V06 (GPM\_3IMERGHH). Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 April 2020, https://doi.org/10.5067/GPM/IMERG/3B-HH/06.
- —, —, —, and —, 2020c: GPM IMERG Early Precipitation L3 Half Hourly 0.1° x 0.1° V06 (GPM\_3I-MERGHHE). Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 1 June 2020, https://doi.org/10.5067/GPM/IMERG/3B-HH-E/06.
- IDMC, 2021: GRID 2021 internal displacements in a changing climate. Global Report on International Displacement 2021, 160 pp., https://www.internal-displacement.org/sites/default/ files/publications/documents/grid2021\_idmc.pdf.
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, https://doi.org/10.1175/1520-0493(1999) 127<2204:FASOCR>2.0.CO;2.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeor., 5, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2. 0.CO;2.
- Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteor. Appl.*, 18, 334–353, https://doi.org/10.1002/met.284.
- Kizza, M., A. Rodhe, C. Y. Xu, H. K. Ntale, and S. Halldin, 2009: Temporal rainfall variability in the Lake Victoria Basin in East Africa during the twentieth century. *Theor. Appl. Climatol.*, **98**, 119–135, https://doi.org/10.1007/s00704-008-0093-6.
- —, I. Westerberg, A. Rodhe, and H. K. Ntale, 2012: Estimating areal rainfall over Lake Victoria and its basin using groundbased and satellite data. J. Hydrol., 464–465, 401–411, https:// doi.org/10.1016/j.jhydrol.2012.07.024.
- Kummerow, C., W. S. Olson, and L. A. Giglio, 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans.*

Geosci. Remote Sens., 34, 1213-1232, https://doi.org/10.1109/ 36.536538.

- —, and Coauthors, 2001: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. J. Appl. Meteor. Climatol., 40, 1801–1820, https://doi.org/10.1175/1520-0450(2001)040<1801:TEOTGP>2. 0.CO;2.
- Lakew, H. B., S. A. Moges, and D. H. Asfaw, 2020: Hydrological performance evaluation of multiple satellite precipitation products in the upper Blue Nile basin, Ethiopia. J. Hydrol. Reg. Stud., 27, 100664, https://doi.org/10.1016/j.ejrh.2020.100664.
- Le Coz, C., and N. van de Giesen, 2020: Comparison of rainfall products over sub-Saharan Africa. J. Hydrometeor., 21, 553– 596, https://doi.org/10.1175/JHM-D-18-0256.1.
- Legates, D. R., and G. J. McCabe, 1999: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, **35**, 233–241, https://doi. org/10.1029/1998WR900018.
- Lemos, I. P., and Coauthors, 2020: Thresholdmodeling: A Python package for modeling excesses over a threshold using the Peak-Over-Threshold Method and the Generalized Pareto Distribution. J. Open Source Softw., 5, 2013, https://doi.org/ 10.21105/joss.02013.
- Love, T. B., V. Kumar, P. P. Xie, and W. Thiaw, 2004: A 20-year daily Africa precipitation climatology using satellite and gauge data. *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 5.4, https://ams.confex.com/ams/84Annual/ techprogram/paper\_67484.htm.
- Maidment, R. I., D. I. F. Grimes, R. P. Allan, H. Greatrex, O. Rojas, and O. Leo, 2013: Evaluation of satellite-based and model re-analysis rainfall estimates for Uganda. *Meteor. Appl.*, 20, 308–317, https://doi.org/10.1002/met.1283.
- —, and Coauthors, 2017: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. Sci. Data, 4, 170063, https://doi.org/10.1038/sdata.2017.63.
- Maranan, M., A. H. Fink, P. Knippertz, L. K. Amekudzi, W. A. Atiah, and M. A. Stengel, 2020: A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the West African forest zone. J. Hydrometeor., 21, 729–749, https://doi.org/10.1175/JHM-D-19-0257.1.
- McCollum, J. R., A. Gruber, and M. B. Ba, 2000: Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa. J. Appl. Meteor., 39, 666–679, https://doi.org/10. 1175/1520-0450-39.5.666.
- Monsieurs, E., and Coauthors, 2018: Evaluating TMPA rainfall over the sparsely gauged East African Rift. J. Hydrometeor., 19, 1507–1528, https://doi.org/10.1175/JHM-D-18-0103.1.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, **10**, 282–290, https://doi.org/10.1016/0022-1694(70) 90255-6.
- Nicholson, S. E., 2017: Climate and climatic variability of rainfall over eastern Africa. *Rev. Geophys.*, 55, 590–635, https://doi. org/10.1002/2016RG000544.
- —, D. Klotter, and A. T. Hartman, 2021: Lake-effect rains over Lake Victoria and their association with Mesoscale Convective Systems. J. Hydrometeor., 22, 1353–1368, https://doi.org/ 10.1175/JHM-D-20-0244.1.
- Novella, N. S., and W. M. Thiaw, 2013: African rainfall climatology version 2 for famine early warning systems. J. Appl. Meteor. Climatol., 52, 588–606, https://doi.org/10.1175/JAMC-D-11-0238.1.

- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. J. Mach. Learn. Res., 12, 2825–2830, https://doi. org/10.5555/1953048.2078195.
- Phillips, J., and B. McIntyre, 2000: ENSO and interannual rainfall variability in Uganda: Implications for agricultural management. Int. J. Climatol., 20, 171–182, https://doi.org/10.1002/ (SICI)1097-0088(200002)20:2<171::AID-JOC471>3.0.CO;2-O.
- Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. J. Climate, 30, 1643–1664, https://doi.org/10.1175/ JCLI-D-16-0570.1.
- Satgé, F., D. Defrance, B. Sultan, M. Bonnet, F. Seyler, N. Rouche, F. Pierron, and J. Paturel, 2019: Evaluation of 23 gridded precipitation datasets across West Africa. *J. Hydrol.*, 581, 124412, https://doi.org/10.1016/j.jhydrol.2019.124412.
- Seregina, L. S., A. H. Fink, R. van der Linden, N. A. Elagib, and J. G. A. Pinto, 2019: A new and flexible rainy season definition: Validation for the Greater Horn of Africa and application to rainfall trends. *Int. J. Climatol.*, **39**, 989–1012, https:// doi.org/10.1002/joc.5856.
- Sungmin, O., and P. E. Kirstetter, 2018: Evaluation of diurnal variation of GPM IMERG-derived summer precipitation over the contiguous US using MRMS data. *Quart. J. Roy. Meteor. Soc.*, 144, 270–281, https://doi.org/10.1002/qj.3218.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106, 7183– 7192, https://doi.org/10.1029/2000JD900719.
- Thiemig, V., R. Rojas, M. Zambrano-Bigiarini, V. Levizzani, and A. De Roo, 2012: Validation of satellite-based precipitation products over sparsely gauged African river basins. J. Hydrometeor., 13, 1760–1783, https://doi.org/10.1175/JHM-D-12-032.1.
- Thorndike, R. L., 1953: Who belongs in the family? *Psychometrika*, **18**, 267–276, https://doi.org/10.1007/BF02289263.
- Tian, Y., and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. J. Geophys. Res., 114, D24101, https://doi.org/10.1029/2009JD011949.
- UBOS, 2020: Population and census. Uganda Bureau of Statistics, accessed 7 December 2020, https://www.ubos.org/explorestatistics/20/.
- Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting, 2018: Skill of global raw and postprocessed ensemble predictions of rainfall over Northern Tropical Africa. *Wea. Forecasting*, **33**, 369–388, https://doi.org/10.1175/WAF-D-17-0127.1.
- Washington, R., and Coauthors, 2006: Africa climate change: Taking the short route. *Bull. Amer. Meteor. Soc.*, 87, 1355–1366, https://doi.org/10.1175/BAMS-87-10-1355.
- Wilks, D. S., 2011: Statistical Methods in the Atmospheric Sciences. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- World Bank, 2020: Climate Change Knowledge Portal. Accessed 7 December 2020, https://climateknowledgeportal.worldbank. org/country/uganda/vulnerability.
- Young, M. P., C. J. R. Williams, J. C. Chiu, R. I. Maidment, and S.-H. Chen, 2014: Investigation of discrepancies in satellite rainfall estimates over Ethiopia. J. Hydrometeor., 15, 2347– 2369, https://doi.org/10.1175/JHM-D-13-0111.1.
- Zhang, Y., S. Moges, and P. Block, 2016: Optimal cluster analysis for objective regionalization of seasonal precipitation in regions of high spatial-temporal variability: Application of western Ethiopia. J. Climate, 29, 3697–3717, https://doi.org/10. 1175/JCLI-D-15-0582.1.