



Evaluation of Global Gridded Precipitation and Temperature Datasets against Gauged Observations over the Upper Tekeze River Basin, Ethiopia

Kidane Welde REDA, Xingcai LIU, Qiuhong TANG, Tesfay Gebretsadkan GEBREMICHAEL

Citation: , 2021: Evaluation of Global Gridded Precipitation and Temperature Datasets against Gauged Observations over the Upper Tekeze River Basin, Ethiopia. *J. Meteor. Res.*, 35(4), 673–689, doi: [10.1007/s13351-021-0199-7](https://doi.org/10.1007/s13351-021-0199-7)

View online: <http://jmr.cmsjournal.net/article/doi/10.1007/s13351-021-0199-7>

Related articles that may interest you

[Evaluation and Hydrological Application of CMADS Reanalysis Precipitation Data against Four Satellite Precipitation Products in the Upper Huaihe River Basin, China](#)

Journal of Meteorological Research. 2020, 34(5), 1096 <https://doi.org/10.1007/s13351-020-0026-6>

[Evaluation of the Biodiversity Conservation Function in Liaohe Delta Wetland, Northeastern China](#)

Journal of Meteorological Research. 2020, 34(4), 798 <https://doi.org/10.1007/s13351-020-9186-7>

[Evaluation of the CAM and PX Surface Layer Parameterization Schemes for Momentum and Sensible Heat Fluxes Using Observations](#)

Journal of Meteorological Research. 2018, 32(6), 1026 <https://doi.org/10.1007/s13351-018-8009-6>

[Evaluation of TIGGE Daily Accumulated Precipitation Forecasts over the Qu River Basin, China](#)

Journal of Meteorological Research. 2019, 33(4), 747 <https://doi.org/10.1007/s13351-019-8096-z>

[Arctic Climate Changes Based on Historical Simulations \(1900–2013\) with the CAMS-CSM](#)

Journal of Meteorological Research. 2018, 32(6), 881 <https://doi.org/10.1007/s13351-018-7188-5>

[Seasonal Prediction of Boreal Winter Rainfall over the Western Maritime Continent during ENSO](#)

Journal of Meteorological Research. 2020, 34(2), 294 <https://doi.org/10.1007/s13351-020-9181-z>

Evaluation of Global Gridded Precipitation and Temperature Datasets against Gauged Observations over the Upper Tekeze River Basin, Ethiopia

Kidane Welde REDA^{1,2,3}, Xingcai LIU^{1*}, QiuHong TANG^{1,2}, and Tesfay Gebretsadkan GEBREMICAEL³

¹ Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Tigray Agricultural Research Institute, Mekelle 492, Ethiopia

(Received December 3, 2020; in final form April 12, 2021)

ABSTRACT

The availability of satellite and reanalysis climate datasets and their applicability have been greatly promoted in hydro-climatic studies. However, such climatic products are still subject to considerable uncertainties and an evaluation of the products is necessary for applications in specific regions. This study aims to evaluate the reliability of nine gridded precipitation and temperature datasets against ground-based observations in the upper Tekeze River basin (UTB) of Ethiopia from 1982 to 2016. Precipitation, maximum temperature (T_{\max}), minimum temperature (T_{\min}), and mean temperature (T_{mean}) were evaluated at daily and monthly timescales. The results show that the best estimates of precipitation are from the Earth2Observe, WFDEI, and ERA-Interim reanalysis data Merged and Bias-corrected for the Inter-Sectoral Impact Model Intercomparison Project (EWEMBI), and the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) datasets. The percentage biases and correlation coefficients (CCs) are within $\pm 15\%$ and > 0.5 , respectively, for both EWEMBI and CHIRPS at the two timescales. All products underestimate the drought conditions indicated by the standardized precipitation index (SPI), while the EWEMBI and CHIRPS datasets show higher agreement with the observations than other datasets. The T_{mean} estimates produced by the ECMWF Re-Analysis version 5 (ERA5) and the Climate Hazards Group InfraRed Temperature with Station data (CHIRTS) are the closest to the observations, with CCs of 0.65 and 0.55, respectively, at the daily timescale. The CHIRTS and EWEMBI datasets show better representations of T_{\max} (T_{\min}), with CCs of 0.69 (0.72) and 0.62 (0.68), respectively, at the monthly timescale. The temperature extremes are better captured by the ERA5 (T_{mean}), CHIRTS (T_{\max}), and EWEMBI (T_{\min}) datasets. The findings of this study provide useful information to select the most appropriate dataset for hydrometeorological studies in the UTB and could help to improve the regional representation of global datasets.

Key words: climatic datasets, evaluation, rainfall, temperature, Tekeze River basin

Citation: Reda, K. W., X. C. Liu, Q. H. Tang, et al., 2021: Evaluation of global gridded precipitation and temperature datasets against gauged observations over the upper Tekeze River basin, Ethiopia. *J. Meteor. Res.*, **35**(4), 673–689, doi: 10.1007/s13351-021-0199-7.

1. Introduction

Climate variables play various essential roles in the components of the global water cycle. Specifically, rainfall has dominant influences in the hydrological cycle and thus affects the modeling of extreme hydrological events (floods and droughts), water management, and natural disasters (landslides and earthquakes) (Tan et al., 2018; Gyasi-Agyei, 2020; Satgé et al., 2020). Similarly,

temperature has significant influences on heat and cold waves (Costa et al., 2020). However, the variabilities of precipitation and temperature in space and time cause their measurements and estimations to be difficult.

The temporal and spatial coverages of gauge observations and their quality standards highly influence the accuracy of hydrometeorological studies (Li et al., 2020). The quality and stability of climatic data from ground meteorological stations may not be sufficient due to a limited number of stations, uneven spatial distribution,

Supported by the National Natural Science Foundation of China (41790424, 41730645, and 41877164), Strategic Priority Research Program of the Chinese Academy of Sciences (XDA20060402), and International Partnership Program of the Chinese Academy of Sciences (131A11KYSB20170113).

*Corresponding author: xingcailiu@igsnrr.ac.cn

© The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2021

and vulnerability to human and environmental factors (Wang et al., 2020). These issues are more pronounced in developing countries (Boke, 2017). Economic and political problems often result in sparse and nonuniformly-distributed gauge networks that are unable to capture temporal and spatial climatic variability (Le Coz and van de Giesen, 2020). For example, the recommended rain gauge density is one station per 600–900 km² for flatlands and one station per 100–250 km² for topographically rugged areas (Zeng et al., 2018). Africa has only 744 stations, of which only a quarter of them meet the international standards. However, Africa ideally requires 10,000 uniformly distributed standard stations to capture its spatiotemporal climate variabilities (Satgé et al., 2020). Moreover, the gauges that exist in Africa are concentrated in a few regions, such as in South Africa (Le Coz and van de Giesen, 2020). In Ethiopia, the gauge networks are very sparse and are of low quality. For example, in the upper Tekeze River basin (UTB), one rain gauge station covers an area of approximately 1400 km², which is far below the recommended density standard (Gebremicael et al., 2019). Thus, investigating relevant data sources that can capture the spatiotemporal climatic variability of the UTB is required for high-quality hydro-meteorological studies.

Recently, the availability of high-quality gridded climatic dataset products (satellite and reanalysis-based) has increased, and these products are increasingly used in hydrological and water management studies (Hu et al., 2017; Fang et al., 2019; Azimi et al., 2020; Gebremicael et al., 2020; Wang et al., 2020). In this context, gridded climatic datasets at both regional and global scales become essential alternative data sources (Li et al., 2020). This could be a result of the fast development of remote sensing and data assimilation technologies (Zhang and Ma, 2018). However, the performance of these datasets in representing the ground-based observed data and in hydrological modeling applications is inconsistent between different regions and basins.

Many studies have evaluated and validated different datasets against gauge measurements (Dembélé and Zwart, 2016; Sahlu et al., 2017; Aslami et al., 2019; Lockhoff et al., 2019; Ayoub et al., 2020; Islam and Cartwright, 2020). Moreover, satellite and reanalysis-based datasets have also been used to drive hydrological models in different basins around the world (Zhao et al., 2015; Tan et al., 2017; Li et al., 2018; Roy et al., 2018; Awange et al., 2019; Azimi et al., 2020). These studies suggested that the various datasets may have different applicability across regions. For instance, the climatic research unit (CRU) time series (TS) 2.1 shows good agreement over most parts of China as compared to other

reanalysis rainfall estimates (Zhao and Fu, 2006). A similar study by Dembélé and Zwart (2016) indicated that the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPSv8) precipitation dataset better captured the observed extremes in West Africa. Notable improvements of interpolated surface temperature have been achieved through topographic correction for the 40-yr ECMWF Re-Analysis (ERA-40) temperature dataset over China (Zhao et al., 2008). The different meteorological datasets are inevitably subject to uncertainties resulting from the factors associated with climate zones, seasonal changes, ground-surface conditions, geographical positions, and the embedded algorithms used to derive the datasets (Fang et al., 2019; Gebremicael et al., 2019; Wang et al., 2020). Assessment of the accuracy and uncertainties in historical meteorological dataset products based on ground observations is helpful for water resource and hydrological studies.

Some studies have already evaluated satellite- and reanalysis-based global datasets in Ethiopia, e.g., in the Lake Tana basin (Worqlul et al., 2014), the Gilgel Abay watershed (Lakew et al., 2017), and the upper Blue Nile (Bayissa et al., 2017; Sahlu et al., 2017). These studies focused on rainfall datasets only in the upper Blue Nile basin. Gebremicael et al. (2019) evaluated eight precipitation products in the UTB. However, their study focused only on satellite rainfall products despite the fact that temperature is also a key parameter for hydro-climatic studies in the region. Reanalysis-based climatic datasets often provide better spatial coverage than non-reanalysis-based datasets. Based on the above, our study aims to evaluate nine satellite and reanalysis rainfall and near surface air temperature (SAT) datasets against gauge-observed data in the UTB. The UTB is representative of the mesoscale basins in northern Ethiopia, and the Tekeze River is the most important water source for irrigation and hydropower in both Ethiopia and the downstream countries such as Sudan and Egypt (Fentaw et al., 2018; Gebremicael et al., 2019). With a comprehensive evaluation of the temperature and precipitation products, this study will contribute to UTB water resource management for the sustainable development of the basin region and the downstream countries.

2. Study area and data

2.1 Study area

The Tekeze River basin (Fig. 1), situated in the northwestern part of Ethiopia, is one of the major tributaries of the Nile River (Abraha, 2009). The Tekeze headwaters originate in the Meket Mountains near Lalibela and flow northwards until it turns westward along the Ethiopia-Eritrea

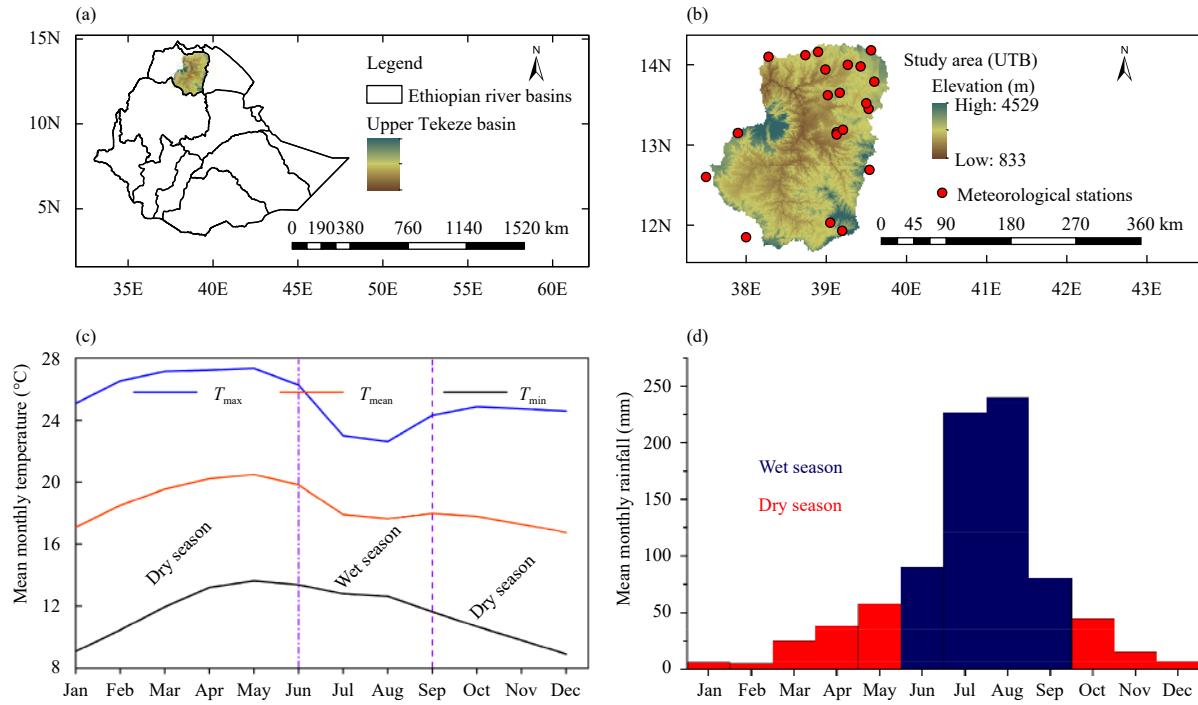


Fig. 1. Map of the study area: (a) Ethiopia, (b) geographical locations and distributions of meteorological stations in the UTB, (c) average monthly maximum, mean, and minimum temperatures (T_{max} , T_{mean} , and T_{min}), and (d) mean monthly rainfall of the UTB for dry and wet seasons based on gauge records from 1980 to 2019.

itrean border, with a distance of 600 km until it crosses the Ethiopia–Sudan border near Humera (MoWR, 1998). The basin is characterized by complex topography consisting of mountains, highlands, and lowlands with gently sloping terrain. The complex topography endows the basin a high potential for hydropower production in the mountainous areas and for irrigation in the lowland areas.

The northern and eastern parts of the UTB are categorized as semi-arid, and the southern part is categorized as semi-humid (Fentaw et al., 2018). The annual rainfall ranges from 400 mm yr⁻¹ in the eastern parts to 1200 mm yr⁻¹ in the southwestern parts of the basin (Gebremicael et al., 2019). The rainy season is from June to September, during which more than 70% of the total annual rainfall occurs in the UTB. The variables T_{max} and T_{min} vary from 3 to 21°C in the high-elevation areas (around the Semien Mountains) and from 19 to 43°C in the flat and low-elevation areas (around Humera). At the basin level, the average minimum and maximum temperatures (T_{min} and T_{max}) of the UTB are 11 and 32°C, respectively (Fentaw et al., 2018).

2.2 Climate data

2.2.1 Gauge observations

Gauge observations of daily maximum and minimum SAT and precipitation were collected from the National

Meteorological Agency of Ethiopia (NMAE) from 21 stations located inside and nearby the UTB (Fig. 1). These daily data span from 1980 to 2019, with the lengths of records varying from station to station. In this study, the stations having at least 35 years of continuous records with less than 2.5% missing values are considered. Most of the meteorological stations are in the northeastern part of the UTB (mostly highland area), and only a few are in the western and southwestern parts (dominantly lowland areas) of the basin (Fig. 1). This may introduce uncertainties into the spatially aggregated values of the climate variables. Thus, the point data are not interpolated into gridded time series. However, for the purpose of evaluation/validation of satellite and reanalysis climatic datasets, the distribution of stations can be considered sufficient for a point-to-pixel comparison.

Quality control such as excluding outliers and homogenization was applied to gauge observation TS data. Monthly precipitation and SAT series were calculated for each gauge station and then tested for homogeneity using the Standard Normal Homogeneity Test. The homogenization procedure is based on the application of the Standard Normal Homogeneity Test. This test assumes that precipitation amounts at the station being tested (test station) and some regional average values are proportional

to each other. In our study, the test was applied to a series of ratios comparing the observations at each gauge station (test stations) with the averaged observations at the four nearest stations to the station to be tested. This relationship is expressed in terms of the ratio between the test station normalized precipitation values and those of a regional TS defined as a weighted average of the four neighboring reference stations. See Section 1 in supplementary information for more details. The test indicated no homogenization problems in the rainfall and temperature gauge observations, which are thus reliable for statistical analyses following screening criteria.

2.2.2 Satellite-based and reanalysis datasets

We selected nine climatic datasets that include satellite, reanalysis, gauge-based temperature and rainfall observations, and the merged estimates based on the three types of datasets (Table 1). The choice of these products was mainly based on their temporal coverage, data integrity, availability of near-real-time data, accessibility, and popularity in the literature.

The CRU gridded TS (CRU TS) dataset is produced by the CRU at the University of East Anglia based on the World Meteorological Organization (WMO) global stations (Harris et al., 2014). Many studies have used this dataset in diverse research areas in Africa since it was first released in 2000 (e.g., Haile et al., 2020; Mahmood et al., 2020; Peng et al., 2020).

ERA5-land is a global reanalysis-based product produced by the Copernicus Climate Change Service (C3S; Maidment et al., 2013). The variable T_{mean} at 2 m above the ground and the total precipitation were obtained from this dataset. These two variables are based on inputs from satellite radiances and in-situ data provided by the WMO information system (WIS). This product has been applied in Africa, e.g., in East Africa (Agutu et al., 2017) and Uganda (Maidment et al., 2013).

The CHIRPS data are a quasi-global rainfall dataset developed by the Climate Hazards Center (CHC) at the

University of California, Santa Barbara and the US Geological Survey (USGS; Funk et al., 2015). This dataset is based on the CHC's precipitation climatology dataset and quasi-global geostationary thermal infrared satellite observations. Rainfall data from the NOAA Climate Forecast System and the TRMM 3B42 (Tropical Rain Measurement Mission level 3 output) product are also the inputs for CHIRPS (Funk et al., 2015). This product has been widely used in East Africa, as reported in many studies (e.g., Bayissa et al., 2017; Lemma et al., 2019).

Climate Hazards Group InfraRed Temperature with Station (CHIRTS) daily SAT data was developed by the CHC at the global scale (60°S – 70°N) with a spatial resolution of $0.05^{\circ} \times 0.05^{\circ}$ (approximately 5 km). This dataset contains daily maximum and minimum SAT and was produced through the disaggregation from monthly fields of the CHIRTS T_{max} (CHIRTSmax1) TS and by synthesizing daily temperatures from the ERA5 (Funk et al., 2019). The input data for CHIRTSmax1 are Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2), CRU TS, and ground station data (of approximately 15,000 stations) from the Berkeley Earth database.

The Global Precipitation Climatology Centre (GPCC) product v2018 of global daily precipitation was produced based on rainfall data provided by national meteorological and hydrological station services, regional and global data collections and WMO GTS (Global Telecommunication System) data (Schamm et al., 2014). More than 85,000 rain gauge stations across the globe were used in the GPCC dataset. This dataset has been widely used in Africa, e.g., in the Horn of Africa (Dinku et al., 2007; Funk et al., 2015) and sub-Saharan Africa (Harrison et al., 2019).

The Climate Prediction Center (CPC) dataset is part of a product suite from the CPC Unified Precipitation Project that is underway at the NOAA (Chen et al., 2008). The precipitation and temperature variables were

Table 1. Summary of dataset characteristics considered in this study

Dataset	Resolution	Frequency	Coverage	Period	Variable	Reference
EWEMBI v1.1	0.5°	Daily	Global	1979–2016	Pre, T_{max} , T_{min} , T_{mean}	Lange (2019)
CRU TS v4.03	0.5°	Monthly	Global	1901–2018	Pre, T_{max} , T_{min} , T_{mean}	Harris et al. (2020)
ERA5-land	0.1°	Hourly	Global	1982–2020	Pre, T_{mean}	C3S (2019)
GPCC v2018	1°	Daily	Global	1982–2016	Pre	Ziese et al. (2018)
CPC	0.5°	Daily	Global	1979–2019	Pre, T_{max} , T_{min} , T_{mean}	Chen et al. (2008)
NCEP Reanalysis 2	2°	6 hourly	Global	1979–2019	Pre, T_{max} , T_{min} , T_{mean}	Kanamitsu et al. (2002)
WFDEI	0.5°	Daily	Global	1979–2016	Pre	Weedon et al. (2018)
CHIRPS v8	0.05°	Daily	50°S – 50°N	1981–2019	Pre	Funk et al. (2015)
CHIRTS	0.05°	Daily	Global	1983–2016	T_{max} , T_{min} , T_{mean}	Funk et al. (2019)

Note: EWEMBI: the Earth2Observe, WFDEI, and ERA-Interim reanalysis data Merged and Bias-corrected for the Inter-Sectoral Impact Model Intercomparison Project; CRU: Climatic Research Unit; ERA5: ECMWF Re-Analysis version 5; GPCC: Global Precipitation Climatology Centre; CPC: Climate Prediction Center; NCEP: National Centers for Environmental Prediction; WFDEI: WATCH Forcing Data methodology applied to ERA-Interim data; CHIRPS: Climate Hazards Group InfraRed Precipitation with Station data; CHIRTS: Climate Hazards Group InfraRed Temperature with Station data. Pre denotes precipitation.

produced by merging observations from gauge stations with precipitation estimates from several satellite-based (infrared and microwave) datasets.

NCEP Reanalysis-2, produced by the NOAA-CIRES Climate Diagnostics Center, is a new version of the NCEP Reanalysis-1 model that has been modified through the update of the physical process parameterization and the fixing of errors (Marques et al., 2009). NCEP Reanalysis-2 performed data assimilation using the Rapid Radiative Transfer Model (RRTM) developed by the Atmospheric and Environmental Research (AER) group using different data sources.

WFDEI was generated by using the same methodology as the WATCH (WATER and global CHange) Forcing Data by making use of the ERA-Interim reanalysis data (Weedon et al., 2014). The ERA-Interim data improved the temperature and precipitation data compared to WATCH. WFDEI has been widely used for studies on climate change and hydrological modeling (Liu et al., 2017; Lange, 2019).

EWEMBI was a newly compiled reference dataset named for EartH2Observe, WFDEI, and ERA-Interim reanalysis data Merged and Bias-corrected for the Inter-Sectoral Impact Model Intercomparison Project (EWEMBI) phase 2b (Lange, 2018). The sources of the precipitation and temperature data of the EWEMBI dataset are the WFDEI and GPCC v5 data over the land surface, and EartH2Observe forcing data (E2OBS) over the ocean (Lange, 2016).

3. Methodology

3.1 Evaluation indices

The climatic variables of precipitation and mean, maximum, and minimum SAT were evaluated at daily and monthly timescales. The temperature and precipitation products with hourly temporal frequencies were aggregated to daily and monthly TS. Evaluation of the selected temperature and precipitation products was performed for 35 yr covered by all the products. A point-to-pixel comparison approach was used to evaluate the products against their corresponding observations at each gauge station. In this point-to-pixel approach, precipitation and temperature products from a grid cell were compared to the observed gauge data located within this pixel. This approach is more applicable in rugged topography and complex terrain such as the UTB than the areas with homogenous topography for the evaluation of such products at small spatial scales with varying climatic factors (Gebremicael et al., 2019; Lemma et al., 2019; Li et al., 2020).

The precipitation products were further evaluated for

monthly cumulative precipitation and different rainfall intensity groups. The monthly standardized precipitation index (SPI) was used to compare the performances of the rainfall datasets in capturing historical drought events. SPI is an index for characterizing and monitoring drought conditions on a range of timescales based on the probability distribution of a long-term precipitation TS (McKee et al., 1995; Liu et al., 2012). The calculation of SPI includes a transformation of one frequency distribution (e.g., gamma) to another frequency distribution (normal or Gaussian). Specifically, a time serial of precipitation is firstly fitted with a gamma (or an incomplete beta) probability distribution, and then transformed to a normal (or Gaussian) distribution. The gamma distribution has been widely used in previous studies as it has been understood as the reliable fit to the rainfall data. The normal distribution has a mean of zero and standard deviation of one and then can be used to indicate dry or wet condition. The detailed SPI calculation procedures can be found in the supplementary information. The SPI was calculated for 3-month (SPI3), 6-month (SPI6), 9-month (SPI9), and 12-month (SPI12) timescales for each precipitation product. A threshold value of -1 is usually used as an indicator of drought conditions for the SPI (Tefera et al., 2019). Hence, $-1.0 \geq SPI > -1.5$ indicates moderate drought, $-1.5 \geq SPI > -2.0$ indicates severe drought, and $SPI < -2.0$ indicates extreme drought (Liu et al., 2012). The trend of extreme heat and the number of hot days per year (number of days scored above the extreme value) were also evaluated for the temperature datasets against the gauge-observed data. In this study, extreme temperature was defined as $SAT >$ the 95th percentile for daily maximum and mean SAT and $SAT <$ the 5th percentile for minimum SAT.

3.2 Evaluation metrics

Four statistical metrics (Table 2), namely, the Pearson correlation coefficient (CC), percentage of bias (Pbias), mean absolute error (MAE), and root mean square error (RMSE), were used for the evaluation of temperature and precipitation products. These statistical measures are commonly used metrics for hydro-climatic dataset evaluation (Gebremicael et al., 2019; Costa et al., 2020; Satgé et al., 2020; Wang et al., 2020). Detailed descriptions of such statistical metrics can be found in Chai and Draxler (2014), Toté et al. (2015), and Saccenti et al. (2020). Negative (positive) values of Pbias indicate underestimation (overestimation) of the observations, respectively. High CC and low RMSE and MAE values indicate that the estimates in a climatic dataset are close to those of gauge-based observations.

Table 2. Statistical indices used for evaluating multiple meteorological products

Metric name	Function	Optimum value
Pearson correlation coefficient (CC)	$CC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2} \sqrt{(y_i - \bar{y})^2}}$	1
Percentage of bias (Pbias)	$Pbias = \left(\frac{\sum y_i - \sum x_i}{\sum x_i} \right) \times 100$	0
Mean absolute error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - x_i $	0
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$	0

Note: x_i is the temperature or precipitation from gauge observations, y_i is the corresponding value from the precipitation or temperature product, n is the number of observations, and \bar{x} and \bar{y} are the averages of the gauge-observed and meteorological product values, respectively. The evaluation was performed at daily and monthly timescales for all the selected gauge stations.

4. Results and discussion

4.1 Comparison of precipitation products with gauge observations

4.1.1 Daily timescale

The statistical metrics from the comparison of precipitation products against the observed data across all gauging stations of the UTB are shown in Fig. 2. The

EWEMBI dataset shows a reasonably good correspondence with the rain gauge observations at the daily timescale, with average values of 0.58, 2.3%, 6.4 mm, and 2.7 mm for CC, Pbias, RMSE, and MAE, respectively. This may be due to the climatological bias-adjustments done for EWEMBI. EWEMBI was produced by using data sources E2OBS, WFDEI, and ERAI, which are bias-corrected by using GPCPv2.1 monthly precipitation total

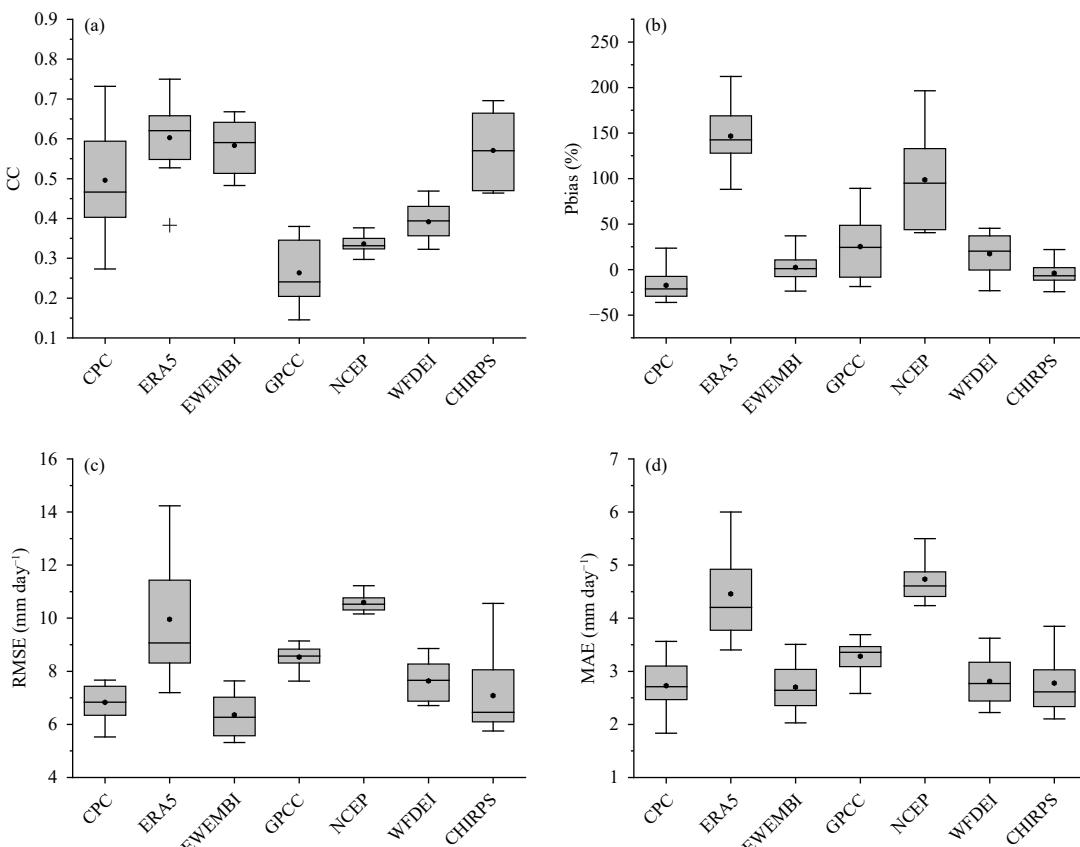


Fig. 2. Boxplots of evaluation metrics for the seven precipitation products at daily timescales during 1982–2016. (a) Pearson correlation coefficient (CC), (b) percentage of bias (Pbias), (c) root mean square error (RMSE), and (d) mean absolute error (MAE). Each boxplot also indicates the range of variation from minimum to maximum (vertical line), the quartiles and median (horizontal lines), mean (points), and outliers (+) of the precipitation estimates at the 21 stations.

over the ocean (Balsamo et al., 2015) and with GPCCv5/v6 monthly precipitation totals over the land (Weedon et al., 2014). The CC, Pbias, RMSE, and MAE values of EWEMBI for the different gauge stations range from 0.48 to 0.67, -23.6% to 37%, 5.3 to 7.6 mm, and 2 to 3.5 mm, respectively. Compared with EWEMBI, the CHIRPS and CPC products perform well, with CC values greater than 0.5. The likely reason for the better performance of CHIRPSv8 could be its high spatial resolution (0.05°) as compared to the other products and the CHIRPS algorithm that combines the bias-corrected CHIRP with station observations. A similar study in the upper Blue Nile basin, Ethiopia by Sahlu et al. (2017), Gebremicael et al. (2019) and Lemma et al. (2019) also showed that the CHIRPSv8 product better captures the ground observations compared to other satellite products. The averages and range values (in brackets) of CC, Pbias, RMSE, and MAE of CHIRPS for all stations are 0.57 (0.45 to 0.7), -4% (-24.4 to 22%), 7.1 mm (5.8 to 10.6 mm), and 2.8 mm (2.1 to 3.8 mm), respectively. Similarly, average values and ranges of 0.5 (0.3 to 0.7), -16.8% (-31.1 to 23.5%), 6.8 mm (5.5 to 7.7 mm), and 2.7 mm (1.8 to 3.6 mm) for CC, Pbias, RMSE, and MAE were obtained for the CPC dataset. The CC (Fig. 2a), Pbias (Fig. 2b), RMSE (Fig. 2c), and MAE (Fig. 2d) values for each precipitation product except for the EWEMBI, CHIRPS, and CPC exhibit poor agreement with the majority of the rain gauge networks at the daily timescale. The larger gap between some of the datasets and the observed data at the daily timescale could be due to the complex and rugged topography in the UTB. This, in turn, could contribute to failures in detecting more localized convective rainfall events.

Among the precipitation datasets, the ERA5 and NCEP products overestimate the rainfall at all stations. The ERA5 products resulted in a high CC (0.6) but largely overestimated the rainfall at all stations, with an average (range) value of Pbias reaching 150% (88% to 242%). This indicates the poor performance of the ERA5 product in capturing the ground rainfall over the basin, which is consistent with the previous study by Lemma et al. (2019). Likewise, NCEP extremely overestimates (Pbias = 98.6%) the daily observed rainfall. NCEP also shows a relatively poor skill throughout the basin, with an average CC value of 0.33. The RMSE and MAE values for both the ERA5 and NCEP products are also very large compared to those of the other datasets (Fig. 2). The performances of the remaining precipitation products are characterized by results in between those of the rainfall products discussed above.

Figure 3 presents a comparison between the precipita-

tion estimates and the observed rainfall using the spatial distribution of CC. The spatial distributions of CC, Pbias, RMSE, and MAE (Table S1a–c) indicated by the range values show that EWEMBI and CHIRPSv8 products have better agreement with ground rainfall, especially at the stations in the northeastern part of the basin. The NCEP, GPCC, and WFDEI show poor agreement with CC values less than 0.5 at almost all the stations.

4.1.2 Monthly timescale

The daily precipitation products were then evaluated at the monthly timescale in order to determine how much the performance of the precipitation products would be improved in the magnitude of the indicators from daily to monthly scales. Meanwhile, they could also be compared with the well-known monthly precipitation product, the CRU TS v4.03. Table 3 summarizes the performances of the different rainfall estimates using the various statistical indices at the monthly timescale. The statistical indices of all products in estimating the monthly rainfall are significantly improved compared to those at the daily timescale. As shown in Table 3, all products have average CC values of greater than 0.7. This result agrees with previous studies showing improvements in detecting rainfall events using satellite and reanalysis-based datasets when the temporal scale was changed from daily to monthly (Roy et al., 2018; Ayoub et al., 2020; Islam and Cartwright, 2020). This could be due to the variabilities being counterbalanced and the errors offset each other when the data are aggregated from shorter to longer timescales. Even though the estimation accuracies are improved for most of the precipitation datasets, the EWEMBI, CHIRPS, CRU TS v4.03, and CPC datasets relatively outperform the others, with higher CC values and lower Pbias, RMSE, and MAE values. In addition, the performance rank of these products at the monthly timescale is in line with the results obtained at the daily timescale.

As the observations have the highest accuracy at the location of stations, a case was added by interpolating grid data to the stations and compare these interpolated-gridded data with station observation directly. A little performance improvement was found for most of the precipitation products (Table 4). CHIRPS (CC = 0.9) and EWEMBI (CC = 0.89) perform better as compared to the others, while the NCEP, ERA, and WFDEI show relatively poor agreement with the ground observations.

4.1.3 Precipitation intensity

To further understand whether the different precipitation datasets could capture rainfall events within various intensity groups, we divided the daily precipitation intensity (PI) of the gauge data (OBS) into 6 groups ($0 \leq$

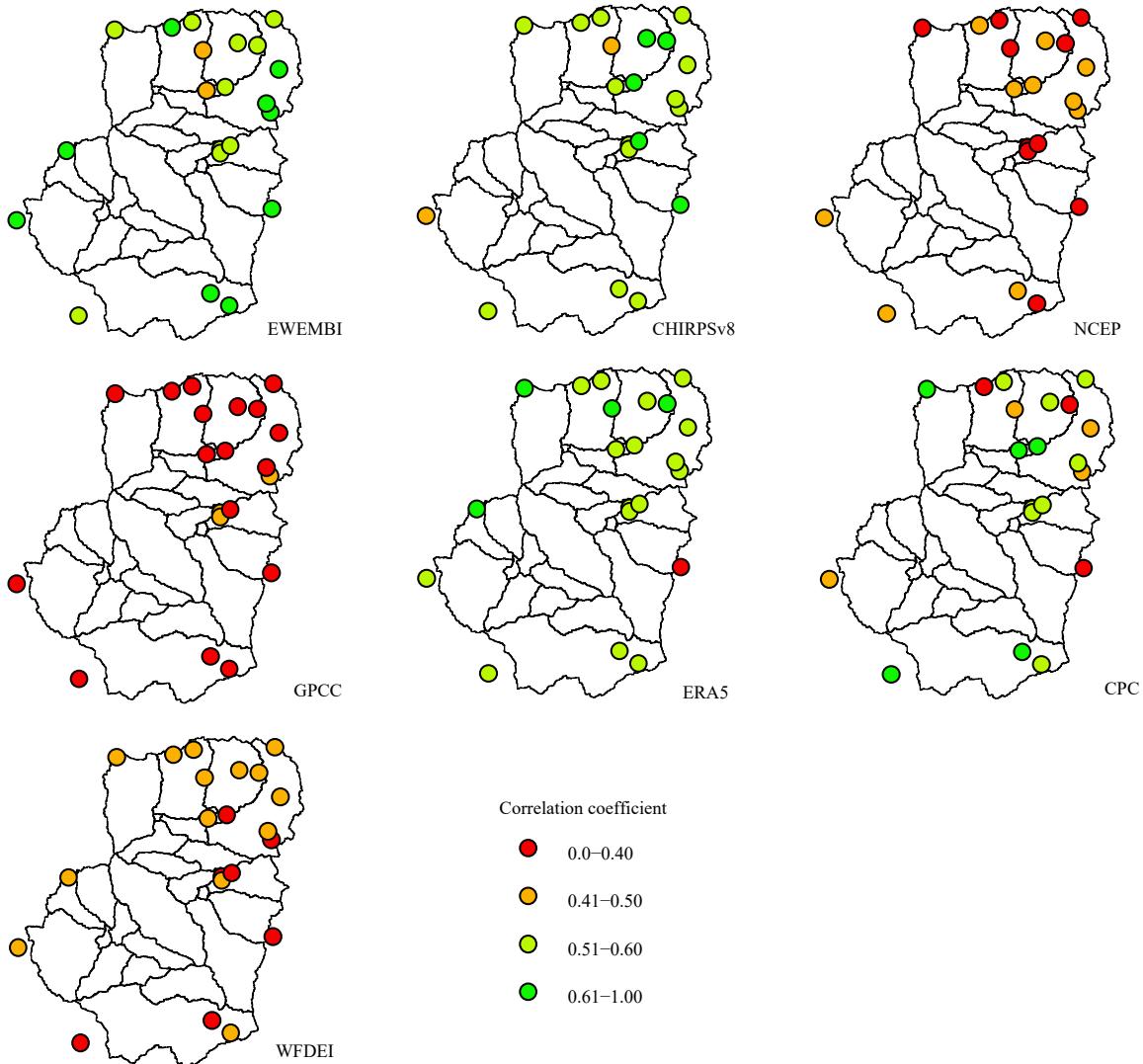


Fig. 3. Spatial distributions of CCs for the seven daily precipitation products.

Table 3. Ranges and means of the statistical metrics derived from the comparison between the different precipitation datasets and the observed data at the monthly timescale. The range values are shown in parentheses

Meteorological forcing	CC	Pbias (%)	RMSE (mm month^{-1})	MAE (mm month^{-1})
CPC	0.78 (0.65, 0.85)	-12 (-36, 75)	62 (40, 80)	37 (22, 48)
ERA5	0.84 (0.7, 0.9)	146 (88, 242)	189 (102, 339)	103 (65, 163)
EWEMBI	0.86 (0.77, 0.93)	5 (-24, 25)	51 (34, 82)	31 (18, 53)
GPCC	0.80 (0.65, 0.87)	25 (-19, 89)	67 (55, 83)	42 (34, 55)
NCEP	0.67 (0.58, 0.75)	99 (41, 196)	122 (101, 139)	80 (65, 91)
WFDEI	0.78 (0.69, 0.85)	13 (-23, 64)	69 (53, 103)	40 (26, 63)
CHIRPS	0.85 (0.78, 0.91)	-7 (-35, 22)	51 (34, 90)	30 (20, 57)
CRU TS	0.81 (0.69, 0.90)	-9 (-35, 54)	57 (38, 76)	34 (21, 43)

$\text{PI} < 1$, $1 \leq \text{PI} < 5$, $5 \leq \text{PI} < 10$, $10 \leq \text{PI} < 20$, $20 \leq \text{PI} < 30$, and $\text{PI} \geq 30 \text{ mm day}^{-1}$ following Wang et al. (2020). Figure 4 shows the total cumulative rainfall values of the eight precipitation products under different PI groups. The PI groups ranging from 0–1 to 1–5 mm day^{-1} are overestimated by all precipitation products. However, it is clearly shown in Fig. 4 that ERA5 and NCEP highly overestimate all rainfall magnitudes (PI groups). In par-

ticular, ERA and NCEP fail to capture the categorized PI groups with values below 20 mm day^{-1} , but reasonably well capture for rain gauge observations of above 20 mm day^{-1} . This result is contrary to the findings by Islam and Cartwright (2020), in which they reported ERA5 outperforming other datasets in predicting rainfall accumulation below 20 mm day^{-1} but seriously underestimating higher rainfall values. This could be an indicator that the

Table 4. Mean of the statistical metrics resulting from the comparison between the different interpolated gridded rainfall datasets and the observed data at the monthly timescale

Meteorologi-cal forcing	CC	Pbias (%)	RMSE (mm month ⁻¹)	MAE (mm month ⁻¹)
CPC	0.76	-15	71	39
ERA5	0.81	127	127	96
EWEMBI	0.89	5.2	47	26
GPCC	0.86	20.5	67	40
NCEP	0.73	102	128	83
WFDEI	0.79	16	60	33
CHIRPS	0.9	-6.6	48	26
CRU TS	0.85	-11	52	28

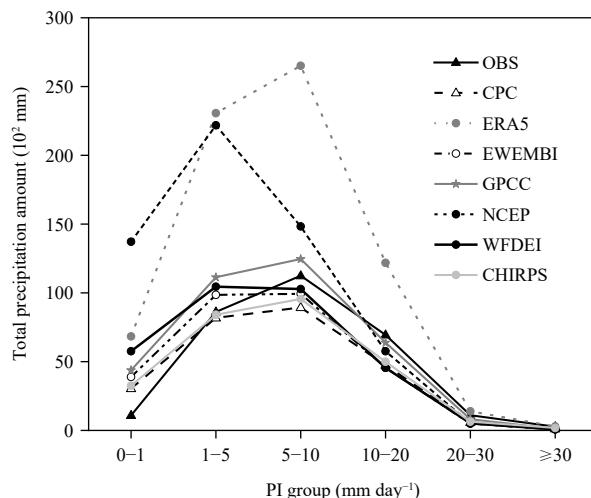


Fig. 4. Cumulative rainfall amounts under different PI groups during 1982–2016.

different datasets could produce different results in different regional studies. Most of the remaining precipitation products, especially EWEMBI, CPC, and CHIRPS, captured the PI groups greater than 5 mm day^{-1} better than ERA5 and NCEP.

The precipitation products were further evaluated by using statistical metrics to examine whether the datasets can capture the rainfall of the different PI groups. The CC and Pbias indices indicate that the different precipitation datasets poorly captured almost all PI groups. However, CHIRPS (for $0 \leq \text{PI} < 1$ and $1 \leq \text{PI} < 5$), EWEMBI (for $5 \leq \text{PI} < 10$ and $10 \leq \text{PI} < 20$), and CPC (for $20 \leq \text{PI} < 30$ and $\text{PI} \geq 30 \text{ mm day}^{-1}$) showed relatively better performance with higher CCs and lower Pbias values (Table 5) compared to the other datasets. This result is in agreement with a previous study by Wang et al. (2020), which showed lower capturing capacities of reanalysis and satellite rainfall products when rainfall events were grouped into different intensity ranges. The results of the statistical metrics (CC, Pbias, RMSE, and MAE) of all the precipitation datasets for the considered PI groups are summarized in the supplementary files (Ta-

Table 5. Statistical summary of EWEMBI, CPC, and CHIRPS precipitation products in capturing different PI groups

PI (mm day^{-1})	EWEMBI		CPC		CHIRPS	
	CC	Pbias	CC	Pbias	CC	Pbias
0–1	0.20	46.5	0.42	32.5	0.52	28
1–5	0.35	14.4	0.15	29.3	0.47	-2.3
5–10	0.57	-11.7	0.08	10.8	0.39	-15
10–20	0.51	-34.3	0.11	-8	0.36	-28.1
20–30	-0.39	-55.3	0.47	-26.1	-0.44	-42.3
≥ 30	-0.46	-72.5	0.48	-38.2	-0.46	-33.6

ble S2a–d).

4.1.4 Historical drought events

The ability of the different rainfall datasets to detect temporal drought was analyzed through comparisons to the drought indicated by the gauge-observed rainfall records (OBS). Figure 5 shows the monthly average SPI values over the UTB during 1982–2016 for each of the precipitation products using the 12-month SPI TS (SPI12). The SPI3, SPI6, and SPI9 TS are also given in the supplementary materials (Fig. S1). The discrepancies in the SPI results among the datasets were further supported by the time-specific historical exceptional drought records during the study period. For example, there was a severe drought in Ethiopia and Sudan during 1984–1985, induced by persistent rainfall shortages, which caused many deaths and migrations (Degefou and Bewket, 2015; Haile et al., 2020). The observed data detected this case as an extreme drought ($\text{SPI} < -2$; Fig. 5). EWEMBI, CRU TS v4.03, and CHIRPS detected a severe drought, while NCEP, ERA5, WFDEI, and GPCC identified a moderate drought for that specific period. Furthermore, CPC did not detect a drought in this particular period. Similarly, during 2002–2003, there was a countrywide drought in Ethiopia that affected more than 14 million people and resulted in severe damages (Muller, 2014; Nicholson, 2017). Almost all products detected this drought event better than the drought event in 1984; the datasets detected drought events for this case from moderate to extreme, except GPCC and WFDEI, which detected a very light drought ($0 > \text{SPI} > -1$).

Figure 6 indicates the statistical metrics (Pbias, CC, RMSE, and MAE) for the drought events (for which the observed SPI is ≤ -1) at the SPI3, SPI6, SPI9, and SPI12 TS for each product. Generally, all precipitation products underestimated the drought pattern compared to that detected in the observed data. This is revealed by the negative values of Pbias for all SPI TS (Fig. 6b). Drought estimates from EWEMBI, CRU TS v4.03, and CHIRPS are closer to the observed drought ($\text{SPI} \leq -1$) for SPI3, SPI6, SPI9, and SPI12 than the other datasets. The Pbias and CC for EWEMBI, CRU TS v4.03, and CHIRPS are estimated as -28%, -33%, and -44%, and 0.71, 0.67, and

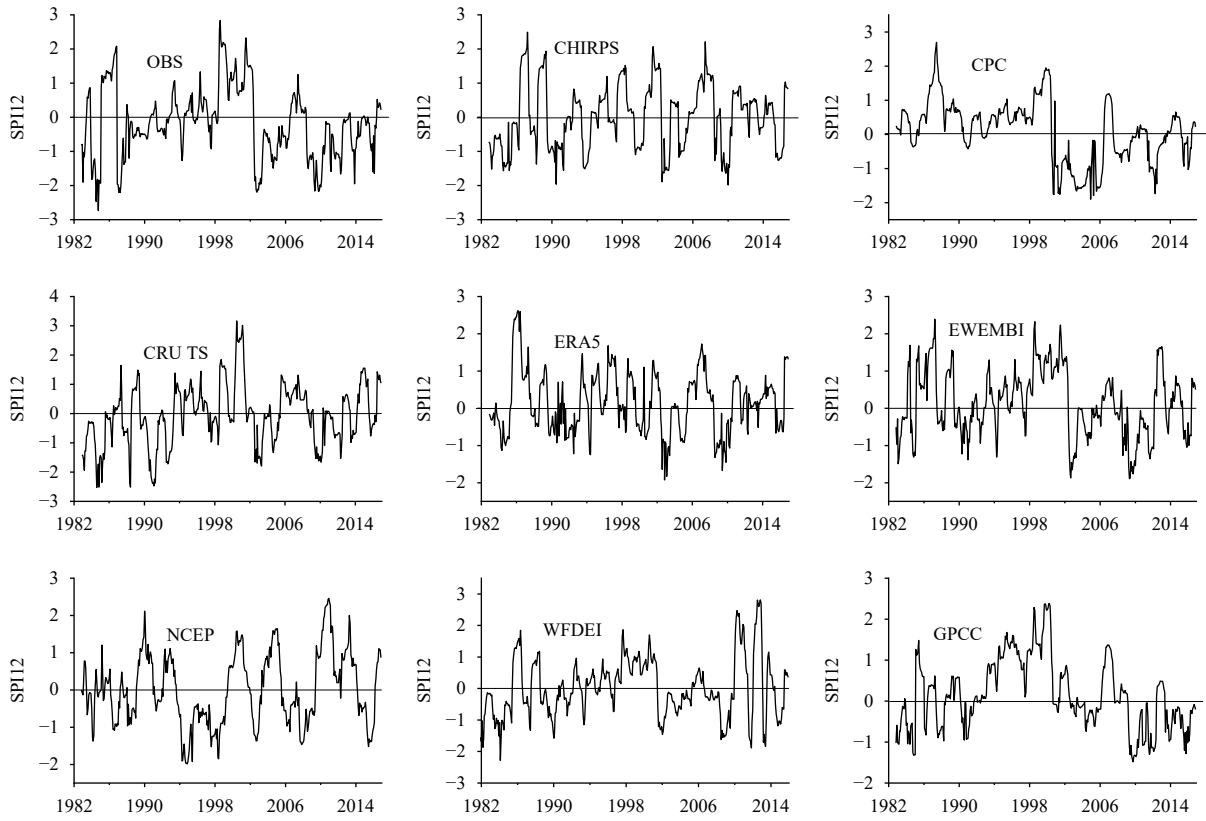


Fig. 5. Temporal variations in the monthly standardized precipitation index (SPI) over the UTB for a 12-month SPI timescale (SPI12) for all precipitation products.

0.73, respectively, for SPI12. The drought metrics at the SPI12 TS driven by NCEP, GPCC, ERA5, and WFDEI are significantly underestimated compared to the drought observations. The underestimates are within the range from -50% (ERA5) to more than -112% (NCEP). This indicates that most precipitation products can reasonably detect the meteorological, agricultural, and hydrological droughts in the UTB.

The Pbias (CC) values of EWEMBI are -40% (0.52), -37% (0.57), -32% (0.61), and -28% (0.71) for drought metrics of SPI3, SPI6, SPI9, and SPI12 TS, respectively. These values clearly show a consistently improving pattern with the increasing months of the SPI TS. This may imply the improvement of the capability in rainfall events detection of the precipitation products from daily to monthly timescales. However, there are some exceptions for WFDEI and NCEP; even though their correlations with the observed data increase with the SPI TS, their underestimations of drought are further accelerated as the SPI increases in the number of months from SPI3 to SPI6, SPI9, and SPI12 (Fig. 6). This may be attributed to the different retrieval methodologies and assumptions of the products. The results suggest that choice of precipitation product, spatial resolution, and record lengths can

vary significantly in precipitation-based drought metrics. These relationships also vary with the severity of drought events.

In summary, the EWEMBI, CRU TS v4.03, and CHIRPS datasets have shown better agreements with the observed drought at each SPI timescale (SPI3, SPI6, SPI9, and SPI 12) across the basin than the other products. The relatively better performance of these products is also confirmed by the minimum Pbias, RMSE, MAE, and maximum CC at the monthly timescale.

4.2 Evaluation of near-surface temperature products

4.2.1 Temperatures at daily and monthly timescales

Comparisons based on the exact daily T_{mean} at each gauge station indicate that the estimates from the ERA5, CHIRPS, and EWEMBI products are in agreement with the corresponding observations. As presented in Fig. 7, the average CC values for the ERA5, CHIRPS, and EWEMBI products are 0.65, 0.55, and 0.55, respectively. The daily T_{mean} patterns gathered from these products have consistent and better agreement with the ground T_{mean} than those of the NCEP and CPC products, which have large RMSE and MAE values. The discrepancies in

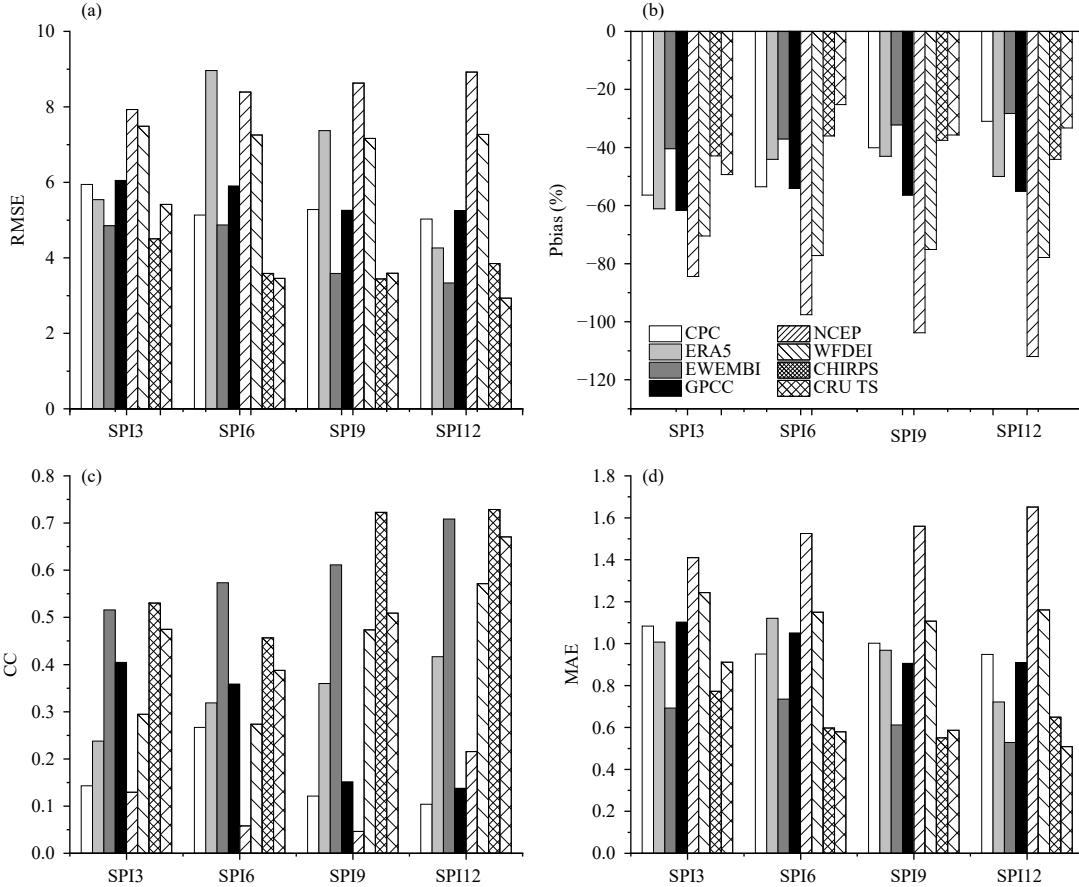


Fig. 6. Evaluation metrics for drought estimated from precipitation products compared with observed drought data ($SPI < -1$). (a) RMSE, (b) Pbias, (c) CC, and (d) MAE.

the metrics are large across stations for NCEP and CHIRTS. In particular, the ranges of the metrics are large for NCEP, e.g., 1.62–6.38°C (MAE) and 2.04–6.82°C (RMSE). The CC metric was also calculated based on the anomaly values of T_{mean} . Similar to the exact values, The ERA5, CHIRTS, and EWEMBI products are in agreement with the corresponding observation of T_{mean} anomalies, with CC value of 0.62, 0.53, and 0.52, respectively (Fig. S2).

Daily T_{max} and T_{min} were evaluated for CHIRPS, CPC, and EWEMBI, and monthly T_{max} and T_{min} were evaluated including CRU TS v4.03. Table 6 indicates the performances of each of these products in capturing their corresponding gauge values; the values shown are the average results of the metrics for all the stations within the basin. The T_{max} and T_{min} estimates of CHIRTS and EWEMBI show relatively more accuracy at the daily scale than the other datasets. Likewise, the T_{max} and T_{min} estimates of CHIRTS, CRU TS v4.03, and EWEMBI indicate better agreement with the observed values at the monthly timescale (Table 6) than the other datasets. However, analogous to the precipitation products, all the

temperature products show poor performances at the daily timescale than at the monthly timescale. The better performance at the monthly timescale could be due to that the errors in daily data offset each other when aggregated to the monthly data.

The CHIRTS, CPC, and EWEMBI products overestimate T_{max} and T_{min} at both daily and monthly timescales. The CRU TS temperature product merely underestimates T_{min} and overestimates T_{max} (Table S3a–c). The temperature products are able to better capture T_{mean} than T_{max} and T_{min} of the gauge observations with relatively higher CC values (Fig. 7 and Table 6). However, SAT estimates from all products show consistently improving patterns with an increase in temporal scale. This improvement could be because the different temperature datasets are relatively better-captured by ground measurements when compared at larger temporal scales than at smaller temporal scales. The temporal variations in T_{mean} , T_{max} , and T_{min} at the seasonal timescale are given in the supplementary information (Table S4) for spring, autumn, winter, and summer. Generally, all the temperature products overestimate the seasonal daily T_{mean} , T_{max} ,

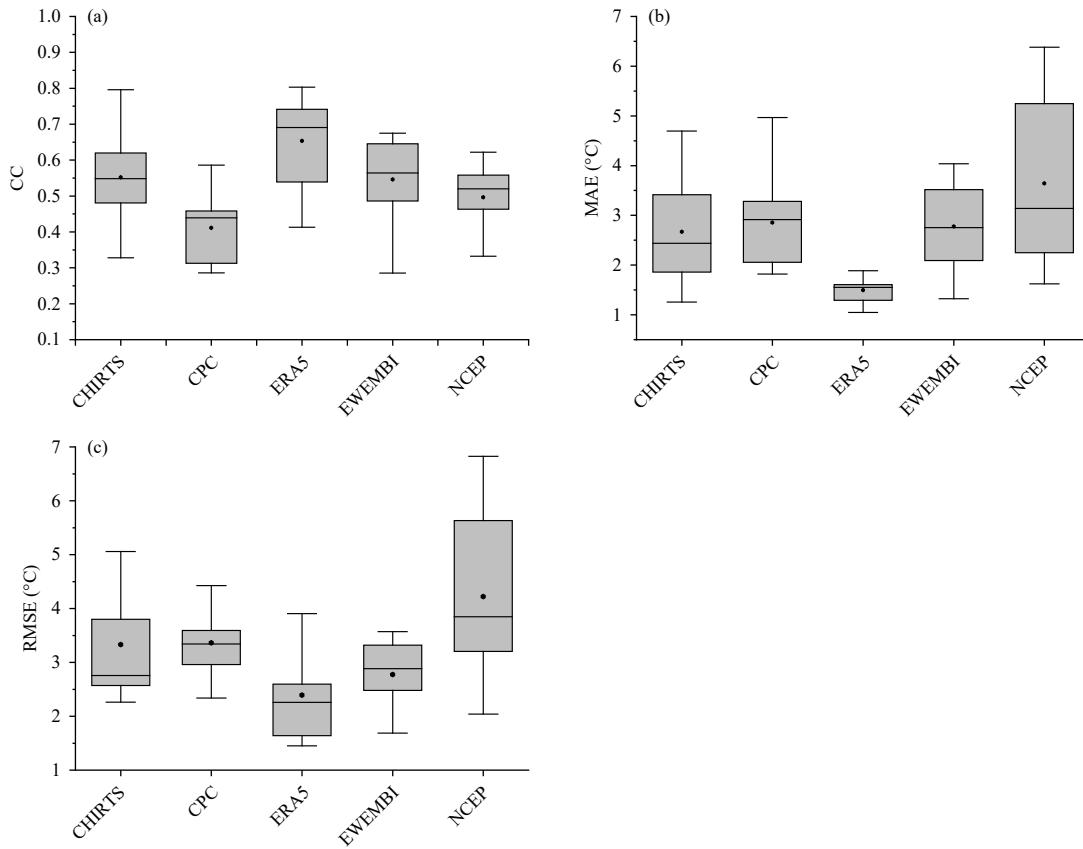


Fig. 7. Boxplots showing comparisons of T_{mean} between the five product estimates and the ground measurements at the daily timescale: (a) Pearson CC, (b) mean absolute error (MAE), and (c) root mean square error (RMSE).

Table 6. Comparison of T_{max} and T_{min} between the four products and the observed values at daily and monthly timescales

Metrics	Timescale	T_{max} (°C)				T_{min} (°C)			
		CHIRTS	CPC	EWEMBI	CRU TS	CHIRTS	CPC	EWEMBI	CRU TS
CC	Daily	0.54	0.36	0.43	—	0.56	0.37	0.49	—
	Monthly	0.69	0.61	0.62	0.68	0.72	0.63	0.68	0.7
RMSE	Daily	3.7	4.9	4.4	—	4.07	4.11	4.6	—
	Monthly	3.0	4.5	3.8	3.2	3.12	3.54	3.45	3.26
MAE	Daily	2.84	3.74	3.45	—	3.41	4.53	3.42	—
	Monthly	2.6	3.4	3.1	2.7	3.0	4.07	3.05	3.0

and T_{min} , except ERA5 (which underestimates summer T_{mean}) and CPC (which underestimates autumn T_{max}).

4.2.2 Temperature extreme indices

Table 7 presents daily values of the 95th percentile of T_{max} and T_{mean} and the 5th percentile values of T_{min} for all the SAT datasets. CPC and ERA5 have relatively closer values to the 95th percentile T_{max} and T_{mean} gauge-observed data, respectively. The high extreme SAT (95th percentile) from all the products are larger than those of the observations. EWEMBI shows a relatively better estimate of low extreme (5th percentile) T_{min} observed data (Table 7) than the other products.

In this study, the SAT values are considered to be highly extreme when their values are greater than the 95th percentile of the gauge SAT data (values greater

than 28.4 and 21.1°C for T_{max} and T_{mean} , respectively). Similarly, a low extreme is considered when the values of T_{min} are less than the 5th percentile of the gauge SAT (values < 8.2°C)

From the perspective of the 95th percentile long-term annual changes in T_{max} and T_{mean} presented in Fig. 8, each product consistently overestimates the SAT variables. With regard to the changing tendencies, OBS, CHIRTS, EWEMBI, ERA5, and NCEP show an increasing pattern, whilst the CPC product shows a significantly decreasing pattern. According to the CC results, the annual TS of the 95th percentile T_{mean} values of ERA5 and EWEMBI shows comparatively better correlations with the observed data than the other datasets, with CC values of 0.71 and 0.66, respectively. CPC shows the

Table 7. 95th percentile record for daily T_{\max} and T_{mean} and 5th percentile record for daily T_{\min} of the SAT products over UTB during 1982–2016

Dataset	5th percentile value ($^{\circ}\text{C}$)		95th percentile value ($^{\circ}\text{C}$)
	T_{\min}	T_{\max}	T_{mean}
OBS	8.2	28.4	21.1
CHIRTS	10.5	31	23.9
CPC	10	29.9	23.9
ERA5	—	—	22
EWEMBI	9	31.8	23.7
NCEP	—	—	25.4

lowest correlation with the observed data, with a CC value of 0.29 for the 95th percentile of T_{mean} (Fig. 8b). CHIRTS exhibits a relatively better accuracy in the 95th percentile annual TS values of T_{\max} (CC value of 0.67). Although the estimated 95th percentile T_{\max} values of CPC are closer than those of CHIRTS and EWEMBI to the observed data, the time-series correlation of CPC shows a poor performance with a CC value of less than 0.26 (Fig. 8a). The SAT datasets have poorly matched temporal variability of the T_{\min} extremes (5th percentile of T_{\min}) compared to the corresponding observed values. Relatively, EWEMBI is considered better (CC = 0.52) than CHIRTS (CC = 0.26) and CPC (CC = 0.30). Similarly, EWEMBI, CHIRTS, and CPC also overestimate the

T_{\min} yearly extremes (Fig. 8).

The number of days per year in which each SAT dataset scores greater than the 95th percentile of T_{\max} (28.4°C) and T_{mean} (21.1°C), and less than the 5th percentile of T_{\min} (8.2°C) is also evaluated. The mean estimated number of days with T_{\max} above 28.4°C is 50, 113, 55, and 126 days per year for OBS, CHIRTS, CPC, and EWEMBI, respectively (Fig. 9a). All SAT products overestimate the number of days with SAT greater than the 95th percentile value of T_{\max} (28.4°C). Among others, the CPC estimate is the closest to the observed values. The highest daily T_{\max} , which reaches a value of 37.7°C , is estimated by EWEMBI, followed by the observed data (36.7°C). Figure 9b shows the average number of days with T_{mean} values higher than 21.1°C ; the highest value is estimated by CHIRTS ($182.9 \text{ days yr}^{-1}$); the minimum number is recorded by ERA (56.6 days yr^{-1}), which is comparatively closer to the ground observations (49.8 days yr^{-1}) than the estimates of the other datasets. The largest daily T_{mean} value from the observed data is 30.8°C , whereas the largest mean daily SAT from the different datasets ranges from 26.6°C (EWEMBI) to 31°C (CPC). Similarly, the average number of days per year of T_{\min} less than 8.2°C is 20.4, 2.5, 4.9, and 10.5

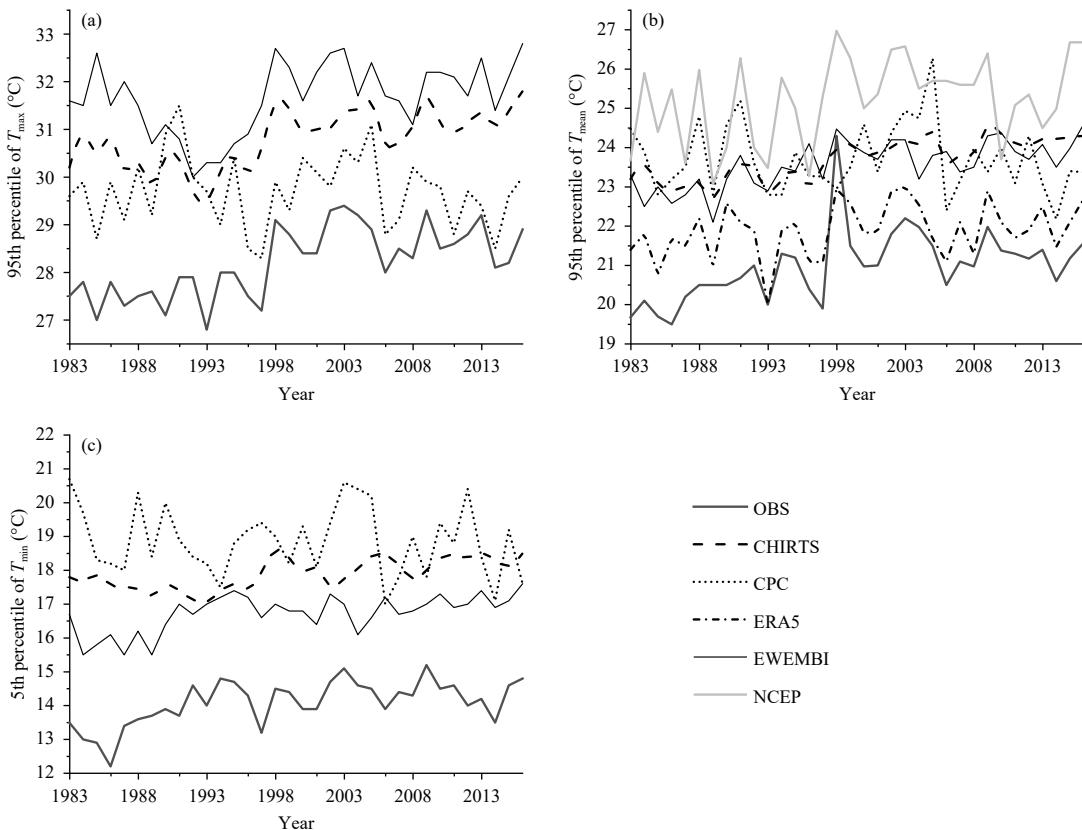


Fig. 8. Temporal changes in the annual 95th percentile values of (a) T_{\max} and (b) T_{mean} , and temporal changes in the 5th percentile for (c) T_{\min} of the different SAT datasets.

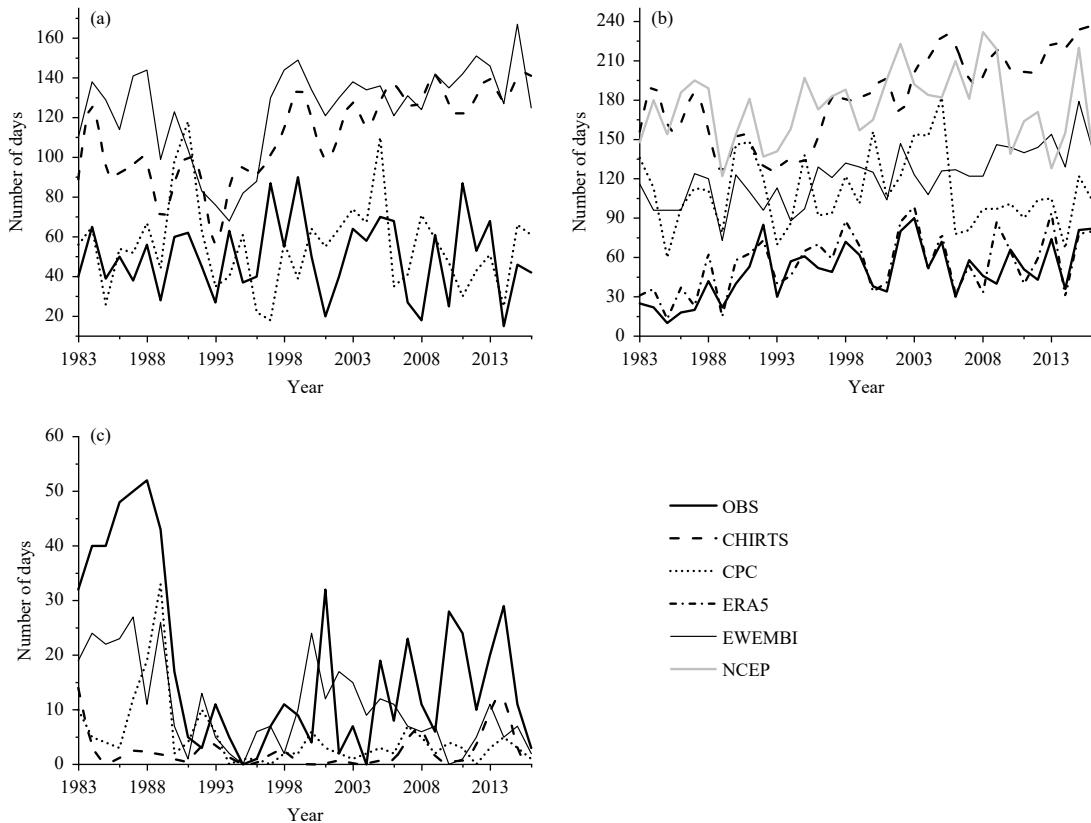


Fig. 9. Comparisons between observed values and the various daily SAT datasets in the number of days estimated above the 95th percentile for T_{\max} and T_{mean} and below the 5th percentile for T_{\min} : (a) annual number of days with daily T_{\max} above 28.4°C , (b) annual number of days with daily T_{mean} above 21.1°C , and (c) annual number of days with T_{\min} below 8.2°C . In this study, the scores over the specified SAT (95th percentile values of the observed data) denote occurrence of extreme SAT.

days for OBS, CHIRTS, CPC, and EWEMBI, respectively. The number of T_{\min} days below the 5th percentile value of the observed data is underestimated by all products, unlike the trends seen with T_{\max} and T_{mean} (Fig. 9c).

In the UTB, T_{mean} is better captured by ERA5, while T_{\max} and T_{\min} estimated by CHIRTS are in good agreement with the observed values. This implies that the best estimates of T_{mean} , T_{\max} , and T_{\min} are from different SAT products. This result agrees with the study of Nechita et al. (2019). The TS of SAT extremes of all the datasets, including the observed data, shows increasing trends. This could be an indication of the impact of climate change in the UTB. This argument could also be supported by Fentaw et al. (2018) in a climate change study in the UTB; the study results show an increasing trend in SAT.

5. Conclusions

In this study, nine satellite and reanalysis-based precipitation and SAT datasets were statistically evaluated

over the UTB. The EWEMBI, ERA5-land, GPCC v2018, CPC, WFDEI, NCEP reanalysis-2, CRU TS v4.03, CHIRPSv8, and CHIRTS products were evaluated against observed data. The precipitation, T_{\max} , T_{\min} , and T_{mean} were evaluated by using different statistical indices, including CC, Pbias, RMSE, and MAE against the ground measurements at daily and monthly timescales. These products were further evaluated by using different precipitation intensity groups and drought indices. Accordingly, the following conclusions are derived from this study.

EWEMBI, CHIRPSv8, and CRU TS v4.03 show good performances in representing daily and monthly precipitation. EWEMBI (CC = 0.86) shows reasonable and good agreement with the observed values, followed by the CHIRPS (CC = 0.85) and CRU TS v4.03 (CC = 0.85) products at the monthly timescale. The rainfall estimates from ERA5 and NCEP show poor performances as the daily rainfall is largely overestimated by these products. The monthly precipitation estimates usually perform better than the daily estimates. The precipitation products overestimate the lower precipitation intensity events (less

than 5 mm day⁻¹) but better estimate higher precipitation intensity events (greater than 20 mm day⁻¹). Moreover, the precipitation products underestimate historical drought events compared to the observed data. Specifically, the EWEMBI, CRU TS v4.03, and CHIRPS products show better agreement with the gauge-based drought ($SPI \leq -1$) estimation for each SPI timescale. Overall, the EWEMBI, CHIRPSv8, and CRU TS v4.03 precipitation products better represent the rainfall in the UTB than the other products.

ERA5, CHIRTS, and EWEMBI provide relatively better SAT estimates than the other products. T_{mean} estimates from ERA5, CHIRTS and EWEMBI show good agreement with gauge measurements. Similarly, T_{max} and T_{min} estimates of CHIRTS and EWEMBI at daily timescales and CHIRTS, CRU TS v4.03, and EWEMBI at monthly timescales have better agreement with observed values over the UTB than the other products. The SAT products have a poorer performance at the daily timescale than at the monthly timescale. In addition, T_{mean} is better represented than T_{max} and T_{min} by these products. SAT extremes are well-represented by ERA5 and EWEMBI (for T_{mean}), CHIRTS (for T_{max}) and EWEMBI (for T_{min}), while the CPC product shows poor performance in capturing the temperature extremes. The number of days with extreme temperature values is overestimated by all products. The overall deviation between the SAT products and the observed values is higher in the T_{min} estimates than in the T_{max} and T_{mean} estimates. In summary, depending on factors such as time series (daily and monthly) agreement and the estimation accuracy of the products for temperature extremes, the ERA5, CHIRTS, and EWEMBI temperature products are relatively better than the other products at representing the gauge-observed values in the UTB.

Acknowledgments. The first author was sponsored by the Chinese Academy of Sciences (CAS)—The World Academy of Sciences (TWAS) President's Fellowship Programme for his PhD study at the University of Chinese Academy of Sciences. The authors appreciate the Ethiopian National Meteorological Agency for providing the weather data.

REFERENCES

- Abrha, A. Z., 2009: Assessment of spatial and temporal variability of river discharge, sediment yield and sediment-fixed nutrient export in Geba River catchment, northern Ethiopia. Ph.D. dissertation, Katholieke Universiteit Leuven, Belgium, 203 pp.
- Agutu, N. O., J. L. Awange, A. Zerihun, et al., 2017: Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa. *Remote Sens. Environ.*, **194**, 287–302, doi: [10.1016/j.rse.2017.03.041](https://doi.org/10.1016/j.rse.2017.03.041).
- Aslami, F., A. Ghorbani, B. Sobhani, et al., 2019: Comprehensive comparison of daily IMERG and GSMAp satellite precipitation products in Ardabil Province, Iran. *Int. J. Remote Sens.*, **40**, 3139–3153, doi: [10.1080/01431161.2018.1539274](https://doi.org/10.1080/01431161.2018.1539274).
- Awange, J. L., K. X. Hu, and M. Khaki, 2019: The newly merged satellite remotely sensed, gauge and reanalysis-based Multi-Source Weighted-Ensemble Precipitation: Evaluation over Australia and Africa (1981–2016). *Sci. Total Environ.*, **670**, 448–465, doi: [10.1016/j.scitotenv.2019.03.148](https://doi.org/10.1016/j.scitotenv.2019.03.148).
- Ayoub, A. B., F. Tangang, L. Juneng, et al., 2020: Evaluation of gridded precipitation datasets in Malaysia. *Remote Sens.*, **12**, 613, doi: [10.3390/rs12040613](https://doi.org/10.3390/rs12040613).
- Azimi, S., A. B. Dariane, S. Modanesi, et al., 2020: Assimilation of Sentinel 1 and SMAP-based satellite soil moisture retrievals into SWAT hydrological model: the impact of satellite revisit time and product spatial resolution on flood simulations in small basins. *J. Hydrol.*, **581**, 124367, doi: [10.1016/j.jhydrol.2019.124367](https://doi.org/10.1016/j.jhydrol.2019.124367).
- Balsamo, G., C. Albergel, A. Beljaars, et al., 2015: ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi: [10.5194/hess-19-389-2015](https://doi.org/10.5194/hess-19-389-2015).
- Bayissa, Y., T. Tadesse, G. Demisse, et al., 2017: Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sens.*, **9**, 669, doi: [10.3390/rs9070669](https://doi.org/10.3390/rs9070669).
- Boke, A. S., 2017: Comparative evaluation of spatial interpolation methods for estimation of missing meteorological variables over Ethiopia. *J. Water Resour. Prot.*, **9**, 945–959, doi: [10.4236/jwarp.2017.98063](https://doi.org/10.4236/jwarp.2017.98063).
- Climate Change Service (C3S), 2019: ERA5-Land reanalysis. Copernicus Climate Change Service. Available online at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview>. Accessed on 20 March 2021.
- Chai, T., and R. R. Draxler, 2014: Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, **7**, 1247–1250, doi: [10.5194/gmd-7-1247-2014](https://doi.org/10.5194/gmd-7-1247-2014).
- Chen, M. Y., W. Shi, P. P. Xie, et al., 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res. Atmos.*, **113**, D04110, doi: [10.1029/2007JD009132](https://doi.org/10.1029/2007JD009132).
- Costa, R. L., G. M. de Mello Baptista, H. B. Gomes, et al., 2020: Analysis of climate extremes indices over northeast Brazil from 1961 to 2014. *Wea. Climate Extremes*, **28**, 100,254, doi: [10.1016/j.wace.2020.100254](https://doi.org/10.1016/j.wace.2020.100254).
- Degefu, M. A., and W. Bewket, 2015: Trends and spatial patterns of drought incidence in the Omo-Ghibe River basin, Ethiopia. *Geogr. Ann. Ser. A Phys. Geogr.*, **97**, 395–414, doi: [10.1111/geoa.12080](https://doi.org/10.1111/geoa.12080).
- Dembélé, M., and S. J. Zwart, 2016: Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *Int. J. Remote Sens.*, **37**, 3995–4014, doi: [10.1080/01431161.2016.1207258](https://doi.org/10.1080/01431161.2016.1207258).
- Dinku, T., P. Ceccato, E. Grover-Kopec, et al., 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, **28**, 1503–1526, doi: [10.1080/01431161.2007.9507901](https://doi.org/10.1080/01431161.2007.9507901).

- [01431160600954688](#).
- Fang, J., W. T. Yang, Y. B. Luan, et al., 2019: Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China. *Atmos. Res.*, **223**, 24–38, doi: [10.1016/j.atmosres.2019.03.001](#).
- Fentaw, F., D. Hailu, A. Nigussie, et al., 2018: Climate change impact on the hydrology of Tekeze Basin, Ethiopia: Projection of rainfall-runoff for future water resources planning. *Water Conserv. Sci. Eng.*, **3**, 267–278, doi: [10.1007/s41101-018-0057-3](#).
- Funk, C., P. Peterson, M. Landsfeld, et al., 2015: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data*, **2**, 150,066, doi: [10.1038/sdata.2015.66](#).
- Funk, C., P. Peterson, S. Peterson, et al., 2019: A high-resolution 1983–2016 T_{\max} climate data record based on infrared temperatures and stations by the climate hazard center. *J. Climate*, **32**, 5639–5658, doi: [10.1175/JCLI-D-18-0698.1](#).
- Gebremicael, T. G., Y. A. Mohamed, P. van der Zaag, et al., 2019: Evaluation of multiple satellite rainfall products over the rugged topography of the Tekeze-Atbara basin in Ethiopia. *Int. J. Remote Sens.*, **40**, 4326–4345, doi: [10.1080/01431161.2018.1562585](#).
- Gebremicael, T. G., Y. A. Mohamed, P. van der Zaag, et al., 2020: Change in low flows due to catchment management dynamics—Application of a comparative modelling approach. *Hydrol. Process.*, **34**, 2101–2116, doi: [10.1002/hyp.13715](#).
- Gyasi-Agyei, Y., 2020: Identification of the optimum rain gauge network density for hydrological modelling based on radar rainfall analysis. *Water*, **12**, 1906, doi: [10.3390/w12071906](#).
- Haile, G. G., Q. H. Tang, G. Y. Leng, et al., 2020: Long-term spatiotemporal variation of drought patterns over the Greater Horn of Africa. *Sci. Total Environ.*, **704**, 135299, doi: [10.1016/j.scitotenv.2019.135299](#).
- Harris, I., P. D. Jones, T. J. Osborn, et al., 2014: Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.*, **34**, 623–642, doi: [10.1002/joc.3711](#).
- Harris, I., T. J. Osborn, P. Jones, et al., 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, doi: [10.1038/s41597-020-0453-3](#).
- Harrison, L., C. Funk, and P. Peterson, 2019: Identifying changing precipitation extremes in Sub-Saharan Africa with gauge and satellite products. *Environ. Res. Lett.*, **14**, 085007, doi: [10.1088/1748-9326/ab2cae](#).
- Hu, S., H. J. Qiu, D. D. Yang, et al., 2017: Evaluation of the applicability of climate forecast system reanalysis weather data for hydrologic simulation: A case study in the Bahe River Basin of the Qinling Mountains, China. *J. Geogr. Sci.*, **27**, 546–564, doi: [10.1007/s11442-017-1392-6](#).
- Islam, A., and N. Cartwright, 2020: Evaluation of climate reanalysis and space-borne precipitation products over Bangladesh. *Hydrol. Sci. J.*, **65**, 1112–1128, doi: [10.1080/02626667.2020.1730845](#).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, et al., 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, doi: [10.1175/BAMS-83-11-1631](#).
- Lakew, H. B., S. A. Moges, and D. H. Asfaw, 2017: Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile Basin: A case study of Gilgel Abbay. *Hydrology*, **4**, 39, doi: [10.3390/hydrology4030039](#).
- Lange, S., 2016: Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI). GFZ Data Services. Available online at <https://dataservices.gfz-potsdam.de/pik/showshort.php?id=escidoc:1809891>. Accessed on 20 March 2021.
- Lange, S., 2018: Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth Syst. Dyn.*, **9**, 627–645, doi: [10.5194/esd-9-627-2018](#).
- Lange, S., 2019: Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI). V. 1.1. GFZ Data Services. Available online at <http://doi.org/10.5880/pik.2019.004>, doi: [10.5880/pik.2019.004](#). Accessed on 20 March 2021.
- Le Coz, C., and N. van de Giesen, 2020: Comparison of rainfall products over sub-Saharan Africa. *J. Hydrometeor.*, **21**, 553–596, doi: [10.1175/JHM-D-18-0256.1](#).
- Lemma, E., S. Upadhyaya, and R. Ramsankaran, 2019: Investigating the performance of satellite and reanalysis rainfall products at monthly timescales across different rainfall regimes of Ethiopia. *Int. J. Remote Sens.*, **40**, 4019–4042, doi: [10.1080/01431161.2018.1558373](#).
- Li, C. X., T. B. Zhao, C. X. Shi, et al., 2020: Evaluation of daily precipitation product in China from the CMA global atmospheric interim reanalysis. *J. Meteor. Res.*, **34**, 117–136, doi: [10.1007/s13351-020-8196-9](#).
- Li, D., G. Christakos, X. X. Ding, et al., 2018: Adequacy of TRMM satellite rainfall data in driving the SWAT modeling of Tiaoxi catchment (Taihu Lake basin, China). *J. Hydrol.*, **556**, 1139–1152, doi: [10.1016/j.jhydrol.2017.01.006](#).
- Liu, X. C., Z. X. Xu, and R. H. Yu, 2012: Spatiotemporal variability of drought and the potential climatological driving factors in the Liao River basin. *Hydrol. Process.*, **26**, 1–14, doi: [10.1002/hyp.8104](#).
- Liu, X. C., Q. H. Tang, H. J. Cui, et al., 2017: Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations. *Environ. Res. Lett.*, **12**, 025009, doi: [10.1088/1748-9326/aa5a3a](#).
- Lockhoff, M., O. Zolina, C. Simmer, et al., 2019: Representation of precipitation characteristics and extremes in regional reanalyses and satellite- and gauge-based estimates over western and central Europe. *J. Hydrometeor.*, **20**, 1123–1145, doi: [10.1175/JHM-D-18-0200.1](#).
- Mahmood, R., S. F. Jia, T. Mahmood, et al., 2020: Predicted and projected water resources changes in the Chari Catchment, the Lake Chad basin, Africa. *J. Hydrometeor.*, **21**, 73–91, doi: [10.1175/JHM-D-19-0105.1](#).
- Maidment, R. I., D. I. F. Grimes, R. P. Allan, et al., 2013: Evaluation of satellite-based and model re-analysis rainfall estimates for Uganda. *Meteorol. Appl.*, **20**, 308–317, doi: [10.1002/met.1283](#).
- Marques, C. A. F., A. Rocha, J. Corte-Real, et al., 2009: Global atmospheric energetics from NCEP–Reanalysis 2 and ECMWF–ERA40 Reanalysis. *Int. J. Climatol.*, **29**, 159–174, doi: [10.1002/joc.1704](#).
- McKee, T. B., N. J. Doesken, and J. Kleist, 1995: Drought Monitoring with Multiple Time Scales. Proc. 9th Conference on Applied Climatology, American Meteorological Society, Dallas, TX, 233–236.
- Ministry of Water Resources (MoWR), 1998: *Tekeze river basin integrated development master plan project*. Sectoral Report.

- Vol. VII, Addis Ababa, Ethiopia. Available online at <https://entrospace.nilebasin.org/handle/20.500.12351/383?show=full>. Accessed on 20 March 2021.
- Muller, J. C.-Y., 2014: Adapting to climate change and addressing drought-learning from the Red Cross Red Crescent experiences in the Horn of Africa. *Wea. Clim. Extremes*, **3**, 31–36, doi: [10.1016/j.wace.2014.03.009](https://doi.org/10.1016/j.wace.2014.03.009).
- Nechita, C., K. Čufar, I. Macovei, et al., 2019: Testing three climate datasets for dendroclimatological studies of oaks in the South Carpathians. *Sci. Total Environ.*, **694**, 133,730, doi: [10.1016/j.scitotenv.2019.133730](https://doi.org/10.1016/j.scitotenv.2019.133730).
- Nicholson, S. E., 2017: Climate and climatic variability of rainfall over eastern Africa. *Rev. Geophys.*, **55**, 590–635, doi: [10.1002/2016RG000544](https://doi.org/10.1002/2016RG000544).
- Peng, J., S. Dadson, F. Hirpa, et al., 2020: A pan-African high-resolution drought index dataset. *Earth Syst. Sci. Data*, **12**, 753–769, doi: [10.5194/essd-12-753-2020](https://doi.org/10.5194/essd-12-753-2020).
- Roy, A., P. K. Thakur, N. Pokhriyal, et al., 2018: Intercomparison of different rainfall products and validation of WRF modelled rainfall estimation in N-W Himalaya during monsoon period. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **IV-5**, 351–358, doi: [10.5194/isprs-annals-IV-5-351-2018](https://doi.org/10.5194/isprs-annals-IV-5-351-2018).
- Saccenti, E., M. H. W. B. Hendriks, and A. K. Smilde, 2020: Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models. *Sci. Rep.*, **10**, 438, doi: [10.1038/s41598-019-57247-4](https://doi.org/10.1038/s41598-019-57247-4).
- Sahlu, D., S. A. Moges, E. I. Nikolopoulos, et al., 2017: Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa. *Adv. Meteor.*, **2017**, 4957960, doi: [10.1155/2017/4957960](https://doi.org/10.1155/2017/4957960).
- Satgé, F., D. Defrance, B. Sultan, et al., 2020: Evaluation of 23 gridded precipitation datasets across West Africa. *J. Hydrol.*, **581**, 124412, doi: [10.1016/j.jhydrol.2019.124412](https://doi.org/10.1016/j.jhydrol.2019.124412).
- Schamm, K., M. Ziese, A. Becker, et al., 2014: Global gridded precipitation over land: a description of the new GPCC First Guess Daily product. *Earth Syst. Sci. Data*, **6**, 49–60, doi: [10.5194/essd-6-49-2014](https://doi.org/10.5194/essd-6-49-2014).
- Tan, M. L., P. W. Gassman, and A. P. Cracknell, 2017: Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins. *Water*, **9**, 229, doi: [10.3390/w9030229](https://doi.org/10.3390/w9030229).
- Tan, X. H., B. Yong, and L. L. Ren, 2018: Error features of the hourly GSMAp multi-satellite precipitation estimates over nine major basins of China. *Hydrol. Res.*, **49**, 761–779, doi: [10.2166/nh.2017.263](https://doi.org/10.2166/nh.2017.263).
- Tefera, A. S., J. O. Ayoade, and N. J. Bello, 2019: Comparative analyses of SPI and SPEI as drought assessment tools in Tigray Region, Northern Ethiopia. *SN Appl. Sci.*, **1**, 1265, doi: [10.1007/s42452-019-1326-2](https://doi.org/10.1007/s42452-019-1326-2).
- Toté, C., D. Patricio, H. Boogaard, et al., 2015: Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sens.*, **7**, 1758–1776, doi: [10.3390/rs70201758](https://doi.org/10.3390/rs70201758).
- Wang, N., W. B. Liu, F. B. Sun, et al., 2020: Evaluating satellite-based and reanalysis precipitation datasets with gauge-observed data and hydrological modeling in the Xihe River basin, China. *Atmos. Res.*, **234**, 104,746, doi: [10.1016/j.atmosres.2019.104746](https://doi.org/10.1016/j.atmosres.2019.104746).
- Weedon, G. P., G. Balsamo, N. Bellouin, et al., 2014: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, doi: [10.1002/2014WR015638](https://doi.org/10.1002/2014WR015638).
- Weedon, G. P., G. Balsamo, N. Bellouin, et al., 2018: The WFDEI meteorological forcing data. Research data archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Available online at [doi: 10.5065/486N-8109](https://doi.org/10.5065/486N-8109). Accessed on 20 March 2021.
- Worqlul, A. W., B. Maathuis, A. A. Adem, et al., 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, doi: [10.5194/hess-18-4871-2014](https://doi.org/10.5194/hess-18-4871-2014).
- Zeng, Q., H. Chen, C. Y. Xu, et al., 2018: The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. *J. Hydrol.*, **563**, 106–122, doi: [10.1016/j.jhydrol.2018.05.058](https://doi.org/10.1016/j.jhydrol.2018.05.058).
- Zhang, Y. S., and N. Ma, 2018: Spatiotemporal variability of snow cover and snow water equivalent in the last three decades over Eurasia. *J. Hydrol.*, **559**, 238–251, doi: [10.1016/j.jhydrol.2018.02.031](https://doi.org/10.1016/j.jhydrol.2018.02.031).
- Zhao, H. G., S. T. Yang, Z. W. Wang, et al., 2015: Evaluating the suitability of TRMM satellite rainfall data for hydrological simulation using a distributed hydrological model in the Weihe River catchment in China. *J. Geogr. Sci.*, **25**, 177–195, doi: [10.1007/s11442-015-1161-3](https://doi.org/10.1007/s11442-015-1161-3).
- Zhao, T. B., and C. B. Fu, 2006: Comparison of products from ERA-40, NCEP-2, and CRU with station data for summer precipitation over China. *Adv. Atmos. Sci.*, **23**, 593–604, doi: [10.1007/s00376-006-0593-1](https://doi.org/10.1007/s00376-006-0593-1).
- Zhao, T. B., W. D. Guo, and C. B. Fu, 2008: Calibrating and evaluating reanalysis surface temperature error by topographic correction. *J. Climate*, **21**, 1440–1446, doi: [10.1175/2007JCLI1463.1](https://doi.org/10.1175/2007JCLI1463.1).
- Ziese, M., A. Rauthe-Schöch, A. Becker, et al., 2018: GPCC Full Data Daily Version.2018 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. Available online at https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-daily_v2018_doi_download.html, doi: [10.5676/DWD_GPCC/FD_D_V2018_100](https://doi.org/10.5676/DWD_GPCC/FD_D_V2018_100). Accessed on 20 March 2021.