Weather and Climate Extremes 35 (2022) 100416

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Evaluating performance of 20 global and quasi-global precipitation products in representing drought events in Ethiopia I: Visual and correlation analysis

Mekonnen Adnew Degefu[a,b,*], Woldeamlak Bewket[a], Yosef Amha[c]

[a] Department of Geography & Environmental Studies, Addis Ababa University, P. O. Box: 1176, Addis Ababa, Ethiopia
[b] Department of Geography & Environmental Studies, Debre Markos University, P.O. Box 269, Debre Markos, Ethiopia
[c] European Economic Commission for Africa, African Climate Policy Center (ACPC), United Nations Economic Commission for Africa (UNECA), P.O. Box: 3001, Addis Ababa, Ethiopia

ARTICLE INFO

Keywords:
Global precipitation
Drought utility
Performance evaluation
SPI method
Ethiopia

ABSTRACT

Increased availability and access to satellite observation and the advanced blending techniques for generating reanalysis data have improved accuracy of precipitation products for data poor regions. However, all data products should be evaluated for their performance before being used for specific applications in different regions or locations. In this study, we evaluated 20 precipitation products for their drought monitoring performance over Ethiopia. These datasets are produced by different state-of-the-art techniques, including: 1) gauge-interpolated, 2) satellite-estimate, 3) reanalysis and 4) merged from multi-sources. The Ethiopian gauge-satellite gridded precipitation dataset and gauge records from 126 stations were used for ground truthing. Drought indices were generated for all data products using the Standardized Precipitation Index (SPI) method at 3- and 12-month time scales. We evaluated accuracy of the data products in representing occurrence and spatial and temporal patterns of the 3- and 12-month droughts using visual and correlation analysis techniques. The data were visually compared against reference data in representing four selected major drought episodes in the country (1984, 2002, 2009 and 2015). The Spearman's correlation was used to quantify relationships between the 20 precipitation products and two reference datasets using SPI values generated at 3- and 12-month time scales. Results showed considerable discrepancies and poor performance for most datasets. The ability to represent the spatial pattern and severity of major drought events varied between drought years. Although the correlation result for areal average SPI time series of the 3- and 12-month droughts for all data products showed statistically significant correlation with the reference data, there is discrepancy between the data products across space. Overall, only three out of the 20 products (CHIRPS, FLDAS and GPCC) performed relatively better. Our results provide important information to guide the choice of precipitation products for drought research, and operational drought risk management as well as useful feedback to data developers to further improve their products.

1. Introduction

Precipitation is a key component of the hydrological cycle that is widely used for drought monitoring and forecasting across the world (WMO , 2012). However, ground observations that provide accurate precipitation data are very scarce in many drought-prone parts of the world, and even declining in most parts of Africa (Dinku et al., 2014; Vogel et al., 2019). In addition, available station records are often low quality, and difficult to access due to legal restrictions, lack of dissemination capacity, or high cost of data in many parts of Africa (Dinku et al., 2014). Innovation in satellite observation and the state-of-the-art techniques in precipitation data production such as interpolation from in-situ observations, modeling and reanalysis have created a new avenue to fill data voids in data-poor regions (Le Coz and van de Giesen, 2020). This is because these precipitation products provide geographically and temporally continuous information for ungauged or data-poor regions, and are freely available for end users (AghaKouchak et al., 2015).

* Corresponding author. Department of Geography & Environmental Studies, Addis Ababa University, P. O. Box: 1176, Addis Ababa, Ethiopia.
E-mail addresses: mekonnenadnew@yahoo.com (M.A. Degefu), woldeamlak.bewket@aau.edu.et (W. Bewket), amhay@un.org (Y. Amha).

https://doi.org/10.1016/j.wace.2022.100416
Received 22 July 2021; Received in revised form 11 January 2022; Accepted 31 January 2022
Available online 3 February 2022
2212-0947/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Currently, there are many precipitation products which can be grouped into four classes. (*1) Gauge interpolated*: these are precipitation products created by interpolating/regridding gauge observations. Some of the high resolution gauge interpolated precipitation products are the Climate Prediction Center (CPC; Xie et al., 2007), Climate Research Unit (CRU; Harris et al., 2014), Global Precipitation Climatology Centre (GPCC; Schneider et al., 2014), and PRECipitation REConstruction over Land (PREC/L; Chen et al., 2002). Each of these products are different in terms of spatial resolution, number of gauges used and interpolation methods. (*2) Satellite-only*: these are precipitation products either derived from polar orbiting passive microwave (PMW) or geostationary thermal infrared (TIR) or a combination of the two. Some of the high resolution (<0.5°) satellite-based precipitation products are Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000), PERSIANN-Cloud Classification System (PERSIANN-CCS; Hong et al., 2004), Climate Prediction Center MORPHing (CMORPH; Joyce et al., 2004), Tropical Rainfall Measuring Mission (TRMM; Huffman et al., 2007), and Climate Hazard Group InfraRed Precipitation (CHIRP; Funk et al., 2015). 3) *Satellite-gauge merged*: these precipitation products have been produced by merging TIR or PMW or both with gauge observation and sometimes with radar and reanalysis data using different data merging algorisms (Le Coz and van de Giesen, 2020). Some of the widely used merged products include Africa Rainfall Climatology version 2 (ARC2; Novella and Thiaw, 2013), Climate Hazard Group InfraRed Precipitation with Stations (CHIRPS; Funk et al., 2015), PERSIANN-Climate Data Record (PERSIANN-CDR; Ashouri et al., 2014), TRMM Multi-Satellite Precipitation Analysis V-3B43 (TMPA-3B43; Huffman et al., 2007), Tropical Applications of Meteorology using SATellite and ground-based observations (TAMSAT; Tarnavsky et al., 2014), Rainfall Estimate version 2 (RFE2), and the Tropical Applications of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT; Tarnavsky et al., 2014). Of these ARC2, RFE2 and TAMSAT are developed specifically for Africa. Each precipitation product in satellite-only and satellite-gauge merged groups differ each other in terms of spatial resolution, type and number of data inputs, record length, algorism used to estimate data, etc. 4) *Reanalysis products*: these precipitation products are generated by assimilating information derived from weather generator numerical models with ground or satellite observation (Balsamo et al., 2015). Some of these are Global Land Data Assimilation System (GLDAS: Rodell et al., 2004), Modern-Era Retrospective analysis for Research and Application version 2 (MERRA2: Gelaro et al., 2017) and European Centre for Medium-Range Weather Forecasts ReAnalysis version 5 (ERA5; Albergel et al., 2018) and the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS; McNally et al., 2017). These precipitation data products also differ in many ways (e.g., spatial resolution, parameterization, model algorism and bias correction).

Although these data products largely increased availability and access to many precipitation products, they cannot be directly used for specific purposes before testing for suitability. This is because each precipitation product has its own advantages and weaknesses and its appropriateness for specific applications (such as for drought or flood monitoring or agrometeorological applications) vary from place to place and over time due to many factors (e.g. variations in type and number of data inputs, spatial resolution, and types of data production algorisms) (Duan et al., 2016; Beck et al., 2017; Le Coz and van de Giesen, 2020). In other words, different precipitation products have different quality in representing the various aspects of precipitation occurrences (e.g. amount, frequency, intensity, seasonality, geographical distribution, etc). Hence, different precipitation products serve different specific functions (e.g., climatological studies, hydrological application, agrometeorological monitoring, and drought and flood monitoring and forecasting). For example, the precipitation product commonly recommended for drought monitoring is the one that can accurately represent low precipitation amounts (Tarnavsky et al., 2014). It is therefore imperative to evaluate spatiotemporal representativeness and applicability of available precipitation products for specific purposes before any actual use (Le Coz and van de Giesen, 2020).

During the last two decades, many country scale (e.g., Dinku et al., 2007, 2008, 2011; Diro et al., 2009; Romilly and Gebremichael, 2011; Young et al., 2014; Lemma et al., 2019) and basin scale (e.g., Hirpa et al., 2010; Worqlul et al., 2014; Bayissa et al., 2017; Gebremichael et al., 2017; Sahlu et al., 2017; Ayehu et al., 2018; Belay et al., 2019) studies have evaluated performance of various global and quasi-global precipitation products in capturing various aspects of precipitation occurrences over Ethiopia. The results show large differences between data products in representing rainfall events and amount across space and seasons. For example, Dinku et al. (2008) reported strong agreement between the monthly precipitation for three global gauge-interpolated precipitation products (CPC, CRU, GPCC) and reference stations. On the other hand, different results were reported on the representativeness of different satellite-based precipitation products. For example, Dinku et al. (2007, 2011) and Romilly and Gebremichael (2011) concluded that no single satellite product performs best for Ethiopia, as it varied by location and the specific application considered. Hirpa et al. (2010) found comparable performance for the real time TRMM (3B42RT) and CMORPH products in representing the spatial pattern, bias and elevation-dependent trends. Bayissa et al. (2017); Gebremicael et al. (2017); Lemma et al. (2019) and Belay et al. (2019) identified CHIRPS as the best satellite based precipitation product for Ethiopia. Others such as Young et al. (2014) identified TRMM as the best data in capturing the mean rainfall, rainy events and seasonal variability across different basins, while Sahlu et al. (2017) identified CMORPH to be the best data that accurately represents the wet season rainfall over the Blue Nile Basin. Dinku et al. (2018) compared CHIRP and CHIRPS precipitation products with ARC2 and TAMSAT products at daily, decadal and monthly time scales over the east African region. The results show that the first two precipitation products to be significantly superior to ARC2 and relatively better than TAMSAT products. However, TAMSAT showed better skill in representing precipitation at the daily time scale. On the other hand, almost all available studies concurred that the gridded data products have better skill in representing precipitation amount during the wet season than the dry season, and most satellite-based products underestimated precipitation amount in the highland and overestimated in the lowland parts of the country (Dinku et al., 2007, 2011; Young et al., 2014). To the best of our knowledge, only one study evaluated performance of some of the precipitation products for their drought monitoring suitability (Bayissa et al., 2017). This study found CHIRPS to be more suitable for meteorological drought monitoring over the Upper Blue Nile Basin compared to other four gridded products (ARC2, TMPA-3B42, PERSIANN/CDR and African Rainfall Climatology and Time-series (TARCAT).

The aim of this study was to evaluate multiple precipitation products for drought monitoring performance over Ethiopia. Evaluation of precipitation data products can provide valuable information both for data users and developers. For local data users (e.g., researchers and practitioners), evaluation results are useful to guide selection of suitable data for research on various aspects of drought episodes. For data developers, results provide valuable feedback to identify problems on their respective data products for areas like Ethiopia which is dominated by complex topography and climate system. The feedback is thus useful for data developers to find ways to improve their products.

2. Methodology

2.1. The physiographic context of Ethiopia

This study is conducted in one of the most drought stricken country (Ethiopia) located in the Greater Horn of Africa. Ethiopia is located between 3° and 15° N and 33° and 48° E, and covers 1.12 million km^2 area (Fig. 1). Ethiopia is characterized by very complex topography with

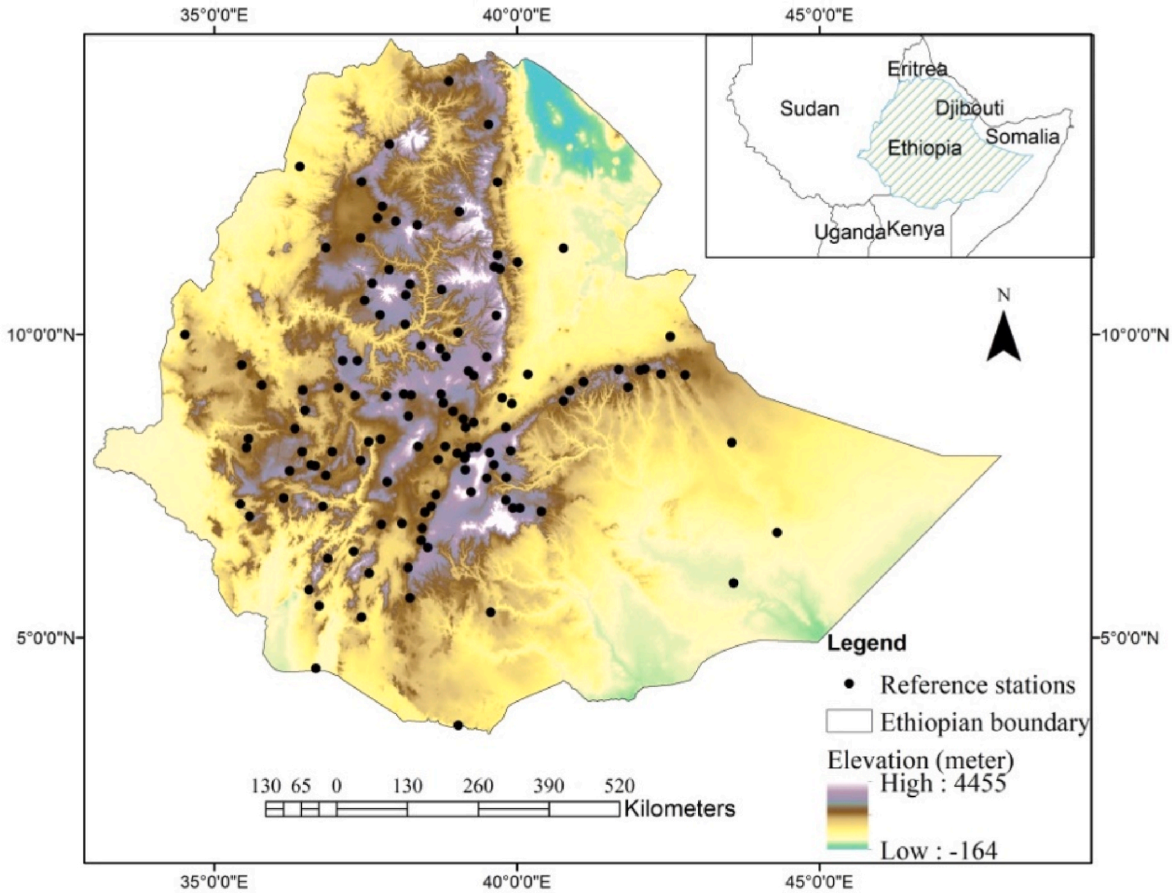

Fig. 1. Map of Ethiopia and the reference stations used in the study.

elevation ranging from below sea level in the northeastern part to more than 4 600 m above sea level at Mount Ras Dejen in the northern part of the country. About 40% of the country is covered by mountains and plateaus with elevation greater than 1 500 m asl, and this is classified as highland. The remaining has elevation of less than 1 500 m asl and it is classified as lowland. The highland part is highly dissected by major river systems that include the Blue Nile, Awash, Baro-Akobo, Genale-Dawa, Omo-Ghibe, Tekezie and Wabeshebele flowing from the central highlands into different directions. The Great East African Rift Valley System that runs from northeast to southwest divides Ethiopia into eastern and western parts. These complex topographic features have created many local climatic conditions that range from hot deserts in the northeast and southeastern lowlands to very cold mountain ranges in the Simien and Arsi-Bale mountain systems (Degefu et al., 2017). The mean annual temperature varies between less than 10 °C in the highlands to more than 40 °C in the hot lowlands, while the mean annual rainfall varies between less than 300 mm in the southeastern and northeastern lowlands to more than 2000 mm in the southwestern highlands. The north-south migration of the Inter-tropical Convergence Zone (ITCZ) and the complex topographic features of the country have large control over the seasonal and spatial distribution of precipitation and other climate elements (Dinku et al., 2011).

In addition to the spatial variation, rainfall also shows very high inter-annual variability, with frequent drought and flood occurrences. The development of drought and flood in Ethiopia is associated with large-scale atmospheric circulation anomalies that take place over the equatorial east Pacific (the El Niño–Southern Oscillation; ENSO) and the Indian Ocean, where warming/cooling events are associated with drought and flood events during the main (*kiremt*) rainfall season (Degefu et al., 2017). Evidences from local chronicles, archival data, historical texts and literature, travelers' dairies and European explorers' notes indicate that drought incidence in Ethiopia dates back to 253 BC (Environmental Protection Authority, 1998). Recent evidences show frequency and impacts of drought have increased since the second half of the 20th century and drought has become the number one environmental problem affecting the subsistence based rain-fed agricultural economy and human livelihood in Ethiopia (Vicente-Serrano et al., 2012).

2.1.1. *Data description*

Twenty global and quasi-global precipitation products were evaluated in this study for their performance in capturing the spatiotemporal conditions of drought episode over Ethiopia (Table 1). These precipitation products varied in terms of spatial resolution (0.04°–1°), number and type of data inputs used to construct the data (TIR, PMW, radar, gauge based and reanalysis or a combination of these in many ways), and the type of methods or algorisms used to construct the data. These data were selected due to their wider applications for research and operational activities, accessibility and high spatial resolution (=<1°). Two precipitation data products that include: 1) quality-controlled gauge records from 126 stations (Fig. 1), and 2) Ethiopian satellite-gauge merged precipitation data at 0.04° (Dinku et al., 2013, 2014) were used as reference datasets. The global and regional precipitation datasets used in this study are from the four types mentioned above, the reference data being the fifth. These are: 1) Reference data, 2) Gauge-interpolated, 3) Satellite data, 4) Reanalysis, and 5) Blended multi-source data. All the 20 precipitation data products were regridded into a common 0.04° spatial resolution to enable comparisons. Brief description of each of these data classes is given below, and detailed information about each individual precipitation data product is

Table 1
Description of reference and global precipitation datasets used in this study.

No.	Dataset	Record length	Temporal resolution	Spatial resolution	Data category	Source of data
Gauge-interpolated						
1	CPC	1979-present	Daily	0.5°	Gauge	https://www.esrl.noaa.gov/psd/
2	CRU	1901-present	Monthly	0.5°	Gauge	http://www.cru.uea.ac.uk/data/
3	GPCC	1901-present	Monthly	1°	Gauge	https://psl.noaa.gov/data/gridded/data.gpcc.html
4	PREC/L	1948-present	Monthly	1°	Gauge	https://psl.noaa.gov/thredds/dodsC/Datasets/precl/
Satellite-based						
5	AIRS	2003-present	Monthly	1°	TIR	http://disc.sci.gsfc.nasa.gov/AIRS/data/
6	CHIRP	1981-present	Monthly	0.05°	TIR	ftp://chg-ftpout.geog.ucsb.edu/pub/org/chg/products/CHIRP/monthly
7	PERSIANN	2001-present	Monthly	0.25°	TIR, PMW	ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN-CDR/mthly
8	PERSIANN-CCS	2003-present	Monthly	0.04°	TIR, PMW	ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN-CDR/mthly
Reanalysis						
9	ERA5	1979-present	Monthly	0.28°	Reanalysis	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5
10	FLDAS	1982-present	Monthly	0.1°	Reanalysis	https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_C_GL_M_001/
11	GLADS	1979-present	Monthly	1°	Reanalysis	https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS
12	MERRA2	1980-present	Monthly	0.66°x0.50°	Reanalysis	https://disc.gsfc.nasa.gov/datasets?keywords=MERRA-2/
Merged from multi-sources						
13	ARC2	1996_present	Daily	0.1°	TIR, gauge	ftp://ftp.cpc.ncep.noaa.gov/fews/fewsdata/africa/arc2
14	CHIRPS	1981-present	Daily	0.05°	TIR, Model, gauge	ftp://chg-ftpout.geog.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_monthly/tifs
15	GPM-IMERG	2001-present	Monthly	0.1°	TIR, PMW, sat-radar, gauge	https://pmm.nasa.gov/data-access/downloads/gpm
16	PERSIANN_CDR	1983-present	Monthly	0.25°	TIR, PMW, gauge	ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN-CDR/mthly
17	TAMSAT	1983-present	Monthly	0.05°	TIR, gauge	https://www.tamsat.org.uk/data/rfe/index.cgi
18	RFE2	2001-present	Monthly	0.1°	TIR, PMW, gauge	https://earlywarning.usgs.gov/fews/datadownloads/Continental%20Africa/Dekadal%20RFE
19	TerraClimate	1958-present	Monthly	0.04°	TIR, gauge, model	https://climate.northwestknowledge.net/TERRACLIMATE/
20	TRMM-3B43	1998-present	Monthly	0.25°	TIR, VIS, PMW, Sat-radar, gauge	https://pmm.nasa.gov/dataaccess/downloads/trmm
Reference data						
1	Ethiopian satellite-gauge merged	1983-present	Monthly	0.04°	TIR, gauge	NMA
2	In-situ stations	1983-present	monthly	point	Station (126)	NMA

available at the sources (see Table 1 for sources).

2.1.2. Reference datasets

We used two reference data products as benchmark to compare the drought monitoring performance of the other studied data products: (i) monthly precipitation data from 126 stations, and (ii) satellite-gauge merged gridded precipitation product; both covered the period from 1983 to 2018. We selected 126 representative stations after checking 250 stations for their data quality (e.g., missing data and length of record). We used these station-based dataset to evaluate the performance of global datasets in capturing locally developed drought indices. The satellite-gauge merged precipitation data is produced by the National Meteorological Agency (NMA) of Ethiopia in collaboration with the International Research Institute (IRI), Columbia University (Dinku et al., 2013). It is produced by merging precipitation observations for more than 600 stations with precipitation estimated from the Meteosat TIR which has a time step of 30 min for the first and 15 min frequency for second generation obtained from European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). It was produced by using locally calibrated TAMSAT algorithm on 0.04° (4 km) spatial resolution and it is available at different time scales (daily, dekdal and monthly) starting from 1983 (Dinku et al., 2013). Both reference datasets were obtained from NMA.

2.1.3. Gauge-interpolated

In this study, we considered the most widely used four gauge-interpolated precipitation products. These are the Climate Prediction Center (CPC; Xie et al., 2007), Climate Research Unit (CRU; Harris et al., 2014), Global Precipitation Climatology Centre (GPCC; Schneider et al., 2014), and the National Oceanographic and Atmospheric Administration (NOAA) PRECipitation REConstruction over Land (PREC/L; Chen et al., 2002). The spatial resolution for CPC and CRU datasets is 0.5° and for GPCC and PREC/L is 1°.

2.1.4. Satellite data

Currently, there are many precipitation datasets retrieved from passive microwave (PMW) information obtained from polar orbiting satellites and Thermal InfraRed (IR) information that is frequently measured from geostationary satellites, and a hybrid from PMW and TIR using different data retrieval algorisms (Duan et al., 2016). In this study, we evaluated the performance of four satellite observation only precipitation datasets that include the Atmospheric Infrared Sounder-only version 6 (AIRS v6 AO; Aumann et al., 2004), the Climate Hazard Group InfraRed Precipitation (CHIRP; Funk et al., 2015), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000) and PERSIANN-Cloud Classification System (PERSIANN-CCS: Nguyen et al., 2019). The spatial resolution is 1° for ARIS, 0.25° for PERSIANN, 0.05° for CHIRP and 0.04° for PERSIANN-CCS data products.

2.1.5. Reanalysis products

Reanalysis precipitation data are produced by assimilating information from weather generation numerical models with ground or satellite based observation. In this study, we evaluated the performance of four reanalysis precipitation products that are widely used for climate system study and drought monitoring across the world. These reanalysis products include European Centre for Medium-Range Weather Forecasts (ECMWF) ReAnalysis (ERA5; Albergel et al., 2018), Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS; Rui and McNally, 2019), Global Land Data Assimilation System (GLDAS; Rodell et al., 2004), and Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2; Gelaro et al., 2017). These datasets are different in many ways such as variable inputs, spatial and temporal resolutions, and model simulations used to estimate precipitation amount. The spatial resolutions of these datasets are 0.28°, 0.1°, 1° and 0.58°, respectively.

2.1.6. Merged precipitation from multi-sources

There has been an increasing effort to produce high quality precipitation data by merging precipitation products obtained from two or more sources (satellite, gauge, radar and mode reanalysis) using the state-of-the-art data merging methods (AghaKouchak et al., 2015). This method enables to produce high quality precipitation product as it helps to reduce inherent weaknesses and adds advantages from each data input (Dinku et al., 2014). For example, gauge-based measurements provide the most accurate precipitation estimate over relatively long period of time, but precipitation data from gauge observations cannot provide spatially continuous information as gauge based observations are too sparse or absent over large areas and temporally inconsistent due to missing data. On the other hand, satellite based rainfall observation provides spatially and temporally consistent and continuous precipitation estimates, however, the available satellite based rainfall products are less accurate and have coarse spatial resolution (AghaKouchak et al., 2015). In this study, we evaluated seven precipitation products produced from multi-sources. These are the Africa Rainfall Climatology version 2 (ARC2; Novella and Thiaw, 2013), Climate Hazard Group InfraRed Precipitation with Stations version 2, (CHIRPS v2; Funk et al., 2015), Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals (GPM-IMERG; Huffman et al., 2019), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR; Ashouri et al., 2014), Rainfall Estimate version 2 (RFE2: Xie et al., 2007), Tropical Applications of Meteorology using SATellite and ground-based observations (TAMSAT; Tarnavsky et al., 2014), the Tropical Rainfall Measuring Mission (TRMM-3B43; Huffman et al., 2007), and TerraClimate (Abatzoglou et al., 2018). Spatial resolutions for these datasets vary between 0.04 (TerraClimate) and 0.25 (PERSIANN-CDR and TRMM-3B43).

2.2. Standardized Precipitation Index (SPI)

We used the Standardized Precipitation Index (SPI; McKee et al., 1993) to detect drought development over Ethiopia. SPI is acknowledged for its multiple advantages (e.g., temporal flexibility, spatial consistency and low data requirement) and it is widely used for operational actions (e.g., drought monitoring and forecasting) in many National Meteorological and Hydrological Agencies (NMHA) across the world (WMO , 2012; AghaKouchak et al., 2015). This method is also widely used to evaluate the performance of precipitation products in representing drought episodes in different parts of the world (e.g., Katiraie-Boroujerdy et al., 2016; Golian et al., 2019). The World Meteorological Organization (WMO , 2012) recommended the SPI as the best drought monitoring tool to be used for operational activities. SPI provides information on the probability of precipitation at different time scales (3-, 6-, 12-, and 24-month) (McKee et al., 1993). It compares the precipitation amount for a specific period (e.g., 3-month) with the precipitation totals from the same period for all the years considered in the historic records. For example, a 3-month SPI at the end of March accounts for the sum of precipitation for the months of January, February and March for a particular year and compares with the January to March precipitation total of all the years. Similarly, a 12-month SPI at the end of December compares the January to December (annual) precipitation total in that particular year with the January to December precipitation totals of all the years for a given geographical location (Shahid, 2008; WMO , 2012).

Monthly precipitation data for 20–30 years or longer is the input data needed to compute the SPI. Different distribution functions such as parametric (e.g., gamma; McKee et al., 1993) and non-parametric (e.g., Gringorten plotting position formula; Hao and AghaKouchak, 2013) can be used to calculate SPI values. In this study we used the gamma distribution to determine the probability density function that describes the long-term time series of precipitation at 3- and 12-month time scales and for each grid. Hence, the monthly precipitation amount for each grid and month was transformed into running time series of total precipitation for 3- and 12-month time series to determine the cumulative probability density function. Then the inverse normal function was applied to the cumulative probability distribution, with mean zero and standard deviation of one to generate SPI values. The SPI calculation produces positive and negative values; positive SPI values indicate wet condition, while negative values represent dry or drought condition. Both the wet and dry conditions can then be classified into different intensity classes as presented in Table 2. In this study, we considered three drought severity classes: Moderate, Severe and Extreme that accounted for 9.2%, 4.4% and 2.3% of the analysis time, respectively (McKee et al., 1993). These percentages are expected to occur from a normal distribution of the SPI time series as the SPI is standardized. A threshold is needed to define drought event/month in the SPI time series, hence a drought month is defined as when the SPI value becomes equal or less than −1.0.

2.3. Data evaluation methods

We used mapping and visual, and correlation analysis approaches to evaluate the performance of precipitation products in representing geographical coverage and severity of major drought episodes, and occurrence of drought events calculated at 3- and 12-month time scales over Ethiopia. The visual analysis is a quasi-subjective method that enables comparison of the ability of data products in capturing the magnitude and spatial patterns of four major historical drought events (1984, 2002, 2009 and 2015) against the reference data. We also calculated and mapped the spatial variability of annual rainfall departure from long-term mean (in percent) for the two major drought years (2009 and 2015). The spatial distribution of SPI values and percent departures of rainfall of the studied data products were compared against the Ethiopian gridded precipitation reference data. In addition to these, Pearson correlation coefficient was used to see the agreement between the evaluated and reference datasets for 3- and 12-month SPI time series. The SPI time series of the 3- and 12-month droughts at 126 stations and the Ethiopian gridded precipitation data products were used to verify occurrence or absence of drought events during the study period. Furthermore, the studied data products were compared against the Ethiopian gridded data for their skill in capturing the spatial distribution of long-term mean and standard deviation of precipitation amounts over Ethiopia.

Table 2
Classification of SPI values into different wet and dry magnitude conditions.

SPI value	Classification	SPI value	Classification
0 to 0.99	Near normal	0 to −0.99	Near normal
1 to 1.49	Moderately wet	−1 to −1.49	Moderate drought
1.5 to 1.99	Severely wet	−1.5 to −1.99	Severe drought
>2	Extremely wet	< −2	Extreme drought

3. Results

3.1. Performance of different datasets in detecting major drought episodes over Ethiopia

Four major drought events that occurred in Ethiopia in 1984, 2002, 2009 and 2015 were used to evaluate skills of the studied data products in capturing drought spatial patterns and severity. The 12-month SPI values for December were used to produce maps for each dataset and drought year to support visual comparisons (Fig. 2a-u). As shown by the reference maps (Fig. 2a), drought occurred over large parts of the country during those years, although the spatial distribution and severity varied between those drought years. For example, it was observed that the 1984 drought was the worst in terms of its large geographical coverage and severity, where large parts of the north-eastern, southern and central parts of the country were affected by severe and extreme intensity droughts. In contrast, the SPI values for some pocket areas in the western and central parts of the country were greater

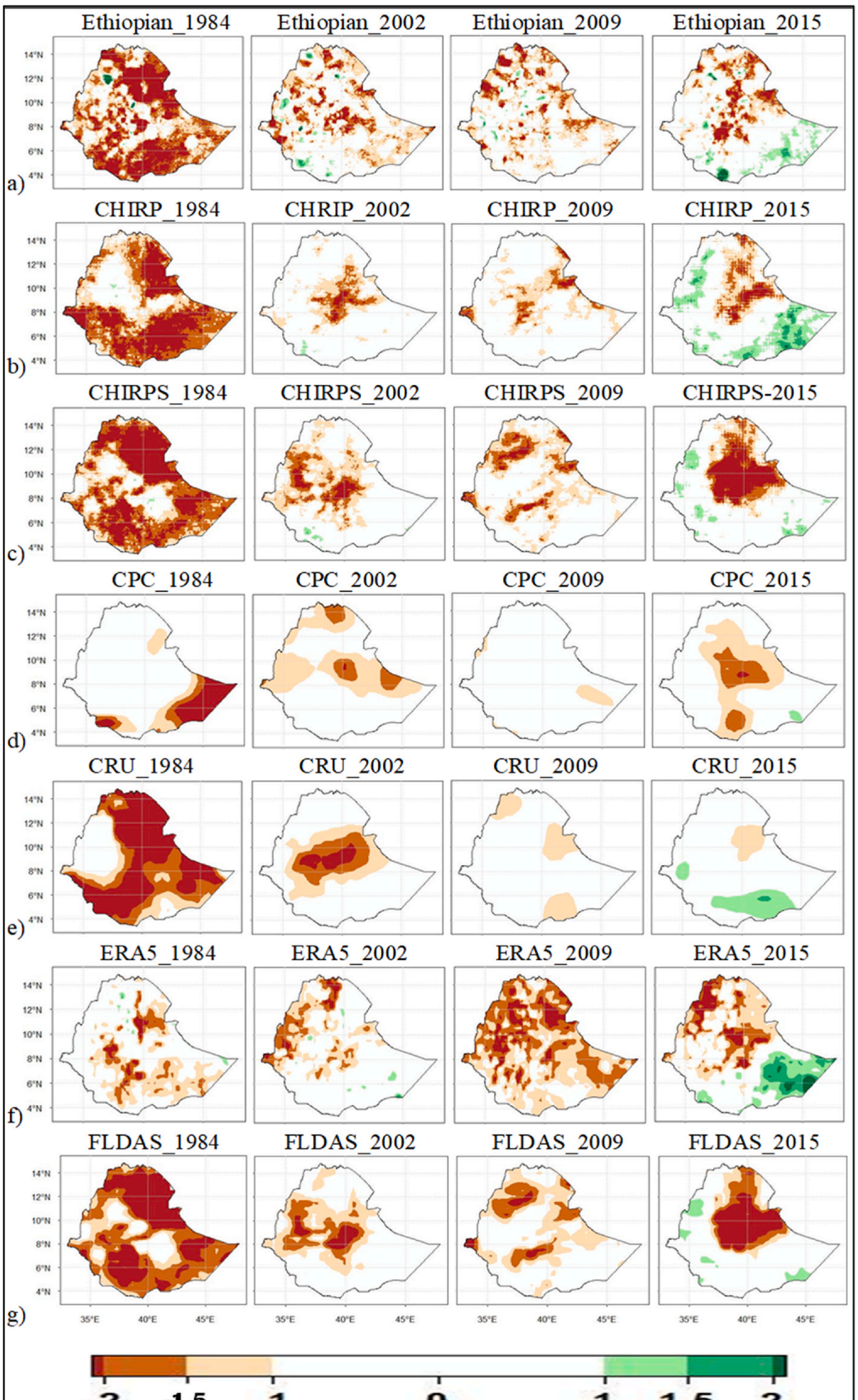

Fig. 2. a-u. Spatial pattern of 12-month SPI values for December of 4 four major drought episodes in Ethiopia from 20 precipitation data products.

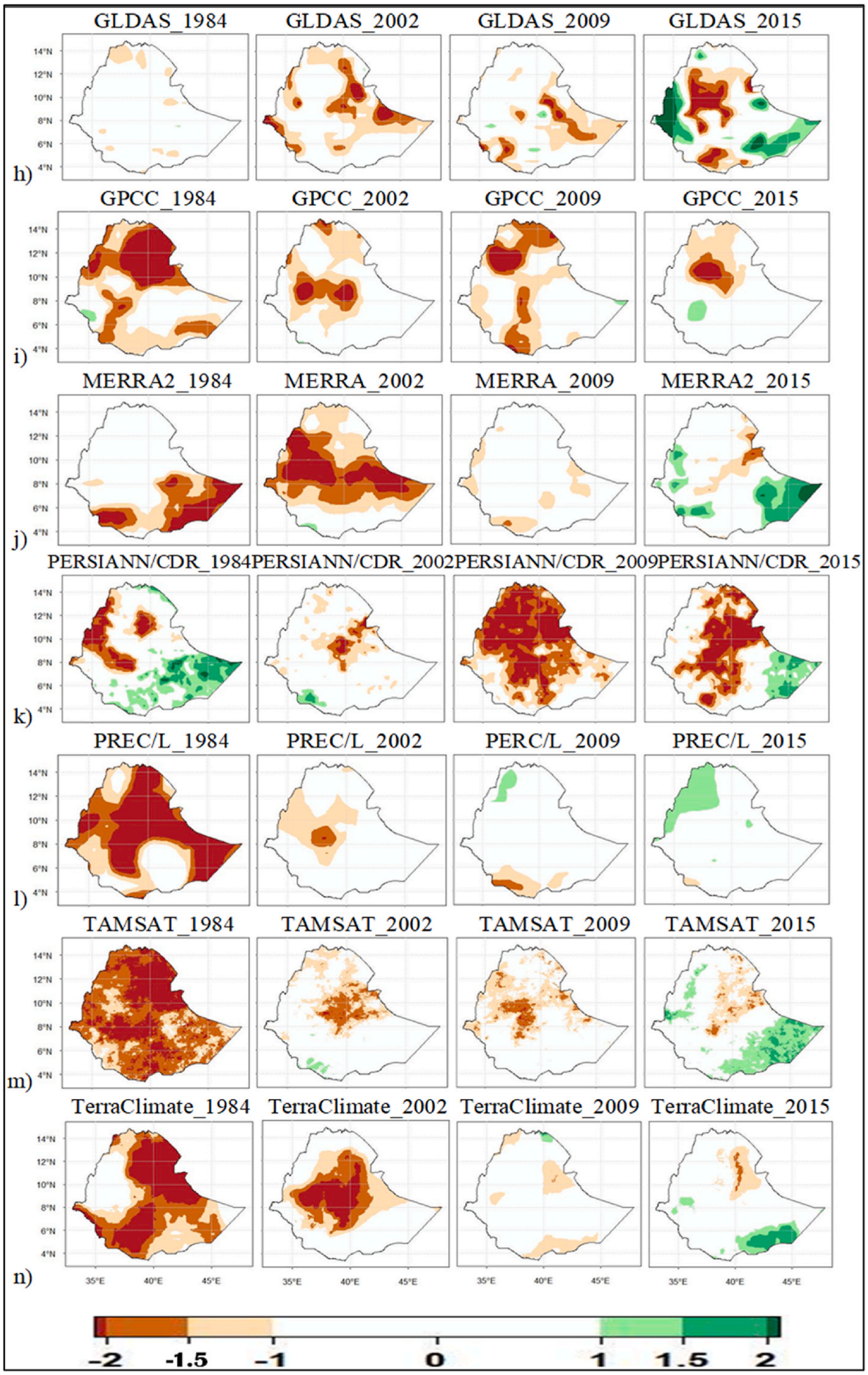

Fig. 2. (*continued*).

than −1. The 2002 and 2009 drought episodes also covered large parts of the country, but different from the 1984 and 2015 drought events; the SPI values did not show continuous geographical pattern for any part of the country. Hence, for both drought years (2002 and 2009), the distribution of SPI values < -1 showed irregular geographical pattern. In addition to this, large areas of the southwestern part of the country in 2002 and southeastern part in 2009 were drought free. Different from the previous three major drought years, the 2015 drought occurred over specific parts of the country (central and northeastern). As shown in Fig. 2a the southeast, south and western parts of the country were not affected by the 2015 drought. There are also few pocket areas in the northeastern part of the country, where the SPI value was greater than −1 for that drought year.

The performance of the 20 gridded precipitation products was inconsistent in representing those four major drought events (Fig. 2b-u). Although there were variations among the data products, eight out of the 13 precipitation products covering the 1980s better captured the 1984 drought. These were CHIRP, CHIRPS, CRU, FLDAS, GPCC, PREC/L, TAMSAT and TerraClimate. CHIRP and CRU in the northwestern and GPCC and PREC/L in the southeastern parts of Ethiopia failed to capture the same drought episode. On the other hand, PREC/L, TAMSAT and TerraClimate amplified the severity and geographical continuity of the

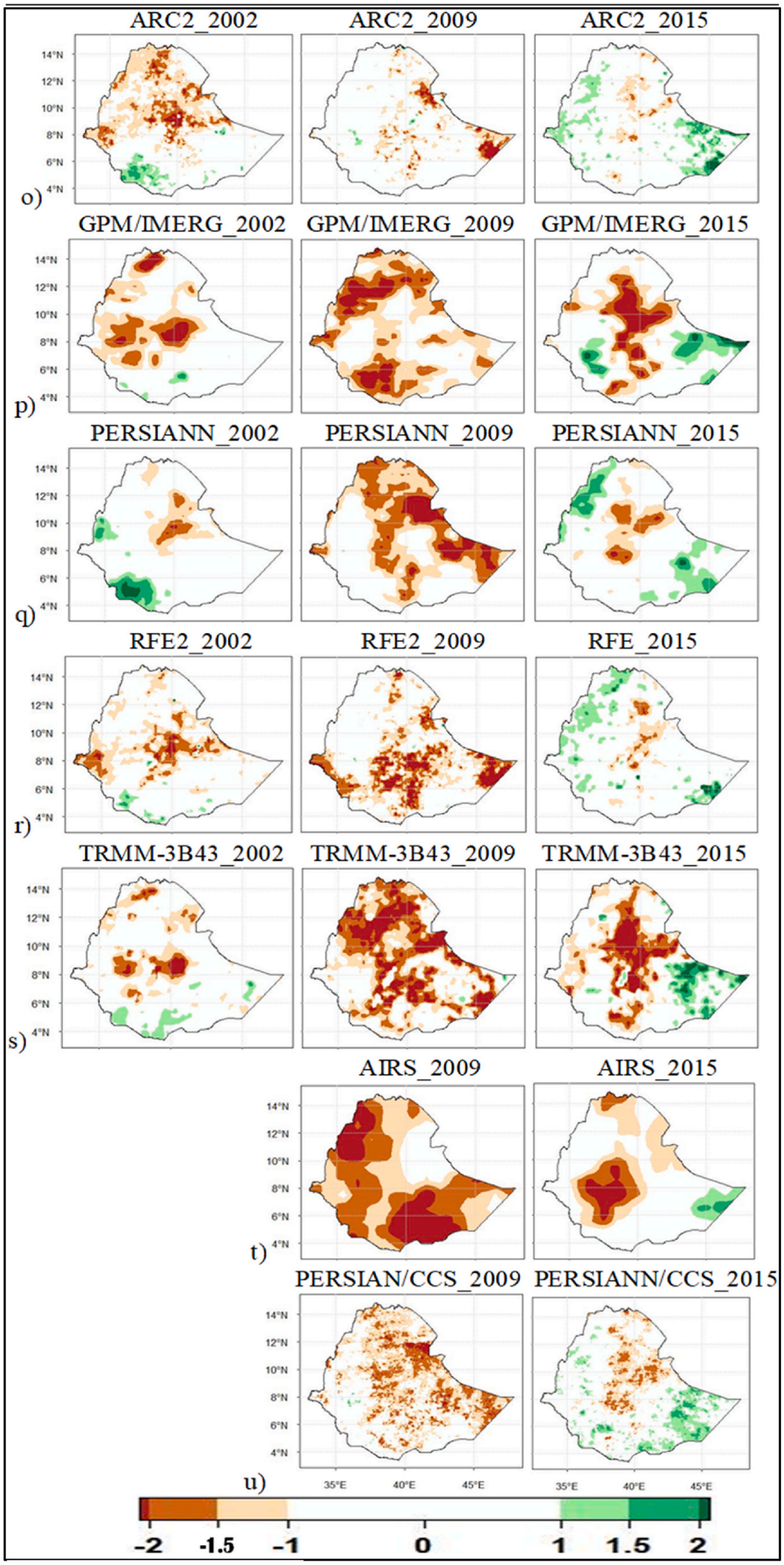

Fig. 2. (*continued*).

same drought event in the northwestern part of Ethiopia. Although, drought intensity was less emphasized, ERA5 also showed some skill in some parts of the country. CPC, GLDAS, MERRA2 and PERSIANN/CDR were poor in representing the 1984 drought. Almost all the products captured the 2002 drought at least for some part of the country. But, it was only MERRA2 that was able to represent the national scale drought occurrence, although severity of the drought was exaggerated for drought free pocket areas compared to the reference data. ARC2, CHIRPS, ERA5, GPCC and TAMSAT captured the 2002 drought for the central and northern parts of the country. Other five data products (CHIRP, CRU, FLDAS, GPM-IMERG, RFE2 and TerraClimate) had good performance over central Ethiopia. It is important to note that all the data products well represented the absence of drought in the southwestern part of the country in 2002.

The 2009 drought was captured by seven of the data products that include AIRS, CHIRPS, ERA5, GPM-IMERG, PERSIANN/CDR, PERSIANN/CCS and TRMM-3B43. Four of these (AIRS, ERA5, PERSIANN and TRMM-3B43) exaggerated intensity of the drought, compared to the reference data (Fig. 2a). FLDAS, GPCC, PERIANN, RFE2 and TAMSAT also showed good skills in representing the 2009 drought, although they missed capturing it over some specific areas. For example, GPCC and FLDAS failed to capture the drought in southeastern Ethiopia, and PERSIANN failed to capture it in the southwest, RFE2 in the northwest and TAMSAT in the south and southwest. Six products (CPC, CRU, GLDAS, MERRA2, PREC/L and TerraClimate) showed poor skill in capturing the 2009 drought. Nine products that include AIRS, CHIRP, CHIRPS, ERA5, FLDAS, PERSIANN/CCS, PERSIANN/CDR, TAMSAT and TRMM-3B43 accurately captured the 2015 drought that occurred over the central and northeastern parts of Ethiopia. In addition to these, CPC, GLDAS, GMP-IMERG and PERSIANN were able to detect the 2015 drought over central Ethiopia, while GPCC detected this drought over northern Ethiopia. ARC2 and RFE2 captured the geographical pattern of the 2015 drought, but the severity underestimated. CRU, MERRA2, PREC/L and TerraClimate showed the least performance in representing the 2015 drought, compared to the other data products. All products except CPC, GLDAS, PERSIANN/CDR, GPM-IMERG and TRMM-3B43 detected areas that were not affected by the 2015 drought (the western, southern and southeastern parts of the country).

3.2. Correlation analysis

We computed the Spearman's correlation coefficient between the 21 precipitation data products and the two reference datasets using SPI values generated at 3- and 12-month time scales. Although, the Ethiopian gridded rainfall data was not independent to most gauge observation considered in this study, we evaluated the agreement between the Ethiopian gridded precipitation product and gauge records obtained from 126 stations. We present the average correlation value in Fig. 3. The spatial distribution of the correlation coefficients is shown in Fig. 4a-r for the 3-month and Fig. 5a-r for the 12-month droughts. The average correlation values presented in Fig. 3 showed the presence of significant differences among the data products in representing drought events over Ethiopia. All the studied precipitation products showed better performance in capturing the occurrence of short-term drought (3-month) than the relatively long-term (12-month) drought occurrences. As expected, the maximum correlation value was found between the Ethiopian gridded data and the reference stations, both for the 3-month (0.94) and 12-month (0.86) time scales. Of the 20 precipitation data products, CHIRPS followed by FLDAS and GPCC showed better agreement with the reference data, both for 3- and 12-month time scales. The mean correlation values for these data products were 0.91, 0.90 and 0.87 for the 3-month SPI time series, respectively. For the 12-month SPI time series, the mean correlation coefficients were 0.74 for CHIRPS and FLDAS and 0.71 for GPCC. CPC, PERSIANN/CCS and GLDAS showed low correlations with the reference data for the 3-month SPI time series. The mean correlation values for these data products were 0.50, 0.56 and 0.58, respectively. AIRS, ARC2 and MERRA2 showed low agreement with the reference data for the 12-month SPI time series. Mean correlation coefficients for these data products were 0.28, 0.30 and 0.32, respectively.

Among the four gauge interpolated data products, GPCC was in closest agreement with the reference data, while CPC was the least, both for 3- and 12-month time scales. The mean correlation values of these data products were 0.87 and 0.5 for the 3month and 0.71 and 0.4 for the 12-month time scales, respectively. Out of the four satellite-only products, PERSIANN and CHIRP for the 3-month and PERSIANN for the 12-month time scales showed better agreement with the reference data. The mean correlation values were 0.73 for the former two and 0.52 for the latter. PERSIANN/CCS for the 3-month and AIRS for the 12-month SPI time series showed the least agreement with the reference dataset. The mean correlation values for these datasets were 0.56 and 0.28, respectively. Of the four reanalysis data products, FLDAS was in closest agreement with the reference data followed by ERA5 for both 3- and 12-month SPI time series. GLDAS and MERRA2 revealed the least agreement for the 3- and 12-month time series, respectively (Fig. 3). CHIRPS showed the best performance out of the eight multisource precipitation products, both for 3- and 12-month time scales. The mean correlation value for CHIRPS was 0.91 for the 3-month and 0.71 for the 12-month

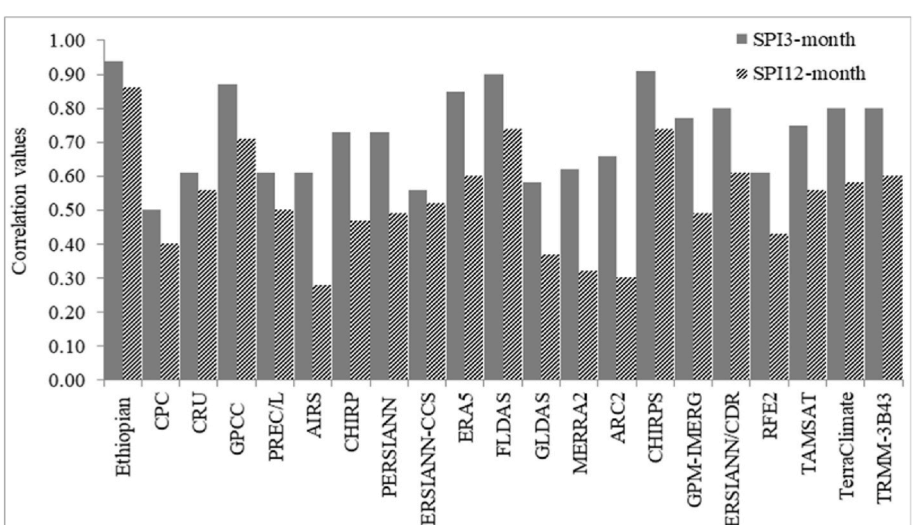

Fig. 3. Correlation coefficients for the 21 precipitation data products.

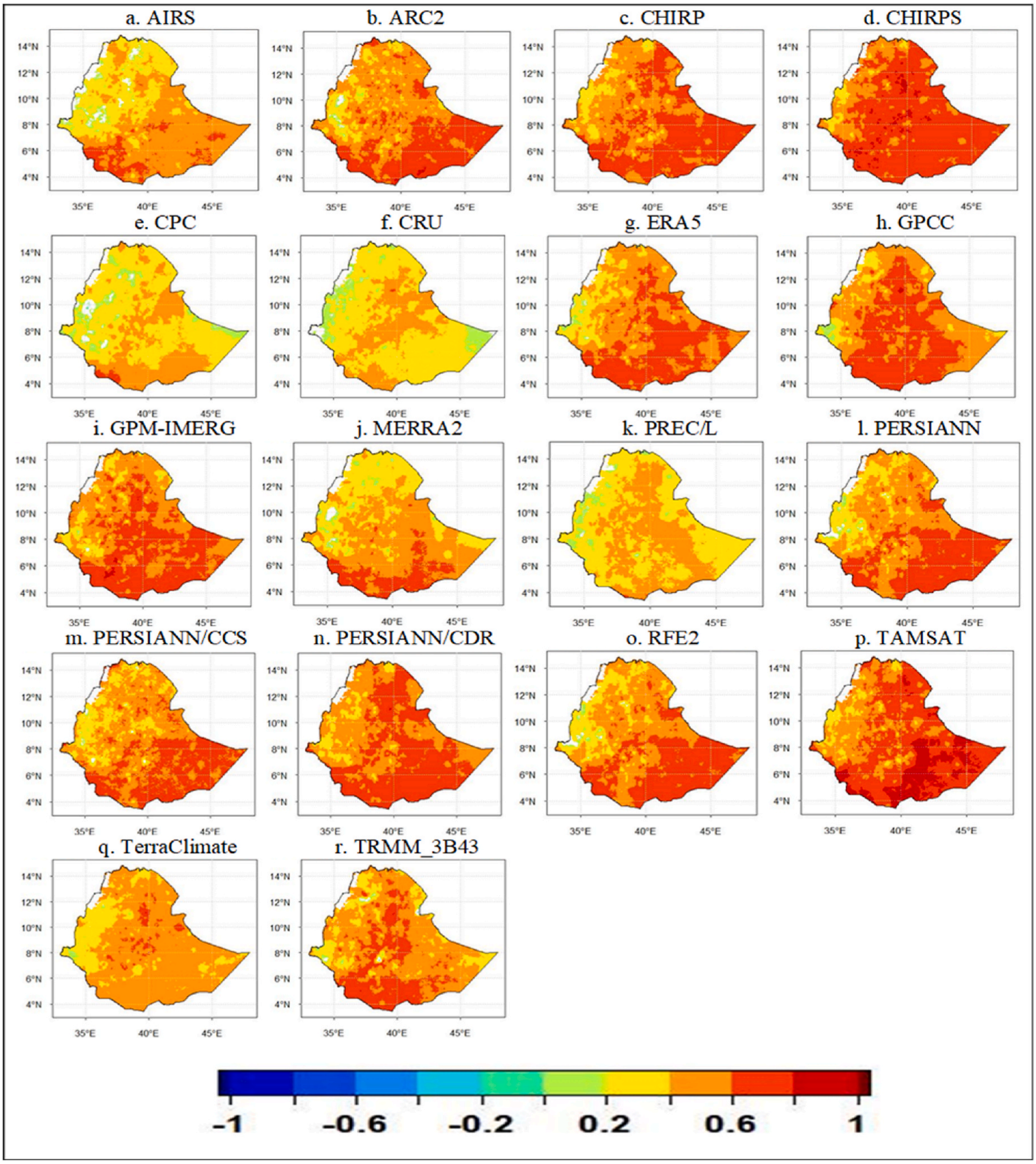

Fig. 4. a-r. Spatial patterns of correlation coefficients between the SPI values of 18 precipitation products and the Ethiopian gridded precipitation data for the 3-month time scale. Correlation values shown by different colors are statistically significant at 0.05 level. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

SPI time series. PERSIANN/CDR, TerraClimate and TRMM-3B43 precipitation products also showed better performance next to CHIRPS. The mean correlation coefficients for these data products were 0.80 for the 3-month SPI time series, and 0.61, 0.58, and 0.60 for the 12-month SPI time series, respectively. The correlation values of RFE2 (0.61) for the 3-month and ARC2 (0.30) for the 12-month SPI time series were the least, compared to the other eight multisource data products.

Fig. 4a-r shows the spatial pattern of correlations between the 3-month SPI values of the studied data products except for FLDAS and GLDAS and the Ethiopian gridded precipitation datasets. FLDAS and GLDAS were excluded because of their missing values for some grid cells in the northeastern part of the country. Consistent with the mean correlation values presented in Fig. 3, all data products showed better skill in representing the spatial patterns of SPI value generated at 3-month (Fig. 4a-r) than the 12-month (Fig. 5a-r) time scale for most data products. Although there are significant differences among data products in representing the spatial variability of SPI values, most data products showed stronger correlation over the southern and south-eastern and weaker values over the northwestern parts of the country. It is CHIRPS that showed the best agreement with the reference data over larger parts of the country, both for the 3-month (Figure 4d) and 12-month (Fig. 5d) time series. Following CHIRPS data TAMSAT, CHIRP, GPCC, GPM-IMERG, and PERSIANN/CDR had relatively better correlation values over extended area, compared to the other data products, for the 3-month SPI time series, while GPCC, GPM-IMERG and TAMSAT had better performance for the 12-month time series. CPC, CRU and

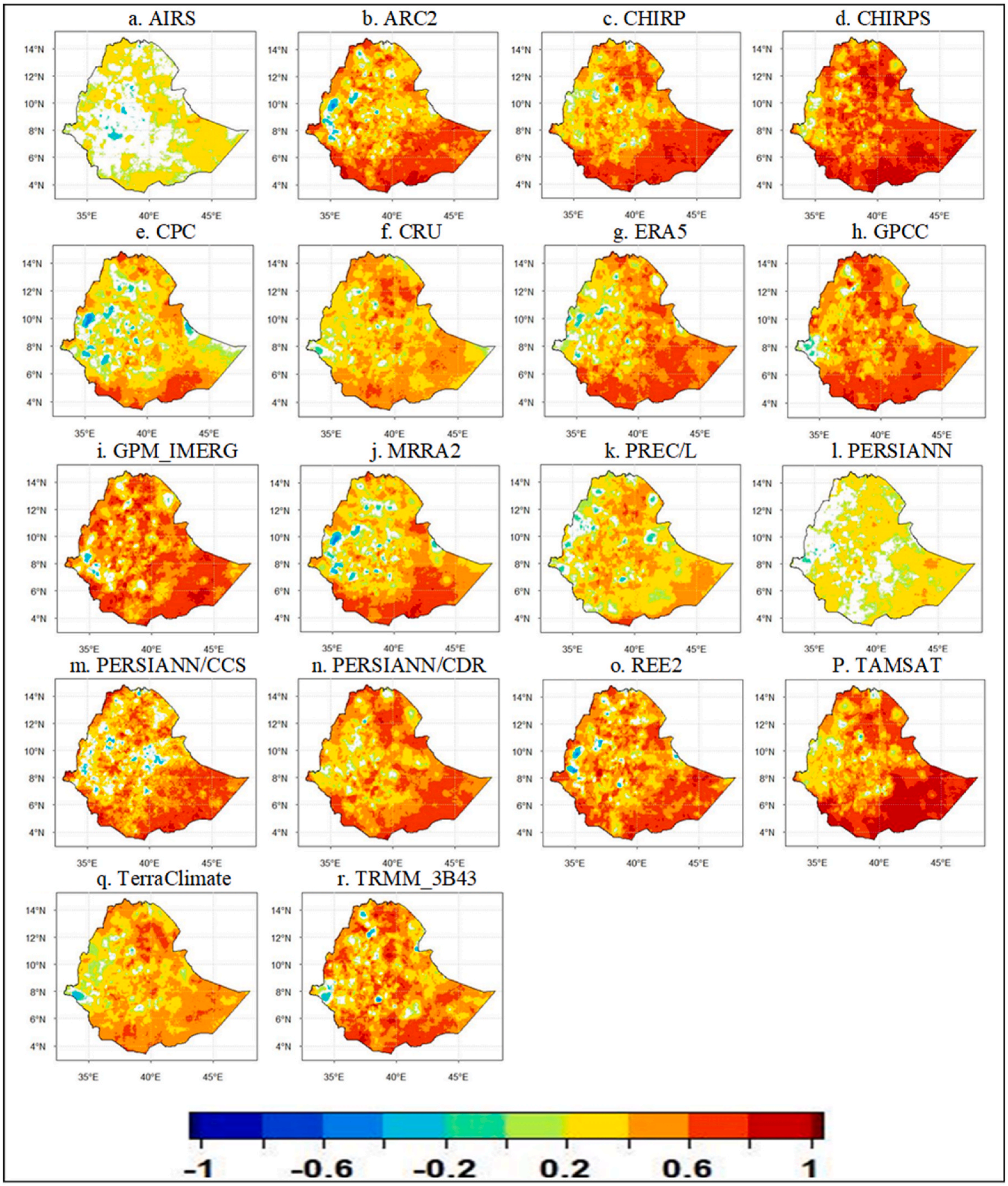

Fig. 5. a-r. Spatial distribution of correlation coefficients between the SPI values of 18 precipitation products and the Ethiopian gridded precipitation data for the 12-month time scale. Correlation values shown by different colors are statistically significant at 0.05 level. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

PREC/L for the 3-month and AIRS and PERSIANN for the 12-month time series showed weakest spatial correlations. It is important to note that most of the thermal infra-red derived products that include ARC2, CHIRP, RFE2, PERSIANN/CCS and TAMSAT show relatively similar and better performance over the southeastern lowland region, which suggests that these data products are better for drier and hotter lowland areas. AIRS stands out as very poor.

3.3. Percent of precipitation departure from the long-term mean for 2009 and 2015 drought years

We analysed annual rainfall departure from the long-term mean for selected two major drought years (2009 and 2015). Percent of annual rainfall deviation from long-term mean for 2009 is presented in Fig. 6a-u. As shown by the reference data (Fig. 6a), the total annual precipitation in 2009 was below normal over most part of Ethiopia. However, the negative precipitation departure was not the same across the country. The 2009 rainfall amount in drought affected areas (areas with SPI <=-1 in Fig. 2a) was less than −40% of the long-term mean. Most of these

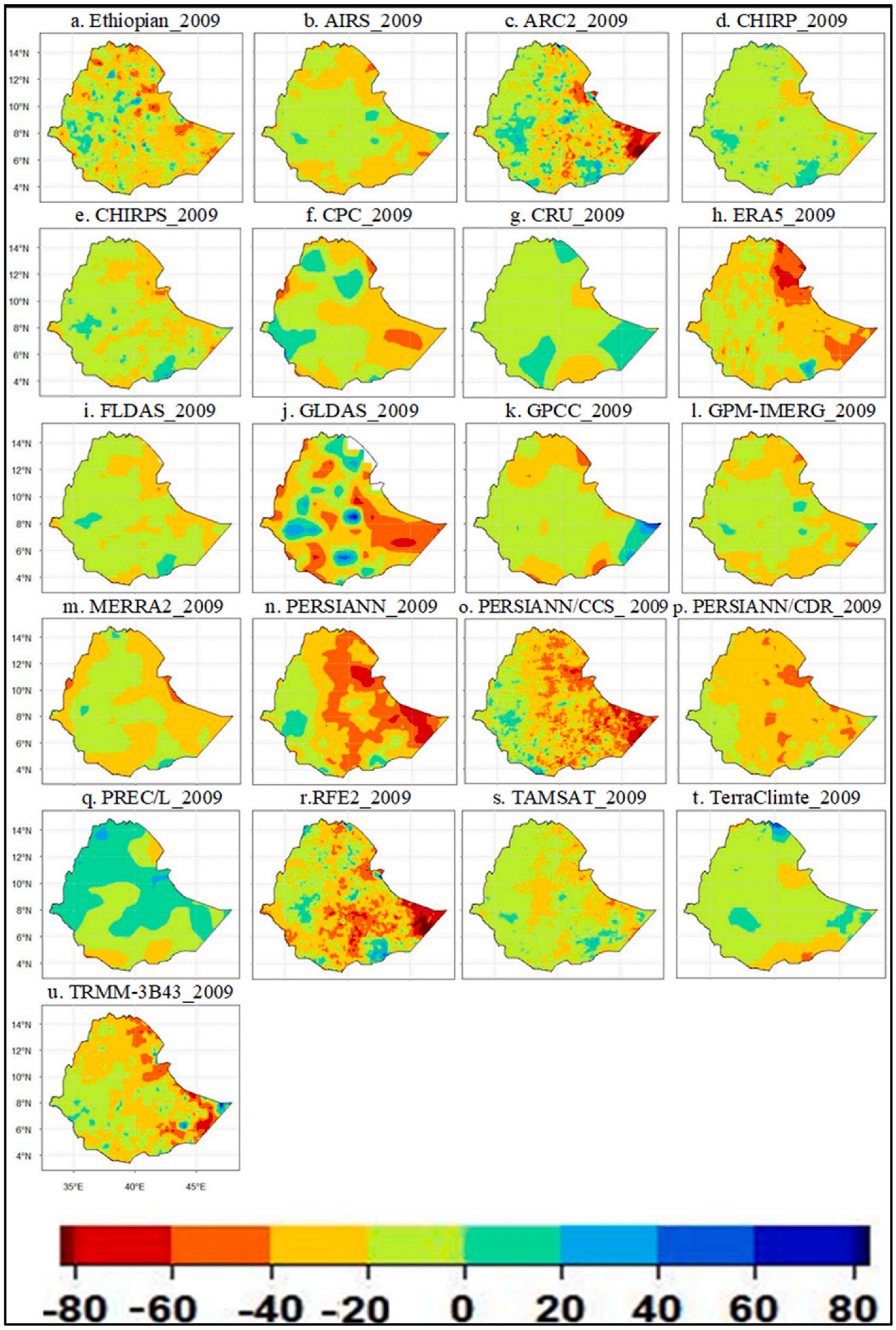

Fig. 6. a-u. Spatial distribution of annual rainfall departure from the mean in percent for 2009 drought year.

areas are found in the north, east and southeast Ethiopia. The negative rainfall departure was larger (less than −60% of the mean) for some areas and most of these areas are found in the northern and eastern parts of the country. In contrast, the 2009 rainfall was above normal for some pocket areas, which are located in the western and north-central parts of the country.

Although the degree of deviation is different between data products, all precipitation products, except PREC/L show negative anomalies for 2009 annual rainfall over large parts of the country. ARC2, CPC, GLDAS, PERSIANN/CCS, RFE2 and TRMM-3B43 precipitation products demonstrate better performance as compared to the reference data by capturing both the positive and negative rainfall deviations and its irregular spatial patterns. Different from the reference data shown in Fig. 6a, all these precipitation products exaggerate the rainfall deficit over southeast Ethiopia. In contrast, the rainfall deficit of less than −40% of the mean is observed in pocket areas by the reference data not well represented by ARC2 and RFE2 in the northwest, by ARC2, PERSIANN/CCS and TRMM-3B43 in the west and northwest and by GLDAS in the northeast Ethiopia. CPC also did not capture well the negative rainfall anomaly for the north-central highland of Ethiopia. On the other hand ERA5, PERSIANN and PERSIANN/CDR are relatively good in representing the ranges of negative anomalies, but these data did not represent well positive rainfall anomalies observed at pocket areas. Compared to the reference data, these data products exaggerate areas with rainfall deficit of less than −40% of the long-term mean. The other precipitation products that include AIRS, CHIRPS, FLDAS, GPCC, GPM-IMERG, MERRA2 and TAMSAT did not effectively capture the rainfall anomalies of less than −40% of the mean and those pocket areas that have positive rainfall anomalies. Similarly, areas with rainfall deficit of less than 20% of the long-term mean were not adequately captured by CHIRP and TerraClimate precipitation products.

Fig. 7a-u shows the spatial distribution of rainfall anomalies for the 2015 drought year. As shown by the reference data (Fig. 7a), the 2015 rainfall was below normal over Ethiopia, except for southern, southeastern and pocket areas in the north-central highland parts of the country. Negative rainfall anomalies of less than −20% of the long-term mean occurred in the central and northeastern parts of Ethiopia. The results show that almost all precipitation data products (Fig. 7b-u), except PREC/L show negative rainfall anomalies for drought affected areas of the central and northeast Ethiopia (Fig. 2a), though there are variations between the different data products. ARC2, CHIRPS, CPC, ERA5, FLDAS, GLDAS, MERRA2, the three PERSIANN data, FRE2 and TRMM-3B43 precipitation products show better performance by capturing the negative rainfall anomalies as shown by the reference data. AIRS, CHIRP, GPM-IMERG and TAMSAT are also relatively good as they are able to represent rainfall deficit of up to 40% of the long-term mean. In contrast, PREC/L followed by CRU, GPCC and TerraClimate did not effectively represent areas that received less than −20% of the long term mean. Similarly, some of the data products that showed better performance in the other areas (e.g. ARC2, CHIRP, CHIRPS, FLDAS, GLDAS, MERRA2, PERSIANN, PERSIANN/CCS, RFE2 and TAMSAT) missed to represent the negative rainfall anomalies for the western and northwestern parts of the country. On the other hand, it is important to mention that most of the precipitation products showed good performance in representing the positive rainfall anomalies observed in the reference data for the southern and southeastern parts of Ethiopia, although the positive rainfall anomalies for some data products (e.g. ARC2, ERA5, GLDAS, MERRA2 and TAMSAT) is over estimated. In contrast, the other data products that include CPC, GLDAS, PERSIANN/CDR, PREC/L, RFE2 and TRMM-3B43 were not good in capturing the

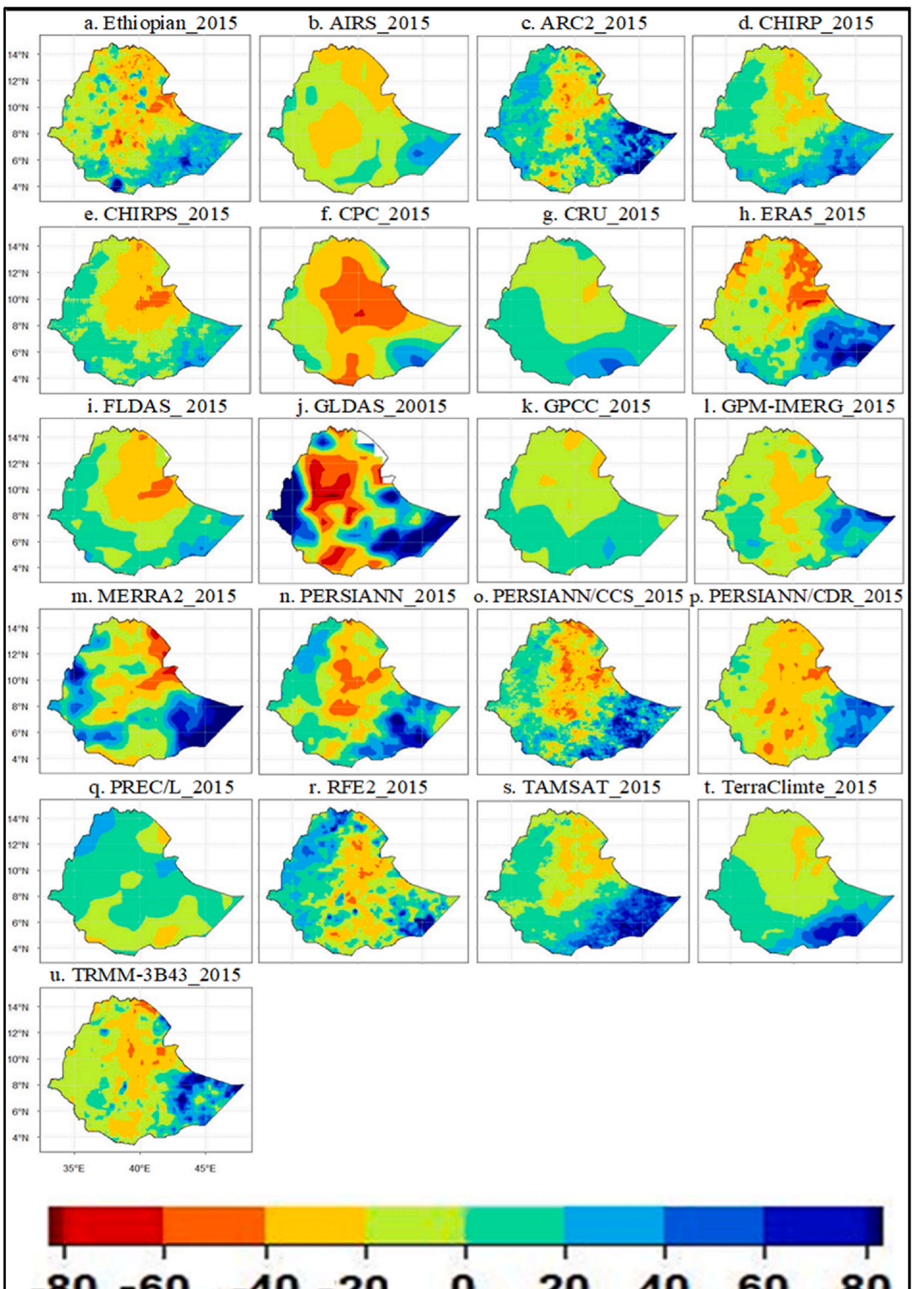

Fig. 7. a-u. Spatial distribution of annual rainfall departure from the mean in percent for 2015 drought year.

positive rainfall anomalies for the southern part of the country.

3.4. Spatial variability of mean annual and standard deviation of precipitation products

We further evaluated performance of the data products for their capacity in representing the spatial patterns of mean annual and standard deviation of precipitation over Ethiopia. As shown by the reference data (Fig. 8a), the long-term mean annual rainfall varies between less than 500 mm over the northeastern and southern lowlands and greater than 2 500 mm over the western and southwestern highlands. Generally, the mean annual rainfall amount decreases as one moves from western and southwestern highlands to the northeast and southeastern parts of the country. It is also observed that the proportion of areas that have high mean rainfall amount (>2 500 mm) is relatively smaller as compared to areas that have lower (<500 mm) mean annual rainfall amount.

The results in Fig. 8a-u shows that most of the precipitation products are able to represent the general spatial locations for high and low mean annual rainfall. However, they showed quite different performance and or biases in representing the mean annual rainfall amount, particularly for areas (western and southwestern Ethiopia) that receive high rainfall amount. Generally, CHIRP followed by CHIRPS and TAMSAT rainfall products show better agreement to the reference data in capturing the spatial distribution of mean annual rainfall over Ethiopia. In contrast, AIRS, ARC2, CPC, CRU, ERA5 and RFE2 precipitation products show least performance in representing mean annual rainfall of greater than 1 500 mm for the western and southwestern parts of the country. These data products underestimated the mean annual rainfall of between 500 and 1500 mm over these areas. Similarly, the other data products that include FLDAS, GLDAS, GPCC, GPM-IMERG, the three PERSIANN datasets, PREC/L, TerraClimate and TAMSAT precipitation products did

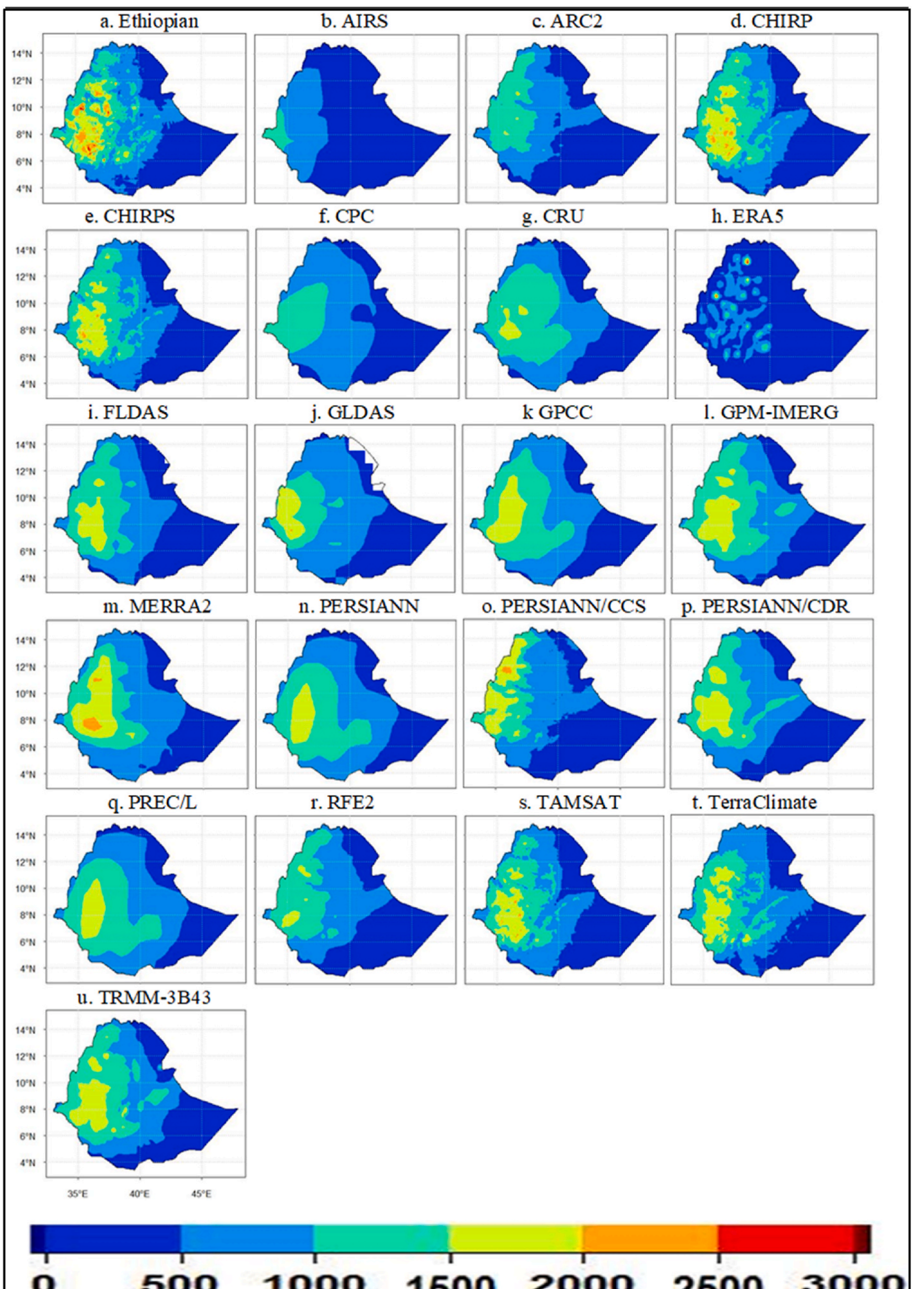

Fig. 8. a-u. The spatial distribution of mean annual rainfall for 21 data products.

not capture the mean annual rainfall of greater than 2000 mm, indicating that these products underestimate the mean annual rainfall of between 500 and 1 000 mm. All the studied data products showed good performance in representing the lowest mean annual rainfall (<500 mm) over the northeast and southeastern parts of the country.

The spatial distribution of standard deviation for annual rainfall time series for all precipitation products is shown in Fig. 9a-u. As shown by the reference data (Fig. 9a), the standard deviation of annual rainfall is relatively higher over the highland part of the country, where the mean annual rainfall is relatively higher, and lower over lowland parts, where mean annual rainfall is low. The standard deviation over the highland regions, particularly in the southwest Ethiopia varied between 200 mm and 900 mm. However, high rainfall deviation (>500 mm) is observed for isolated pocket areas. In contrast, the standard deviation over lowland parts, particularly the northeast and southeast Ethiopia is not more than 100 mm.

The spatial distribution of standard deviation for the studied data products is shown in Fig. 9b-u. The result shows that GLDAS followed by MERRA2 and CPC are relatively better in representing the wider ranges of rainfall deviation from long-term means. However, they did not show the same irregular pattern shown by the reference data. In addition to this, the rainfall deviation shown by these three precipitation products for the northern and northwestern Ethiopia is not in agreement with the reference data. The standard deviation of annual rainfall in these areas is over estimated by 300–600 mm. Furthermore, it is only GLDAS that is able to show areas with standard deviation of greater than 600 mm. ARC2, PERSIANN/CCS and RFE2 precipitation products are also able to represent areas that have standard deviation of less than 400 mm in the western part of Ethiopia. Different from the reference data, these data products overestimate the rainfall deviation for the northwestern part of Ethiopia. The standard deviation for the other precipitation products is less than 200 mm. It is less than 100 mm over most parts of the country

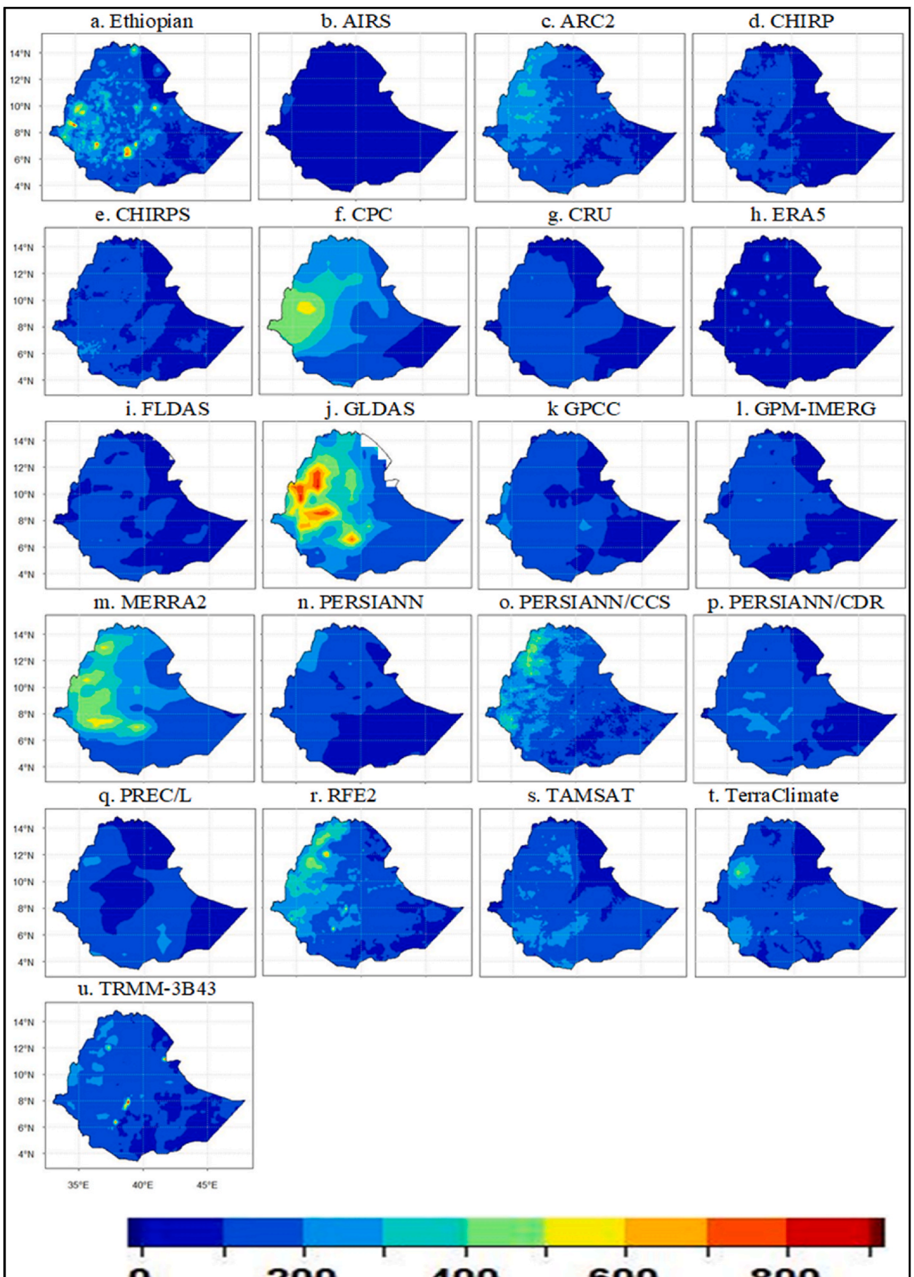

Fig. 9. a-u. The spatial distribution of standard deviation for annual rainfall for 21 data products.

for AIRS and ERA5. On the other hand, most of the precipitation data products show better skill in capturing the low standard deviation observed in the northeastern and southeastern lowland parts of the country.

4. Discussion and conclusion

We evaluated 20 global and regional scale precipitation products for their drought monitoring performance over Ethiopia. These precipitation products are different in terms of spatial resolution (0.04°–1°), record length and the techniques of data construction: 1) gauge-interpolated (CPC, CRU, GPCC and PREC/L), 2) retrieved from satellite observation (AIRS, CHIRP, PERSIANN and PERSIANN/CCS, 3) reanalysis (ERA5, FLDAS, GLDAS and MERRA2) and 4) merged from various sources (ARC2, CHRIPS, GPM-IMERG, PERSIANN/CDR, RFE2, TAMSAT, TerraClimate and TRMM-3B43). We used the Ethiopian gauge-satellite merged precipitation data at 0.04° spatial resolution and in-situ records of 126 stations for ground truthing. We used a quasi-objective (visual inspection) and the Spearman's correlation analysis to evaluate the data products.

The results showed that the studied precipitation products differ in representing the geographical distribution and severity of drought episodes and the climatology of the study area as shown by the long-term mean and the standard deviation of the data series. CHIRPS followed by FLDAS and GPCC showed relatively better and comparable with each other performance in representing the spatial distribution of major drought events, as also indicated by the Spearman's correlation coefficient (Table 2). CHIRP, ERA5, PERSIANN/CDR, TAMSAT and TRMM-3B43 showed better skill for two out of the four major drought years considered. These datasets also showed some skill in capturing the remaining two major droughts for some part of the country. In addition to these, AIRS, PERSIANN/CCS and TRMM-3B43 that covered the period 2000–2018 accurately represented two out of the three droughts. CRU, PREC/L, MERRA2, GPM-IMERG and TerraClimate captured well only one of the four droughts, while CPC and GLDAS showed the worst performance. Regarding rainfall departures from long-term mean for the 2009 and 2015 drought years, all the studied precipitation products except PREC/L showed negative anomalies for most drought affected areas, but showed different performance in capturing areas with rainfall departures of less than −20% of the mean. Some of the data products that include ARC2, ERA5, GLDAS, the three PERSIANN datasets, RFE2 and TRMM-3B43 for both drought years and CHIRPS, CPC, FLDAS and MERRA2 for the 2015 drought year showed better performance in capturing areas with rainfall departures of <40% from the long-term mean.

On the other hand, most of the studied data products are in agreement with the reference data in representing the geographical locations for high (western and southwestern Ethiopia) and low (northeastern and southeastern lowlands) mean annual rainfall areas. The mean annual rainfall is underestimated by most data products over high rainfall regions (western and southwestern Ethiopia). Only CHIRP followed by CHIRPS, MERRA and TAMSAT precipitation products are relatively better in representing areas with mean annual rainfall of greater than 1 500 mm. In contrast, AIRS, CRU, CPC, ERA5, and RFE2 precipitation products showed weak performance. The standard deviation of annual rainfall was not captured satisfactorily by most of the data products and it was generally lower than that of the reference data except for GLDAS, CPC, MERRA2, PERSIANN/CCS, RFE2 and ARC2. This indicates that the data products were better in representing annual rainfall amounts around the mean, but are not good at estimating the extremes; i.e., high and low annual rainfall amounts.

Performance of precipitation products in drought detection is generally determined by their capability in representing amount and geographical distribution, and its ability to capture high and low intensity rainfall events (Le Coz and van de Giesen, 2020). Precipitation products that are recommended for drought monitoring and operational activities are those that have good representation of small rainfall amounts, low overestimation, and good representation of spatial distribution of low precipitation amount and events (Maidment et al., 2014; Le Coz and van de Giesen, 2020). Previous research works have reported a complex picture of performance for different rainfall products (Dinku et al., 2007, 2009, 2011, 2018; Diro et al., 2009; Hirpa et al., 2010; Worqlul et al., 2014; Bayissa et al., 2017; Sahlu et al., 2017). This is reconfirmed by the results of our study. The underestimation of mean annual rainfall found by our study for most data products over the highland areas is also in agreement with those previous studies. On the other hand, Dinku et al. (2008) found monthly precipitation data of CRU, CPC and GPCC to be in good agreement with selected reference stations over Ethiopia for the period 1981–2000, but this was not supported by our finding. In this study, it was only GPCC that showed superior capacity in capturing droughts among the gauge interpolated data products. Similarly, many recent studies that include Bayissa et al. (2017), Ayehu et al. (2018), Dinku et al. (2018) and Belay et al. (2019) reported that CHIRPS precipitation data outperforms in representing the various rainfall characteristics considered at multiple time scales (daily, decadal, monthly and seasonal) than many other satellite and model data products that they used for their study. Our study also found that CHIRPS is a relatively good precipitation data product that showed better performance in capturing drought development followed by FLDAS and GPCC.

Dinku et al. (2011, 2014) note that in the mountainous east African region, the density of observation stations is low and unevenly distributed, hence the estimated areal average rainfall through interpolation from point data may not provide accurate estimate. Even worse, observation points have declined in number over this and other parts of Africa after the second half of the 20th century (Harris et al., 2014; Dinku et al., 2018; Le Coz and van de Giesen, 2020). For example, the average number of stations used to produce the GPCC v7 data over Africa was reduced from over 3 000 in the 1960s to less than 500 in 2000s (Dinku et al., 2014; Zhao and Ma, 2019). Similarly, the average number of stations used to produce CRU precipitation data decreased from around 2000 in the 1960s to less than 200 in the post 2000s (Harris et al., 2014). For Ethiopia, the number of stations used by CPC, CRU and GPCC declined, on average, from 20 per 2.5° grid box in 1981 to less than 5 in the post 2000 (Dinku et al., 2008). The average number of stations used to construct PREC/L precipitation at the global scale declined from about 17,000 in 1970 to about 2000 in 1996, and no new station was added for interpolation since 1997 (Chen et al., 2002). The implication is that the representativeness of gauge interpolated data products to various rainfall characteristics decreases with the decreasing station numbers used for the interpolation. It is also evident in our study that except CPC, the other three gauge-interpolated precipitation products (CRU, PREC/L and TerraClimate) show better performance in capturing the magnitude and spatial patterns of the 1984 drought than the 2002, 2009 and 2015 droughts. Besides, the quality of data and the different interpolation methods are also sources of discrepancies among gauge interpolated data products (Le Coz and van de Giesen, 2020).

Performance evaluation of satellite based and satellite-gauge merged precipitation products showed discrepancies among previous studies as well. For example, according to Dinku et al. (2007) TRMM-3B43 and TAMSAT showed reasonably well performance over Ethiopia, while ARC2 and RFE2 revealed poor performance. Similarly, Young et al. (2014) reported that TAMSAT has good performance in detecting rainy events over Ethiopia, while it underestimated rainfall amounts. PERSIANN showed reasonable accuracy over lowland regions, but underestimated over highland area (Romilly and Gebremichael, 2011). Gebremichael et al. (2017) on the other hand has found good performance for CHIRPS, TRMM-3B43 and RFE2 precipitation products in the Tekezie-Atbara River Basin, northern Ethiopia. These study also indicated that these satellite based products underestimated the rainfall amount over highland areas (>2 500 m a.s.l) and overestimated over lowland regions by 35%. Many recent studies that include Bayissa et al.

(2017); Gebremichael et al. (2017); Ayehu et al. (2018); Dinku et al. (2018); Lemma et al. (2019) and Belay et al. (2019) identified CHIRPS as the best satellite based precipitation product over Ethiopia. For example, Dinku et al. (2018) reported that CHIRP and CHIRPS precipitation products are significantly superior than ARC2 and TAMSAT precipitation products at the dekadal and monthly time scales, while TAMSAT is better in representing daily precipitation. Lemma et al. (2019) identified CHIRPS followed by TRMM-3B43 to have good performance in representing rainfall amount and regimes for three wet seasons (June–September, February–May and September–November). According to this study, CMORPH, TAMSAT and ARC2 showed moderate performance. For Upper Blue Nile Basin, CHIRPS showed better performance over TAMSAT and ARC2 precipitation products (Ayehu et al., 2018). Most of the previous studies highlight that most satellite products showed a tendency to underestimate and overestimate rainfall amounts in highland and lowland parts of Ethiopia, respectively (Bitew and Gebremichael, 2009; Hirpa et al., 2010; Dinku et al., 2011; Young et al., 2014; Gebremicael et al., 2017).

Poor performance (large underestimation) was reported for ARC2 and RFE2 for the daily and dekadal rainfalls across the country (Dinku et al., 2011). Dinku et al. (2008, 2011) and Young et al. (2014) argued that satellite rainfall retrieval algorisms have difficulty to accurately capture low convection warm-cloud orographic rainfall formation process and difficulty of setting accurate thresholds to distinguish rain-giving from non-rain giving clouds. On the other hand, the overestimation of rainfall over lowland areas is attributed to evaporation of sub-cloud before giving rainfall in the region (Dinku et al., 2008) and cold temperature on the top of cirrus non-raining clouds (Young et al., 2014). In addition to these, satellite-based products differ each other by the type of data input (whether it is IR, PWM or the combination of the two or with gauges) used by their algorithms deriving the final precipitation estimates. As consequence, the satellite-based product error can be partially attributed to the error in the retrieval algorithm (i.e., estimating precipitation from the sensors measurements) and partially due to the merging algorithm (i.e., combining the different estimates in the final precipitation estimate) (Le Coz and van de Giesen, 2020). In our study, CHIRPS data out performs all the satellite based precipitation products in representing major drought events, drought frequency and in correlation values. TRMM-3B43, TAMSAT, PERSIANN/CDR, GPM-IMERG, CHRIP and PERSIANN showed medium comparable performance in terms of correlation values.

The other possible reason for data performance discrepancies as shown in Fig. 10a and b is variation in the spatial resolution of the precipitation products. Most precipitation data products with relatively higher resolution (<0.3°) had better performance compared to precipitation products which had relatively low resolution, both for 3- and 12-month drought time series.

In conclusion, our results demonstrate the need for comprehensive evaluation of global and regional precipitation data products before using them for specific research or operational activities. Performance evaluation evidently provides important information and guidance for selecting the most appropriate product specific applications such as for drought monitoring by policy and development communities. A follow up paper (Degefu and Bewket, submitted to *Weather and Climate Extremes*), has used statistical indices that include Probability of Detection (POD), Missing Rate (MR), False Alarm Ratio (FAR) and Critical Success Index (CSI) to undertake verification of how these data products could detect the 3- and 12-month drought episodes over Ethiopia.

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis and first draft writing were performed by Mekonnen Adnew Degefu. Research project management, supervision, second draft wring and edition were made by Woldeamlak Bewket. Yosef Amha has engaged in conceptualization of the paper, visualization, edition and validation activities. All authors read and approved the final manuscript.

Submission declaration and verification

The submission is the independent work of the authors. It has not been submitted and not published or accepted for publication, and is not under consideration for publication, in another journal or book. The submission has been approved by all relevant authors, and all persons entitled to authorship have been so named. All authors have seen and agreed to the submitted version of the manuscript.

Availability of data and material

All the data inputs in for our study, except for the reference (gridded and point precipitation) data products obtained from globally open data sources, and one can easily access using the given data portal links in the paper. However, the reference data that we obtained from National Meteorological Agency are confidential by policy. But we can submit the processes data upon request.

Ethics approval

Not applicable to this manuscript as there was no potential conflict of interest, not involved human/or animals and no other participant that need informed consent.

Consent to participate

Not applicable to this manuscript as this study did not involve human participants that need informed consent.

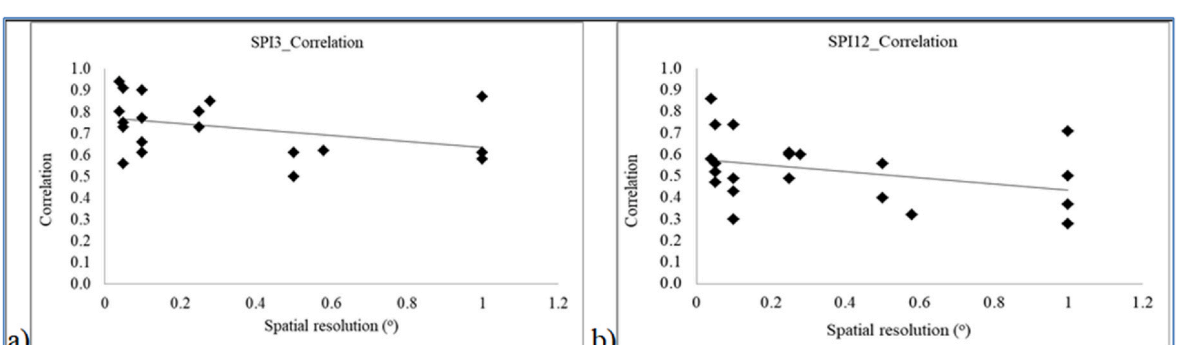

Fig. 10. a & b. Performance of precipitation data in detecting the a) 3- and b) 12-month drought events as a function of products spatial resolutions.

Declaration of competing interest

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

Acknowledgements

This work was supported through the Climate Research for Development (CR4D) Postdoctoral Fellowship [CR4D-19-19] implemented by the African Academy of Sciences (AAS) in partnership with the United Kingdom's Department for International Development (DfID) Weather and Climate Information Services for Africa (WISER) programme and the African Climate Policy Center (ACPC) of the United Nations Economic Commission for Africa (UNECA). Statements made and views expressed in this work are solely the responsibility of the authors. We also acknowledge the National Meteorological Agency of Ethiopia for providing the monthly precipitation products for the study.

References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci. Data 5, 170191.
AghaKouchak, A., Farahmand, A., Melton, F.S., Teixeira, J.M., Anderson, C., Wardlow, B. D., Hain, C.R., 2015. Remote sensing of drought: progress, challenges and opportunities. Rev. Geophys. 53, 452–480.
Albergel, C., Dutra, E., Munier, S., Calvet, J., Munoz-Sabater, J., de Rosnay, P., Balsamo, G., 2018. ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? Hydrol. Earth Syst. Sci. 22, 3515–3532.
Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2014. PERSIANN–CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Am. Meteorol. Soc. 96, 197–210.
Aumann, H.H., Gregorich, D.T., Gaiser, S.L., Chahine, M.T., 2004. Application of Atmospheric Infrared Sounder (AIRS) Data to Climate Research. Proc. SPIE 5570, Sensors, Systems, and Next-Generation Satellites VIII, (4 November 2004).
Ayehu, G.T., Tadesse, T., Gessesse, B., Dinku, T., 2018. Validation of new satellite rainfall products over the upper Blue Nile Basin, Ethiopia. Atmos. Meas. Tech. 11, 1921–1936.
Bayissa, Y., Tadesse, T., Demisse, G., Shiferaw, A., 2017. Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia. Rem. Sens. 9, 669.
Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21, 6201–6217.
Belay, A.S., Fenta, A.A., Yenehun, A., Nigate, F., Tilahun, S.A., Moges, M.M., Dessie, M., Adgo, E., Nyssen, J., Chen, M., Van Griensven, A., Walraevens, K., 2019. Evaluation and application of multi-source satellite rainfall product CHIRPS to assess spatio-temporal rainfall variability on data-sparse western margins of Ethiopian Highlands. Rem. Sens. 11, 2688.
Chen, M., Xie, P., Janowiak, J.E., Arkin, P.A., 2002. Global land precipitation: a 50-yr monthly analysis based on gauge observations. J. Hydrometeorol. 3, 249–266.
Degefu, M.A., Rowell, D.P., Bewket, W., 2017. Teleconnections between Ethiopian rainfall variability and global SSTs: observations and methods for model evaluation. Meteorol. Atmos. Phys. 129, 173–186.
Dinku, T., Block, B., Sharoff, J., Thmoson, M., 2014. Bridging critical gaps in climate services and applications in Africa. Earth Perspect. 1, 1–15.
Dinku, T., Ceccato, P., Connor, S.J., 2011. Challenges of satellie rainfall estimation over mountainous and arid parts of East Africa. Int. J. Rem. Sens. 32, 5965–5979.
Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.J., Ropelewski, C.F., 2007. Validation of satellite rainfall products over East Africa's complex topography. Int. J. Rem. Sens. 28, 1503–1526.
Dinku, T., Connor, S.J., Ceccato, P., Ropelewski, C.F., 2008. Comparison of global gridded precipitation products over a mountainous region of Africa. Int. J. Climatol. 28, 1627–1638.
Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., Ceccato, P., 2018. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Q. J. R. Meteorol. Soc. 144, 292–312.
Dinku, T., Hailemariam, K., Maidement, R., Tarnavsky, E., Connor, S.J., 2013. Combined use of satellite estimates and raingauge observations to produce high-quality historical rainfall time series over Ethiopia. Int. J. Climatol. 34, 2489–2504.
Diro, G.D., Grimes, D.I.F., Black, E., O'Neill, A., Pardo-Iguzquiza, E., 2009. Evaluation of reanalysis rainfall estimates over Ethiopia. Int. J. Climatol. 29, 67–78.
Duan, Z., Liu, J., Tuo, Y., Chiogna, G., Disse, M., 2016. Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. Sci. Total Environ. 573, 1536–1553.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, J., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2, 150066.
Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30, 5419–5454.
Golian, S., Javadian, F., Behrangi, A., 2019. On the use of satellite, gauge, and reanalysis precipitation products for drought studies. Environ. Res. Lett. 14, 075005.
Hao, Z., AghaKouchak, A., 2013. Multivariate standardized drought index: a multi-index parametric approach for drought analysis. Adv. Water Resour. 57, 12–18.
Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations-the CRUTS3.10 Dataset. Int. J. Climatol. 34, 623–642.
Hirpa, F.A., Gebremichael, M., Hopson, T., 2010. Evaluation of high resolution satellite precipitation products over very complex terrain in Ethiopia. J. Appl. Meteorol. Climatol. 49, 1044–1051.
Hong, Y., Hsu, K.-L., Sorooshian, S., Gao, X., 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. J. Appl. Meteorol. 43, 1834–1853.
Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38–55.
Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019. GPM IMERG Final Precipitation L3 1 Month 0.1 Degree X 0.1 Degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD.
Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5, 487–503.
Katiraie-Boroujerdy, P., Nasrollahi, N., Hsu, K., Sorooshian, S., 2016. Quantifying the reliability of four global datasets for drought monitoring over a semi-arid region. Theor. Appl. Climatol. 123, 387–398.
Le Coz, C., van de Giesen, N., 2020. Comparison of rainfall products over sub-Saharan Africa. J. Hydrometeorol. 21, 553–596.
Lemma, E., Upadhyaya, S., Ramsankarn, R.A.A.J., 2019. Investigating the performance of satellite and reanalysis rainfall products at monthly timescales across different rainfall regimes of Ethiopia. Int. J. Rem. Sens. 40, 4019–4042.
McKee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales'. In: Proceedings of the Eighth Conference on Applied Climatology. American Meteorological Society, Boston, pp. pp179–184.
McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. Sci. Data 4, 170012.
Nguyen, P., Shearer, E.J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., Huynh, P., Updegraff, G., Hsu, K., Kuligowski, B., Logan, W.S., Sorooshian, S., 2019. The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. Sci. Data 6, 180296.
Novella, N.S., Thiaw, W.M., 2013. African rainfall climatology version 2 for famine early warning systems. J. Appl. Meteorol. Climatol. 52, 588–606.
Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosogrove, B., Radakovich, J., Bosilovich, M., Entin, J., Walker, J., Lohmann, D., Toll, D., 2004. The Global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394.
Romilly, T.G., Gebremichael, M., 2011. Evaluation of satellite rainfall estimates over Ethiopian river basins. Hydrol. Earth Syst. Sci. 15, 1505–1514.
Rui, H., McNally, A., 2019. README: Document for Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) Products. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA Goddard Space Flight Center, USA.
Sahlu, D., Moges, S.A., Nikolopoulos, E.I., Anagnostou, E.N., Hailu, D., 2017. Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa, 2017 Adv. Meteorol., 4957960
Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2014. GPCC's new land surface precipitation climatology based on quality-controlled in-situ data and its role in quantifying the global water cycle. Theor. Appl. Climatol. 115, 15–40.
Shahid, S., 2008. Spatial and temporal characteristics of droughts in the western part of Bangladesh. Hydrol. Process. 22, 2235–2247.
Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. Bull. Am. Meteorol. Soc. 81, 2035–2046.
Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R.P., Stringer, M., Chadwick, R., Kayitakire, F., 2014. Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. J. Appl. Meteorol. Climatol. 53, 2805–2822.
Vicente-Serrano, S.M., Begueria, S., Gimeno, L., Eklundh, L., Giuliani, G., Weston, D., El Kenawy, A., Lopez-Moreno, J.I., Nieto, R., Ayenew, T., Konte, D., Ardö, J., Pegram, G.G.S., 2012. Challenges for drought mitigation in Africa: the potential use of geospatial data and drought information system. Appl. Geogr. 34, 471–486.
Vogel, C., Steynor, A., Manyuchi, A., 2019. Climate services in Africa: Re-imagining an inclusive, robust and sustainable service. Clim. Serv. 15, 100107.
WMO (World Meteorological Organization), 2012. Standardized Precipitation Index User Guide (WMO-No.1090), Geneva. http://www.droughtmanagement.info/literature/WMO_standardized_precipitation_index_user_guide_en_2012.pdf.

Worqlul, A.W., Maathuis, B., Adem, A.A., Demissie, S.S., Langan, S., Steenhuis, T.S., 2014. Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. Hydrol. Earth Syst. Sci. 18, 4871–4881.

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S., 2007. A gauge-based analysis of daily precipitation over East Asia. J. Hydrometeorol. 8, 607–626.

Young, M.P., Williams, C.J.R., Christine Chiu, J., Maidment, R.I., Chen, S., 2014. Investigation of discrepancies in satellite rainfall estimates over Ethiopia. J. Hydrometeorol. 15, 2347–2369.

Zhao, H., Ma, Y., 2019. Evaluating the drought-monitoring utility of four satellite-based quantitative precipitation estimation products at global scale. Rem. Sens. 11, 2010.