Novel approach to integrate daily satellite rainfall with in-situ rainfall, Upper Tekeze Basin, Ethiopia

Mewcha Amha Gebremedhin¹,a, Maciek W. Lubczynski¹, Ben H.P. Maathuis¹, Daniel Teka¹,b,c

¹ Department of Water Resources, Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, Enschede 7500AA, the Netherlands
² Institute of Geoinformation and Earth Observation Sciences (I-GEOS), Mekelle University, P.O.Box 231, Mekelle, Ethiopia
³ Department of Land Resource Management and Environmental Protection (LaRMEP), Mekelle University, P.O.Box 231, Mekelle, Ethiopia

**ARTICLE INFO**

**Keywords:** Satellite-rainfall, Downscaling, Validation, Bias-correction, Geographically-weighted-regression

**ABSTRACT**

The daily rainfall is the most important and demanded input of water resources studies, challenged by typically low density and/or poor quality of in-situ observations. However, the satellite earth observation, through freely available web-based products, can provide complementary rainfall data. Such data is however, typically affected by substantial error, particularly at daily temporal resolution. Therefore, effective methods and protocols of rainfall downscaling, validation, and bias-correction are needed. The aims of this study were to: i) validate two downscaled satellite-derived daily rainfall products, CHIRPS and MPEG, against in-situ observations; ii) merge the downscaled products with in-situ observations to improve their accuracy and evaluate them to select better performing one. This study was conducted at topographically complex, Upper Tekeze Basin (UTB), separately for the wet and dry seasons, within 1 January 2015 – 31 December 2018. Validation of the products, downscaled by nearest-neighbor (NN) and bilinear (BL) methods, was carried out using descriptive statistics, categorical statistics and bias decomposition methods, introducing novel protocol with new bias indicators for each of the evaluation methods. The validation showed large biases of CHIRPS and of MPEG, larger for CHIRPS than for MPEG, larger in dry than in wet season and slightly larger for NN than for BL. To correct biases of the downscaled products with in-situ observations to improve their accuracy and evaluate them to select better performing one, this study was conducted at topographically complex, Upper Tekeze Basin (UTB), separately for the wet and dry seasons, within 1 January 2015 – 31 December 2018. Validation of the products, downscaled by nearest-neighbor (NN) and bilinear (BL) methods, was carried out using descriptive statistics, categorical statistics and bias decomposition methods, introducing novel protocol with new bias indicators for each of the evaluation methods. The validation showed large biases of CHIRPS and of MPEG, larger for CHIRPS than for MPEG, larger in dry than in wet season and slightly larger for NN than for BL. To correct biases of the downscaled CHIRPS and MPEG, each was merged with the in-situ observed rainfall applying Geographically Weighted Regression (GWR) algorithm and using rainfall dependence on altitude as explanatory variable. The GWR-merging method substantially improved the accuracy of the MPEG and CHIRPS, with slightly better final accuracy of MPEG than of CHIRPS, better in wet than in dry season. This study confirmed that GWR-merged method could substantially reduce daily bias of satellite rainfall products, even in topographically complex areas, such as the UTB. Further improvement of the method application, can be achieved by densifying rain-gauge network and eventually by adding accuracy-effective explanatory variable(s).

1. Introduction

Water resources studies are challenged by the need of accurate data acquisition, at suitable, for a required problem, spatial and temporal data coverage (Gebere et al., 2015). Rainfall is the key driving force in hydrological studies, required to understand the complexity of water cycle and to manage water resources (Haile et al., 2009; Zambrano-Bigiarini et al., 2017). African agriculture and other water-related activities are dependent on quantity of rainfall and also on its spatio-temporal variability (Maidment et al., 2014). Though good quality and distribution of rainfall measurement is critical, the meteorological gauging station networks in Africa are not sufficient in quantity and also not well spatially distributed (Luetkemeier et al., 2018), i.e. particularly scarce in mountainous areas, where rainfall intensity is substantially higher than in lowlands and more variable in time and space (Rahmawati and Lubczynski, 2017). Therefore, obtaining high-quality rainfall data in mountainous areas, at fine spatial and temporal (e.g. daily) resolution, is a demanding but also challenging task (Hu et al., 2019).

In Developing Countries, particularly in high elevated areas with complex topography where rainfall is highly spatio-temporally variable but in-situ observations are scarce or unavailable (if available, then representative only locally), installing dense, spatially representative rain gauge network is not realistic as it requires large number of rain

https://doi.org/10.1016/j.atmosres.2020.105135
Received 4 May 2020; Received in revised form 20 June 2020; Accepted 12 July 2020
Available online 16 July 2020
0169-8095/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
gagues locally (Dinku et al., 2014; Alazzy et al., 2017). A common practice in water resource studies is interpolation of in-situ rainfall observations (Kebelotui et al., 2012; Camera et al., 2013; Bárdossy and Pegram, 2013; Gebremedhin et al., 2018; Kabasy et al., 2019). However, the interpolated rainfall from sparse observations is uncertain (Rahmawati and Lubczynski, 2017). This limitation is particularly critical in Ethiopian highlands where this study is situated and where there is relatively high and spatially variable rainfall (Haile et al., 2009), but sparsely distributed rainfall gauging network. Moreover, in Ethiopian highlands, obtaining continuous daily rainfall from in-situ observations is a challenge due to frequent rain gauge measurement discontinuities (Fenta et al., 2018). Hence, alternative data sources and methods of rainfall assessment are needed for water resources management and research, which require reliable spatio-temporal water input estimates.

Satellite-derived rainfall is a potential alternative or complementary source of spatio-temporal rainfall data, especially in complex and inaccessible terrains. Thermal Infrared (TIR) and Passive Microwave (PMW) sensors use standard algorithms to retrieve rainfall from satellites (Kidd, 2001). The TIR-based rainfall rates are inferred from cloud top temperatures assuming that rainfall and cold cloud duration are linearly correlated while the PMW-based approach provides atmospheric liquid water content and rain rates by penetrating clouds. However, the TIR and the PMW have their own limitations. The TIR sensors do not penetrate clouds (Kidd, 2001), underestimate warm orographic rain and have weak performance to discriminate cirrus clouds from rain clouds (Thiemig et al., 2013; Dinku et al., 2014), while the PMW sensors can confuse very cold surface, like mountain tops covered by ice, with rainfall (Tót et al., 2015).

Despite ongoing improvement of satellite-derived rainfall products, a site-specific validation of preselected rainfall products, to choose the most appropriate for a given study area is needed (Zambrano-Bigiarini et al., 2017; Fenta et al., 2018). Once the type of satellite product is selected, then usually, that product is downscaled and bias-corrected using ground-based rainfall measurements, before using it in hydrological studies. Satellite-derived rainfall products are subjected to errors reflected by the weak relationships between the remotely retrieved, pixel-wise and ground-based point-wise rainfall rates. Such errors can be due to: i) scale difference between the point-wise in-situ observations and the pixel-wise satellite rainfall products (Dinku et al., 2014; Rahmawati and Lubczynski, 2017; Lekula et al., 2018); ii) technical satellite sensor constrains (mentioned above); iii) impact of environmental factors; iv) temporal satellite sampling constrains, i.e. decline of accuracy with the increase of temporal resolution (Kidd, 2001).

The satellite-derived rainfall products are typically available at lower resolution (larger pixel size) than required by water resources projects, so they have to be downscaled. The nearest neighbor, bilinear, and bicubic are the common methods used to downscale the spatial resolution of pixels (Getreuer, 2011). For instance, Kimani et al. (2017) downscaled satellite-derived rainfall products using nearest neighbor and bilinear methods to enhance the spatial scale of satellite-derived products over east Africa. Their result indicated that the nearest neighbor outperformed bilinear downscaling method to represent the ground measured monthly rainfall. Other studies favored bilinear downscaling method to generate smooth interpolated satellite-derived rainfall before merging them with in-situ observations (e.g. Ulloa et al., 2017; Chen et al., 2019). So, the selection of a downscaling method requires testing at site-specific study.

Merging of satellite-derived rainfall with locally available in-situ observations can improve the accuracy of rainfall product (Dinku et al., 2011). There are several approaches to merge satellite-derived rainfall products and in-situ rainfall, including mean bias correction (Lekula et al., 2018), Bayesian merging approach (Todini, 2001; Ma et al., 2018; Kimani et al., 2018), objective analysis (Rozante et al., 2010) and optimal interpolation (Shen et al., 2014). However, such approaches lack consideration of explanatory variables that have a great impact on rainfall such as orographic influence (Shi et al., 2020). In contrast, the regression-based algorithms can consider explanatory variables during merging approach. Recent literature introduced a new regression-based method, i.e. Geographically Weighted Regression (GWR), also applied in this study, that can appropriately describe the spatial nonstationary relationship between rainfall and influencing explanatory variables (Hu et al., 2019). The GWR is one of the robust approaches, capable to efficiently improve quality of the satellite-derived rainfall (Wu et al., 2015; Chao et al., 2018).

Studies in Ethiopia, validated satellite-derived rainfall by comparing in-situ point station records with pixel rainfall values at different spatial and temporal scales. Tropical Rainfall Measuring Mission (TRMM), Climate Prediction Center Morphing technique (CMORPH) and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) are the most frequently validated satellite-derived products (Fenta et al., 2018). Dinku et al. (2008) evaluated these products over very complex topography in Ethiopia and less rugged topography of Zimbabwe. Their result indicated that all these satellite products had poor performance at the daily resolution, particularly over complex topography of Ethiopia. According to Dinku et al. (2011), that poor performance could be attributed to coarse spatial resolution, in which pixels averaged for the wet and dry areas might be misidentified as non-raining pixel.

The performance accuracy of satellite products to detect rainfall, increases towards coarser temporal resolution (Zambrano-Bigiarini et al., 2017; Gebrechorkos et al., 2018) due to temporal data accumulation (in a week or even better in a month) and the systematic error (Tian et al., 2007; Aghakouchak et al., 2012) cancelation (Fenta et al., 2018). Therefore, it is a challenge to find even a relatively accurate satellite-derived rainfall product at a daily time scale (Chen et al., 2018). However, it is the daily rainfall that is nowadays the most demanded in hydrology and water resources management, because the daily rainfall, is typically required as driving force input of integrated hydrological models (HM) used in water management.

There are only few satellite-based studies focused on daily rainfall assessment and even less focusing on Africa. Example of such studies includes Multi-Sensor Precipitation Estimate-Geostationary satellite (MPEG) by Worqulu et al. (2018). That study was conducted on the highlands of the upper Blue Nile Basin. The results indicated that before any bias correction, the MPEG poorly captured the in-situ observations, but after mean bias correction (MBC), it still indicated substantial rainfall underestimation. Another study on daily rainfall assessment was on Climate Hazards Group Infrared Rainfall with Stations (CHIRPS) by Fenta et al. (2018) in Lake Tana Basin (source of Blue Nile Basin). The result remarked that CHIRPS, with its relatively high spatio-temporal resolution, weakly captured rainfall variability at daily time scale in the topographically complex study area. The hitherto validation approaches, especially in Blue Nile Basin (Ethiopia), were not suitable for the Upper Tekeze Basin (UTB), because none of them fulfilled all the conditions required by this study, i.e.: i) high accuracy of rainfall data in topographically complex (mountainous) terrains; ii) daily temporal resolution; and iii) at least 1 km spatial resolution.

Evaluation of daily satellite rainfall products by comparing satellite with in-situ rainfall observations used to indicate large biases. The weak performance of satellite-derived products at daily time scale was mainly reasoned by the fine temporal resolution, high rainfall variability under topographically complex areas and the scale difference between satellite-derived rainfall products and in-situ observations (Rahmawati and Lubczynski, 2017). Moreover, selection of not optimal satellite rainfall evaluation or of stationary bias correction (e.g. MBC) methods, can lead to incorrect conclusion on the performance of satellite products. Hence, direct inputting of daily rainfall from satellite-derived rainfall products into hydrological models can propagate errors affecting models tremendously. As such, there is a need for development of a method allowing to merge daily satellite with ground-based data that will substantially improve the daily satellite rainfall estimate.
This study focused on daily rainfall data acquired from two satellite rainfall products, MPEG and CHIRPS providing rainfall estimates at 3 km and about 5 km spatial resolution respectively. These rainfall products were selected because they provide daily data at the highest available spatial resolution over the UTB study area. The objectives of this study were to evaluate the performance of the two satellite-derived rainfall products on daily basis by downscaling the daily satellite rainfall estimates to 1 km spatial resolution and to improve their accuracy by merging the downscaled rainfall products with in-situ observations and evaluate them to select better performing product to create accurate daily input data for the IHM (not part of this paper) of the UTB.

The main novelty of this study is in: i) new testing indices introduced to all steps of the validation of the daily satellite rainfall products; ii) use of the GWR method under semi-arid, complex topography in a data-scarce area; iii) important, performance-related information for potential users of MPEG and CHIRPS products; and iv) first time of daily satellite rainfall assessment in UTB.

2. Study area and data acquisition

2.1. Study area

The study area, UTB (Northern Ethiopia), is located between latitudes 12° 38′ 12″ and 13° 20′ 16″ N and longitudes 38° 59′ 23″ and 39° 40′ 05″ E. It covers an area of about 3500 km² (Fig. 1). Based on topographic information from the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model (DEM), the elevation of the study area ranges from 1230 m a.s.l. at the western catchment outlet to 3948 m a.s.l. in the eastern mountainous area. More than 50% of the study area has slopes of less than 15°, while ~12% of the area has slopes greater than 30°.

The rainfall of the study area is highly spatially and temporally variable. According to in-situ gauge measurements (Fig. 1) from 2015 to 2018, the mean annual rainfall varied from ~400 mm in the western lowland area, to ~940 mm in the eastern mountainous area. The elevation and mean annual rainfall are positively correlated in the study area, with \( r = 0.65 \) if considering all in-situ observations and much higher correlated (\( r = 0.9 \)) if excluding outlying Adigudem and Debub rainfall station. The study area experiences two main seasons: short, 4 months wet season from June to September, when more than 80% of a yearly rainfall falls with maximum rainfall during August. The remaining months represent dry season. Slightly different is the southeastern area around Maichew and Hashenge, where wet season rainfall is relatively lower, representing 73% and 78% of the yearly rainfall, respectively, with relatively higher rainfall during the dry season. The mean annual temperatures vary from about 11 °C in the eastern mountains to about 31 °C in the western lowlands. The highest mean monthly temperature is in May and the lowest in December.

2.2. Data acquisition

Ground-based and remote sensing CHIRPS and MPEG rainfall data were sourced for four years from 01 January 2015 to 31 December 2018 at a daily time resolution.

2.2.1. Ground-based data

Daily meteorological data for the ten stations (Fig. 1) were obtained from the Ethiopian National Meteorological Agency (NMA) at Tigray regional state. The nine observations were obtained from 1 January

Fig. 1. Location of the UTB with its elevation and location of in-situ rainfall gauging stations used in this study.
2.2.2. Remote sensing data

The CHIRPS satellite rainfall is a quasi-global (50° S - 50° N), gridded product with a spatial resolution of 0.05° (~5 km) and daily, pentadal, gridded, and monthly resolution (Funk et al., 2015). The CHIRPS algorithm takes advantage of integrating data sources from the Climate Hazards Group Precipitation Climatology (CHPclim), TIR-based satellite rainfall and in-situ measurement. The CHPclim uses long-term average satellite rainfall fields as guides to derive climatological surfaces (Funk et al., 2015). The daily CHIRPS rainfall product was freely downloaded from Climate Hazards Group via the link https://data.chc.ucsb.edu.

The MPEG satellite rainfall product covers the Earth from 60° S to 60° N with spatial resolution of 3 km and high temporal resolution of 15 min. It is produced by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) meteorological product extraction facility (MPEF). The MPEG is derived from the infrared data of the EUMETSAT geostationary satellites by continuous rescaling of the algorithm with rain rate data derived from polar-orbiting microwave sensors (Worporul, et al., 2018). The algorithm is based on a combination of Meteosat Second Generation (MSG) images from the infrared IR10.8 μm channel and passive microwave data from the Special Sensor Microwave/Imager (SSM/I) instrument on the United States Defence Meteorological Satellite Program (DMSP) polar satellites. The MPEG data is available through the GEONETCast near-real-time global network of satellite-based data dissemination systems, designed to distribute space-based, airborne and in-situ data. The MPEG product can be sourced freely from the International Institute for Geo-Information Science and Earth Observation (ITC) file transfer server: https://filetransfer.itc.nl/pub/mpe/. For this study, the 15 min time steps of MPEG data, was aggregated to daily time interval (00:00–23:45 UTC) using Integrated Land and Water Information System (ILWIS) software. The EUMETSAT-MPEG data is succeeded by the Hydrology SAF H05B product since 01 July 2019, which can be accessed through http://hsaf.meteoam.it or at https://filetransfer.itc.nl/pub/mpe/.

2.2.2. Remote sensing data

The CHIRPS satellite rainfall is a quasi-global (50° S - 50° N), gridded product with a spatial resolution of 0.05° (~5 km) and daily, pentadal, gridded, and monthly resolution (Funk et al., 2015). The CHIRPS algorithm takes advantage of integrating data sources from the Climate Hazards Group Precipitation Climatology (CHPclim), TIR-based satellite rainfall and in-situ measurement. The CHPclim uses long-term average satellite rainfall fields as guides to derive climatological surfaces (Funk et al., 2015). The daily CHIRPS rainfall product was freely downloaded from Climate Hazards Group via the link https://data.chc.ucsb.edu.

The MPEG satellite rainfall product covers the Earth from 60° S to 60° N with spatial resolution of 3 km and high temporal resolution of 15 min. It is produced by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) meteorological product extraction facility (MPEF). The MPEG is derived from the infrared data of the EUMETSAT geostationary satellites by continuous rescaling of the algorithm with rain rate data derived from polar-orbiting microwave sensors (Worporul, et al., 2018). The algorithm is based on a combination of Meteosat Second Generation (MSG) images from the infrared IR10.8 μm channel and passive microwave data from the Special Sensor Microwave/Imager (SSM/I) instrument on the United States Defence Meteorological Satellite Program (DMSP) polar satellites. The MPEG data is available through the GEONETCast near-real-time global network of satellite-based data dissemination systems, designed to distribute space-based, airborne and in-situ data. The MPEG product can be sourced freely from the International Institute for Geo-Information Science and Earth Observation (ITC) file transfer server: https://filetransfer.itc.nl/pub/mpe/. For this study, the 15 min time steps of MPEG data, was aggregated to daily time interval (00:00–23:45 UTC) using Integrated Land and Water Information System (ILWIS) software. The EUMETSAT-MPEG data is succeeded by the Hydrology SAF H05B product since 01 July 2019, which can be accessed through http://hsaf.meteoam.it or at https://filetransfer.itc.nl/pub/mpe/.

3. Methodology

3.1. General approach

To assess correlation among in-situ UTB rainfall observations, reflecting spatial rainfall variability, Pearson’s product-moment correlation coefficient (r) at daily time interval was used. Also, the relationship between correlation coefficient (r) attributing different pairs of gauges and corresponding distances was tested.

Preprocessing and analysis of the daily satellite rainfall were performed using the ILWIS open-source Water and Food Security Ethiopia Toolbox (Maathuis et al., 2018) and ArcGIS. As the performance of satellite-derived rainfall has to be validated and bias-corrected before using it in hydrological applications (Rahmawati and Lubczynski, 2017; Kimani et al., 2017; Lekula et al., 2018), the MPEG and CHIRPS satellite products were first validated using ground measurements to evaluate and compare their performances at the UTB. Considering CHIRPS satellite rainfall estimate, Navas et al. (2019) recommended that despite inherent blending procedure, the CHIRPS rainfall product can be still improved by merging with unused in the blending, in-situ observations. The list of rainfall stations used in this study was checked and except Maichew gauge record in 2018, none of the in-situ observations as per Fig. 1 and Table 1, acquired in the period 2015–2018, was used in the inherent CHIRPS blending procedure. A flowchart in Fig. 2 illustrates the main steps followed to validate satellite-derived rainfall and merging of the products with the in-situ observations. The evaluation methods used during the validation step, were again used after merging process to evaluate the performance of merged rainfall data.

3.2. Satellite-derived rainfall evaluation

The daily MPEG and CHIRPS rainfall products were downscaled to 1 km resolution using nearest neighbor (NN) and bilinear (BL) interpolation methods. These downsampling methods were used to enhance the spatial resolution of the satellite-derived rainfall products. To evaluate the downsampling approaches, the original and downscaled pixel-wise rainfall estimates of MPEG and CHIRPS, were compared with the coinciding, point-wise rainfall measured at each in-situ observation; in case in-situ observation was not available, the corresponding satellite data pixel was discarded from analysis. Descriptive statistics, categorical statistics and bias decomposition methods were applied to evaluate the performance of satellite rainfall products (Ayebu et al., 2018; Dinku et al., 2018; Lekula et al., 2018). That performance evaluation was carried out on daily basis and per season, i.e. separately for the wet (1 June – 30 September) and dry (1 October - 31 May) seasons, because the UTB is characterized with large amount of rain during the short-wet season compared to the long dry season.

3.2.1. Descriptive statistics

The descriptive statistics was used to compare the satellite-derived rainfall at pixel level with corresponding in-situ observations. The descriptive statistics used in this study include: i) Pearson’s product-moment correlation coefficient (r); ii) Mean Error (ME); iii) Mean Absolute Error (MAE); iv) Root Mean Square Error (RMSE); v) Normalized Mean Absolute Error (NMAE); and vi) Normalized Root Mean Square Error (NRMSE). The equations of the descriptive statistics are as follow:
3.2.2. Categorical statistics

The categorical statistics is a standard way to evaluate the detection capability of satellite-derived rainfall (Fenta et al., 2018; Lekula et al., 2018). Accordingly, the categorical statistics includes the following indices: i) Probability of Detection (POD); ii) False Alarm Ratio (FAR); iii) Critical Success Index (CSI); iv) Frequency Bias (FB); and v) Heidke Skill Score (HSS). These statistics are computed based on four combinations between satellite-derived rainfall and in-situ observations to verify the frequency of the correct and incorrect rainfall detection at a given temporal resolution. The four combinations are Hit (event of in-situ measured rainfall and satellite detected rainfall), Miss (event of in-situ measured rainfall, but satellite failed to detect rainfall), False alarm (event of no in-situ measured rainfall, but satellite detected rainfall) and Correct Negative (no in-situ measured rainfall and no rainfall detected by satellite). The expressions of the categorical statistics based on a contingency table are presented in Table 2.

All the categorical statistics indices mentioned above, except POD, are dependent on FA, which is not an adequate indicator to evaluate the performance of satellite-derived rainfall products (Rahmawati and Lubczynski, 2018). This is because a rainfall-cloud may move through an analyzed pixel, not covering the location of the in-situ observation (Lekula et al., 2018); in that case the gauge will not record rainfall while there will be rainfall within that pixel that might be properly recorded by the satellite; however, applying standard categorical statistics, such event will then be interpreted as FA, despite in reality the satellite properly detected a rain event. Hence, this study proposes two new categorical indicators: i) fraction of miss (FM), which is similar to POD but in contrast, in the numerator instead of $H$ includes more adequate $M$ (see Discussion); ii) fraction of detection (FD) indicator that involves $H$, $M$ and Correct Negative (CN) used to evaluate the overall detection capability of satellite product.

$$ FM = \frac{M}{H + M} $$

$$ FD = \frac{H + CN}{H + M + CN} $$

where $P_g$ is satellite-derived rainfall at day $i$; $P_s$ is in-situ rainfall at day $i$; the over bar $\bar{T}$ and $\bar{R}$ represent the average in-situ and average satellite rainfall, respectively; $T$ is the total number of daily data pairs. The range of values of $r$ is from $+1$ to $-1$, in which the value of 0 indicates that there is no linear relationship between the satellite-derived rainfall and in-situ observations, the value 1 indicates perfect positive linear correlation and the value $-1$ implies a perfect negative linear correlation; the ME ranges from $-\infty$ to $\infty$ with positive or negative values, which indicate overestimation or underestimation of rainfall by satellite product, respectively, while values close to zero indicate a close agreement; MAE, RMSE, $N_{MAE}$ and $N_{RMSE}$ range from 0 to $\infty$ with best value at 0. The MAE measures the average magnitude of the errors between the satellite and in-situ rainfall without considering their direction. The RMSE is a quadratic scoring rule, which is more reactive (sensitive) to rainfall outliers than the MAE. The $N_{MAE}$ and $N_{RMSE}$ provide relative to $\bar{P}_g$ measures of MAE and RMSE.
where \( FM \) is the ratio of number of wet days not detected by satellite to the total number of wet days recorded by a gauge \((H + M)\), where the \( FM \) ranges from 0 to 1, the former representing perfect satellite rainfall detection, while the \( FD \) is the overall capability of a satellite to detect the rain and no rain events where the \( FD \) ranges from 0 to perfect value of 1.

### Table 2

| Contingency matrix to evaluate the frequency of rainfall detection of CHIRPS and MPEG at daily time interval. |
|--------------------------------------------------|
| **Satellite detected rainfall** | **In-situ measured rainfall** | **No in-situ measured rainfall** |
| **Satellite detected rainfall** | \( H \) | \( FA \) |
| **No rainfall detected by satellite** | \( M \) | \( M.B \) |

\( H, M, FA \) and \( CN \) represent the number of days corresponding to their respective combinations.

#### 3.2.3. Bias decomposition

To get more insight into the source of satellite rainfall errors, the total bias between satellite-derived rainfall and in-situ observed rainfall, can be decomposed into Hit bias \((HB)\), Miss bias \((MB)\) and False alarm bias \((FAB)\) (Habib et al., 2009). However, the \( FAB \) which includes \( FA \), cannot be considered as a reliable bias indicator because of the same problem with as \( FA \), explained above. Therefore, this study used \( HB \), Absolute Hit Bias \((AHB)\) and \( MB \) to decompose and analyze the biases between satellite-derived rainfall estimates and in-situ observed rainfall. Besides, new bias decomposition indicators, i.e., Normalized Absolute Hit Bias \((N_{AHB})\) and Normalized Miss Bias \((N_{MB})\), are proposed to normalize the Absolute Hit and Miss biases by the total rainfall accumulated at a given gauge.

\[
HB = \sum_{t=1}^{T} (P_i - P_{\delta,i}); (P > 0 & P_{\delta} > 0) 
\]

\[
AHB = \sum_{t=1}^{T} |P_i - P_{\delta,i}|; (P > 0 & P_{\delta} > 0) 
\]

\[
MB = - \sum_{t=1}^{T} P_{\delta,i}; (P = 0 & P_{\delta} > 0) 
\]

\[
N_{AHB} = \frac{AHB}{\sum_{t=1}^{T} P_i} 
\]

\[
N_{MB} = \frac{MB}{\sum_{t=1}^{T} P_i} 
\]

where \( P_i \) and \( P_{\delta} \) are satellite and in-situ rainfall measures, respectively at valid day \( i \); \( T \) is the total number of valid daily data pairs.

\( HB \) is positive if the rainfall detected by satellite-derived product is higher than the rainfall measured at in-situ observation, which can be interpreted as the satellite overestimation of correctly observed ground-based rainfall while the negative value of the \( HB \) indicates an underestimation of correctly observed rainfall by satellite-derived products. The \( MB \) accumulates the missed rainfall by satellite-derived product and its sign is always negative. The signs of the \( AHB \) and \( N_{AHB} \) are positive while the sign of the \( N_{MB} \) is the same as of the \( MB \).

#### 3.3. Merging satellite-derived and in-situ rainfall

This study applied a geographically weighted regression (GWR) merging approach for bias correction of the satellite rainfall product. GWR is a robust algorithm that has been used in merging satellite-derived rainfall products with in-situ rainfall (Hu et al., 2015; Chao et al., 2018). Moreover, GWR is suitable to analyze the spatial relationship between rainfall events and their influencing factors, such as topography (Lv and Zhou, 2016). The general formula of GWR by Brunsdon et al. (1996) is as follows:

\[
Y_i = \beta_0 + \sum_{j=1}^{n} \beta_j (u_j, v_j)X_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n
\]

where \( Y_i \) is dependent and \( X_j \) independent variable at location \( i \); \( u_j \) and \( v_j \) are the geographical coordinates; \( \beta_0 \) is the intercept, \( \beta_j (u_j, v_j) \) is the constant regression coefficient for \( X_{ij} \), \( \varepsilon_i \) is the regression residual at the \( i \)th location, \( m \) is the number of independent variables and \( n \) is the number of observations.

Considering elevation as independent explanatory \((E)_i\), while the bias between satellite-derived rainfall and in-situ observations as dependent variable \((B)_i\), the Eq. 14 can be rewritten into Eq. 15.

\[
B_i = \beta_0 + a(x_i, y_i)E_i + \varepsilon_i
\]

where \( a(x_i, y_i) \) is corresponding regression coefficient at location \( i \) with projected coordinates of \( x_i \) and \( y_i \).

The regression coefficients are estimated using the following equation, after Fotheringham et al. (1998).

\[
a(x_i, y_i) = [E_i^TW(x_i, y_i)E_i]^{-1}E_i^TW(x_i, y_i)B
\]

where \( a(x_i, y_i) \) is regression coefficients at location \( (x_i, y_i) \); \( W(x_i, y_i) \) is weight matrix; \( E \) is explanatory variable (elevation); \( E_i^T \) is transposition of \( E \); \( B \) is the bias vector (difference between downscaled satellite rainfall and in-situ rainfall).

In GWR, the kernel type and bandwidth method need to be specified to assign the geographical weighting function. In this study, adaptive Gaussian kernel function (Fotheringham et al., 1998) was specified to compute the weighting function because that approach adapts well to the sparseness of the observation data and adjusts the bandwidth to optimize the spatial weight matrix (Kerm, 2003). The cross-validation method suggested by Lv and Zhou (2016) for local regression to determine the optimal bandwidth was specified. The optimal bandwidth estimated using leave-one-out cross-validation, which also indicates uncertainty, is inherent in the GWR method. The inverse distance weighted method was applied to interpolate the GWR residuals of each observation location, because this interpolation method can optimally smooth GWR residual in topographically complex area with

### Table 3

| Stations | Adigudem | Debub | Dengolat | Finarawa | Gijet | Hashenge | Maichew | Wedisemero | Hiwane | Borra |
|----------|----------|-------|----------|----------|------|----------|---------|------------|--------|-------|
|          | 1 (0)    | 1 (0) | 0.32 (23) | 0.26 (44) | 1 (0) | 0.32 (55) | 0.29 (67) | 0.21 (40) | 1 (0) | 0.32 (37) |
| Debub    | 0.32 (23) | 0.26 (44) | 0.32 (55) | 0.29 (67) | 0.21 (40) | 1 (0) | 0.32 (37) | 0.26 (82) | 0.22 (76) | 0.26 (88) | 1 (0) |
| Dengolat | 0.38 (23) | 0.26 (44) | 0.32 (55) | 0.29 (67) | 0.21 (40) | 1 (0) | 0.32 (37) | 0.36 (82) | 0.22 (76) | 0.26 (88) | 1 (0) |
| Finarawa | 0.32 (55) | 0.29 (67) | 0.32 (55) | 0.27 (57) | 0.40 (15) | 0.15 (31) | 1 (0) | 0.28 (76) | 0.28 (76) | 0.26 (88) | 1 (0) |
| Gijet    | 0.32 (37) | 0.42 (55) | 0.27 (57) | 0.40 (15) | 0.15 (31) | 1 (0) | 0.28 (76) | 0.28 (76) | 0.26 (88) | 1 (0) | 0.30 (52) |
| Hashenge | 0.28 (71) | 0.32 (54) | 0.36 (82) | 0.36 (82) | 0.22 (76) | 0.26 (88) | 1 (0) | 0.30 (52) | 0.31 (37) | 0.20 (24) | 0.39 (14) |
| Maichew  | 0.32 (51) | 0.32 (55) | 0.27 (63) | 0.31 (65) | 0.28 (72) | 0.44 (20) | 1 (0) | 0.30 (52) | 0.31 (37) | 0.20 (24) | 0.39 (14) |
| Wedisemero | 0.42 (57) | 0.32 (48) | 0.34 (62) | 0.18 (50) | 0.29 (65) | 0.32 (26) | 0.33 (22) | 1 (0) | 0.34 (44) | 0.20 (39) | 0.37 (47) |
| Hiwane   | 0.46 (15) | 0.34 (16) | 0.32 (30) | 0.28 (51) | 0.22 (42) | 0.33 (56) | 0.40 (36) | 0.40 (42) | 1 (0) | 0.34 (30) | 0.34 (30) |
| Borra    | 0.34 (44) | 0.20 (39) | 0.37 (47) | 0.25 (41) | 0.30 (52) | 0.31 (37) | 0.20 (24) | 0.39 (14) | 0.34 (30) | 1 (0) |
sparse in-situ observations (Camera et al., 2013).

The scale difference between point-wise in-situ measured rainfall and pixel-wise satellite rainfall estimate is one of the most important reasons of overestimation or underestimation of rainfall by satellite products (Haile et al., 2013; Dinku et al., 2014). To lower that difference, before the merging approach, it is important to downscale the product to finer resolution (Shi et al., 2020). Accordingly, in this study, after validation of the MPEG and CHIRPS products using the available in-situ, daily rainfall observations is presented in Table 3. The correlations between stations' distances do not explain the spatial variability of rainfall. Thirdly, the spatial regression coefficients and regression residuals were substituted into Eq. 15 to compute the bias at each grid cell – a sample of such uncertainty map, just for the last dates of the wet and dry seasons, can be seen in Appendix 1. Fourthly, the spatial biases were removed by subtracting them from the downscaled, satellite-derived rainfall, in order to obtain the bias-corrected rainfall at a daily time interval.

Finally, the bias corrected CHIRPS and MPEG rainfalls, were validated by the descriptive and bias decomposition statistics, applying the same procedure as in the validation stage, to: i) evaluate degree of improvement of the satellite performance, comparing the merged with non-merged rainfall products; ii) compare performances of the CHIRPS and MPGE to select one that will be further used for rainfall assessment in UTB. The pre-merging processes such as projecting, clipping and downscaling were carried out in ILWIS and ArcGIS environment. ArcGIS model builder was used to execute the GWR merging processes.

4. Results

4.1. In-situ gauge rainfall variability

Pearson’s product-moment correlation coefficient matrix of ten in-situ, daily rainfall observations is presented in Table 3. The correlations range from 0.12 to 0.46 while distances between the stations from 14 km to 88 km. Generally, a decreasing trend is observed between stations’ correlations and distances but this trend is not always followed. For instance, the correlation coefficient of the most distant (88 km) stations, Hashenge and Gijet, is 0.26 while between much closer (31 km), Gijet and Finarawa, is only 0.15. The results show that the stations’ distances do not explain the spatial variability of rainfall.

Table 4

| Station | $\bar{P}$ | ME | MAE | RMSE | ME | MAE | RMSE | ME | MAE | RMSE | NMAE | NRMSE | NMAE | NRMSE |
|---------|-----------|----|-----|------|----|-----|------|----|-----|------|------|-------|------|-------|
| Adigudem | 3.39 | 0.85 | 4.50 | 8.65 | 0.99 | 4.52 | 8.87 | −0.05 | 4.23 | 8.97 | −0.07 | 4.22 | 8.91 | 1.33 | 2.62 | 1.24 | 2.63 |
| Debub | 3.74 | 0.30 | 4.49 | 10.08 | 0.31 | 4.47 | 9.90 | −0.76 | 4.10 | 9.88 | −0.76 | 4.09 | 9.84 | 1.20 | 2.65 | 1.09 | 2.64 |
| Dengolat | 5.96 | −0.71 | 6.82 | 11.96 | −0.58 | 6.65 | 11.33 | −2.30 | 5.78 | 9.63 | −2.20 | 5.75 | 9.59 | 1.12 | 1.90 | 0.96 | 1.61 |
| Finarawa | 2.52 | 1.94 | 4.40 | 8.35 | 1.87 | 4.34 | 8.13 | 1.11 | 3.77 | 7.65 | 1.17 | 3.82 | 7.67 | 1.72 | 3.23 | 1.51 | 3.05 |
| Gijet | 4.84 | 0.77 | 6.60 | 10.72 | 0.70 | 6.49 | 10.55 | −0.35 | 6.22 | 11.02 | −0.40 | 6.20 | 11.02 | 1.34 | 2.18 | 1.28 | 2.28 |
| Hashenge | 5.05 | −0.46 | 5.46 | 10.42 | −0.40 | 5.25 | 9.73 | −2.17 | 4.73 | 9.78 | −2.17 | 4.71 | 9.79 | 1.04 | 1.93 | 0.93 | 1.94 |
| Maichew | 3.45 | 1.25 | 4.74 | 9.07 | 1.21 | 4.63 | 8.70 | −0.51 | 3.46 | 7.62 | −0.53 | 3.46 | 7.61 | 1.34 | 2.52 | 1.00 | 2.21 |
| Wedisemero | 6.14 | −0.75 | 7.21 | 13.22 | −0.75 | 7.08 | 12.96 | −3.02 | 6.17 | 12.13 | −3.00 | 6.17 | 12.14 | 1.15 | 2.11 | 1.00 | 1.98 |
| Hiwane | 2.96 | 0.74 | 3.89 | 8.75 | 0.80 | 3.87 | 8.61 | −0.27 | 3.45 | 7.88 | −0.23 | 3.42 | 7.85 | 1.31 | 2.91 | 1.15 | 2.65 |
| Borra | 4.46 | 0.51 | 5.95 | 10.15 | 0.52 | 5.94 | 10.13 | −1.11 | 5.02 | 9.41 | −1.14 | 5.03 | 9.48 | 1.33 | 2.27 | 1.13 | 2.13 |
| AVG | 4.25 | 0.44 | 5.40 | 10.14 | 0.47 | 5.32 | 9.89 | −0.94 | 4.69 | 9.40 | −0.93 | 4.69 | 9.39 | 1.29 | 2.43 | 1.13 | 2.31 |

Fig. 3. Pearson’s product-moment correlation coefficients between daily in-situ rainfall measurements and NN and BL downscaled CHIRPS and MPEG, for the wet (1 June – 30 September) and dry (1 October – 31 May) seasons within 1 January 2015 – 31 December 2018.
4.2. Satellite-derived rainfall evaluation using descriptive statistics

The variability of the correlation coefficient ($r$) estimated between downscaled CHIRPS and MPEG and the corresponding, ground rainfall measurements during wet and dry seasons, are presented in Fig. 3. It can be observed that: i) during the wet season the correlation is generally better than in the dry season but the spatial correlation patterns, are similar; ii) at a higher elevation, CHIRPS has a better correlation with in-situ observation than MPEG but worse at lower elevation; iii) BL-CHIRPS estimation has slightly higher $r$ than NN-CHIRPS while BL-MPEG and NN-MPEG, similar.

The $ME$, $MAE$ and $RMSE$ for the wet and dry seasons are presented in Table 4 and Table 5 respectively. In the wet season (Table 4), the $ME$ shows rainfall differences at different altitudes of the in-situ observations for both products using both downscaling methods. The $ME$ of CHIRPS shows large overestimation at the lowland (Finarawa) with better BL downscaling values (closer to zero) than with NN. The $ME$ of MPEG shows also overestimation at the lowland (Finarawa), but in contrast to CHIRPS, in all other, higher elevated stations, the rainfall is underestimated. The mean rainfall underestimation was larger at the MPEG than the overestimation at the CHIRPS but in both satellite products, the BL downscaling had slightly better $ME$ than NN downscaling. Considering the wet season average of $MAE$ and $RMSE$ (Table 4), they are generally lower for MPEG than for CHIRPS and
Table 6  
Bias decomposition of the daily CHIRPS and MPEG rainfall (in mm), totalized over wet seasons (1 June – 30 September) within 1 January 2015 – 31 December 2018: \( P_t \) – gauge accumulated total rainfall (mm); \( HB \) - hit bias (mm); \( AHB \) - absolute hit bias (mm); \( MB \) – miss bias (mm); \( N_{AHB} \) - normalized AHB and \( N_{MB} \) - normalized MB.

| Stations | \( P_t = \sum_{i=1}^{N} P_{i,t} \) | CHIRPS | MPEG | CHIRPS | MPEG | CHIRPS | MPEG |
|----------|------------------------------|--------|-------|--------|-------|--------|-------|
|          | \( HB \) | \( AHB \) | \( MB \) | \( HB \) | \( AHB \) | \( MB \) | \( N_{AHB} \) | \( N_{MB} \) | \( N_{AHB} \) | \( N_{MB} \) |
| Adigudem | 1515.2 | 418.7 | 991.9 | -500.3 | -160.0 | 1166.1 | -294.6 | 0.65 | -0.33 | 0.77 | -0.19 |
| Debub    | 1666.1 | 157.1 | 1029.3 | -491.4 | -429.4 | 1396.9 | -166.9 | 0.62 | -0.29 | 0.84 | -0.10 |
| Dengolat | 2635.6 | 467.5 | 1494.0 | -1085.8 | -646.0 | 1883.5 | -493.3 | 0.57 | -0.41 | 0.71 | -0.19 |
| Finarawa | 735.9  | 75.4  | 321.4  | -237.7 | -75.0  | 600.1  | -49.2  | 0.44 | -0.32 | 0.82 | -0.07 |
| Gijet    | 1912.1 | 138.5 | 1121.9 | -652.7 | -572.4 | 1347.3 | -343.5 | 0.59 | -0.34 | 0.70 | -0.18 |
| Hashenge | 2170.8 | -130.5 | 998.6 | -649.8 | -906.7 | 1260.2 | -396.4 | 0.46 | -0.30 | 0.58 | -0.18 |
| Maichew  | 1542.3 | 414.8 | 1091.7 | -425.8 | -172.4 | 1192.9 | -208.5 | 0.71 | -0.28 | 0.77 | -0.14 |
| Wedisemero | 2729.4 | 61.1  | 1676.7 | -939.3 | -1182.8 | 2070.7 | -417.6 | 0.61 | -0.34 | 0.76 | -0.15 |
| Hiwane   | 1223.3 | 66.6  | 533.0  | -400.2 | -222.4 | 818.5  | -232.9 | 0.44 | -0.33 | 0.67 | -0.19 |
| Borra    | 1953.0 | 180.7 | 779.3  | -888.3 | -721.8 | 1424.6 | -279.2 | 0.40 | -0.45 | 0.73 | -0.14 |
| AVG      | 1809.4 | 185.0 | 1003.8 | -627.1 | -508.9 | 1316.1 | -288.2 | 0.55 | -0.34 | 0.74 | -0.15 |

lower for BL than for NN, with some few stations (e.g. Adigudem and Gijet), where small relative differences are observed. Also, the wet season \( N_{MAE} \) and \( N_{MSE} \) clearly show that MPEG outperforms CHIRPS.

In dry season (Table 5), the ME of CHIRPS shows underestimation at the higher elevated stations with slight overestimation at the lowland stations. On average, unlike overestimation in the wet season, CHIRPS underestimated the rainfall during the dry season. The ME of MPEG shows underestimation at all stations on average larger than in CHIRPS. In CHIRPS, the NN downscaling indicated slightly better ME than BL downscaling while in MPEG, the ME was the same. Considering the dry season averages of \( MAE \) and \( RMSE \), they were lower for MPEG than for CHIRPS and lower for BL than for NN, as in the wet season. The dry season normalized \( MAE \) and \( RMSE \) (\( N_{MAE} \) and \( N_{MSE} \)) show that MPEG outperformed CHIRPS but also that these errors were lower during wet seasons than during dry seasons.

4.2.1. Satellite-derived rainfall detection capabilities using categorical statistics

The frequencies of rainfall detection success (\( H \) - hit), failure (\( M \) - miss) and indices (\( FM \) - fraction of miss and \( FD \) - fraction of detection) of CHIRPS and MPEG at the daily resolution for the wet and dry seasons, are presented in Fig. 4. The analysis of that Fig. 4, shows that: i) the MPEG provides better rainfall detection capability than CHIRPS because it has substantially higher \( H \), lower \( M \), substantially lower \( FM \) and higher \( FD \), all these characteristics pointing at superiority of MPEG, more distinct in wet than in dry season; ii) the BL-downscaling in general, provides better rainfall detection capability than NN-downscaling, except of CHIRPS in dry season, when the NN-downscaling shows better rainfall detection. Overall, the BL downscaling method outperforms NN in descriptive and categorical evaluations methods, therefore, in the follow up analysis (bias decomposition and merging process) only the BL downscaling is used.

4.3. Bias decomposition of satellite-derived rainfall

The biases of CHIRPS and MPEG were decomposed into \( HB \), \( AHB \) and \( MB \) to quantify the magnitude of correctly detected rainfall and magnitude of missed rainfall in the wet (Table 6) and dry (Table 7) seasons, separately. In the wet season, CHIRPS overestimates observations except in Hashenge station while MPEG substantially underestimates in all stations. Considering \( AHB \), the overall bias in MPEG is higher by \( \sim31\% \) than in CHIRPS, but the \( MB \) is much lower in MPEG than in CHIRPS at all gauge measurements.

The \( HB \), \( AHB \) and even \( MB \) show large differences between the satellite rainfall products and in-situ observed rainfall, which are attributed not only to satellite performance but also to differences in gauge accumulated total rainfall (\( P_t \)) between stations, which justify introduction of \( N_{AHB} \) and \( N_{MB} \) (Eqs. 12 and 13, see Table 6 and Table 7). In wet season (Table 6), the \( N_{AHB} \) in both rainfall products showed poor performance (>50% bias), being worse in MPEG. In contrast, the \( N_{MB} \) was better in MPEG than in CHIRPS.

During the dry season (Table 7), the average \( HB \) shows that CHIRPS underestimated rainfall, unlike in the wet season when overestimated with small bias at the lowland (Finarawa). The MPEG underestimated the rainfall at all stations, which is consistent with the wet season characteristics. Considering \( AHB \), the overall bias in MPEG is higher by \( \sim50\% \) than in CHIRPS but the \( MB \) at all gauge measurements, in MPEG is much lower than in CHIRPS. The maximum \( MB \) for CHIRPS and
MPEG is observed in the highland stations. In dry season (Table 7), like in wet season, the $N_{\text{MAE}}$ was better in CHIRPS but in contrast, the $N_{\text{MB}}$ was better in MPEG than in CHIRPS.

In the dry season, as expected (because of generally lower $P_d$), the $HB$, $AHB$ and $MB$ are smaller than in the wet season. However, interesting is comparison of the normalized characteristics of the dry and wet seasons. The $N_{\text{MAE}}$ indicates better performance of CHIRPS and of MPEG in the dry season than in the wet season while the $N_{\text{MB}}$ other way round, i.e. it shows better performance in the wet season.

4.4. Merging satellite and in-situ observation rainfall

The correlation coefficient ($r$) estimated between GWR-merged CHIRPS and MPEG rainfall and pixel-corresponding, ground rainfall measurement during wet and dry seasons is presented in Fig. 5. It is observed that the $r$ is noticeably improved as compared to the down-scaled rainfall products, being close to the perfect linear correlation value (~1). This indicates that the daily rainfall is well captured by CHIRPS and MPEG after the GWR merging approach for both, the wet and dry seasons.

The descriptive statistics of error measurement results after GWR merging is demonstrated in Table 8 and Table 9 for the wet and dry seasons, respectively. All the measures presented, show that the GWR merging approach substantially improved the accuracy of both rainfall products. Considering the wet season, the ME shows only a slight overestimation by CHIRPS and MPEG. The MAE and RMSE also demonstrated that the GWR merging approach improved the accuracy of rainfall data at all in-situ observations. The largest MAE and RMSE were at the highlands; for example in wet season, before merging, at Wedisemero, in CHIRPS 7.08 mm day$^{-1}$ and 12.96 mm day$^{-1}$ and in MPEG, 6.17 mm day$^{-1}$ and 12.14 mm day$^{-1}$ (Table 4) and after merging decreased to 0.45 mm day$^{-1}$ and 1.06 mm day$^{-1}$ for CHIRPS and 0.42 mm day$^{-1}$ and 1.04 mm day$^{-1}$ for MPEG (Table 8), respectively. The before-merging, average ME in MPEG indicating underestimation (~0.93 mm day$^{-1}$), was also substantially improved to a slight overestimation (0.11 mm day$^{-1}$). Considering average ME, the GWR-approach showed a similar improvement in CHIRPS and MPEG while regarding the average MAE and RMSE, the GWR-approach improved MPEG slightly better than CHIRPS. The $N_{\text{MAE}}$ and $N_{\text{RMSE}}$ also show that MPEG outperformed CHIRPS after merging during the wet season.

In the dry season, the ME showed slight overestimation by CHIRPS and MPEG after GWR-merging, substantially reducing error in both products as compared to before merging with similar improvement in CHIRPS and MPEG. The MAE and RMSE also demonstrated that the GWR merging improved the accuracy of rainfall at all in-situ observations. For example, before merging, the largest MAE and RMSE were at the highlands; at Hashenge station, in CHIRPS, 1.82 mm day$^{-1}$ and 6.03 mm day$^{-1}$, and in MPEG, 1.34 mm day$^{-1}$ and 5.32 mm day$^{-1}$.

Table 8

| Stations      | $\overline{P}_i$ | ME  | MAE | RMSE | ME  | MAE | RMSE | $N_{\text{MAE}}$ | $N_{\text{RMSE}}$ | $N_{\text{MAE}}$ | $N_{\text{RMSE}}$ |
|---------------|------------------|-----|-----|------|-----|-----|------|------------------|------------------|------------------|------------------|
| Adigudem      | 3.39             | 0.01| 0.09| 0.18 | 0.00| 0.10| 0.40 | 0.03             | 0.05             | 0.03             | 0.12             |
| Debub         | 3.74             | 0.07| 0.22| 0.68 | 0.02| 0.05| 0.14 | 0.06             | 0.18             | 0.01             | 0.04             |
| Dengolat      | 5.96             | 0.05| 0.56| 1.41 | -0.11| 0.25| 0.43 | 0.09             | 0.24             | 0.04             | 0.07             |
| Finarawa      | 2.52             | 0.11| 0.27| 0.86 | 0.01| 0.06| 0.16 | 0.11             | 0.34             | 0.02             | 0.06             |
| Gijet         | 4.84             | 0.32| 0.91| 1.52 | 0.58| 0.90| 2.65 | 0.19             | 0.31             | 0.19             | 0.55             |
| Hashenge      | 5.05             | 0.00| 0.09| 0.17 | -0.03| 0.14| 0.93 | 0.02             | 0.03             | 0.03             | 0.18             |
| Maichew       | 3.45             | 0.18| 0.44| 0.86 | 0.15| 0.26| 0.81 | 0.13             | 0.25             | 0.08             | 0.24             |
| Wedisemero    | 6.14             | 0.12| 0.45| 1.06 | 0.26| 0.42| 1.04 | 0.07             | 0.17             | 0.07             | 0.17             |
| Hiwane        | 2.96             | 0.08| 0.24| 0.66 | 0.00| 0.07| 0.32 | 0.08             | 0.22             | 0.02             | 0.11             |
| Borra         | 4.46             | 0.15| 0.88| 1.75 | 0.24| 0.79| 1.52 | 0.20             | 0.39             | 0.18             | 0.29             |
| AVG           | 4.25             | 0.11| 0.42| 0.92 | 0.11| 0.30| 0.82 | 0.10             | 0.22             | 0.07             | 0.18             |
Table 9
Descriptive statistics of dry seasons daily rainfall (1 October – 31 May) after GWR merging: ME - mean error [mm day⁻¹]; MAE - mean absolute error [mm day⁻¹]; RMSE - root mean square error [mm day⁻¹]; N_MAE - normalized MAE (MAE/PT); and N_RMSE - normalized RMSE (RMSE/PT) analyzed using data within 1 January 2015 – 31 December 2018. PT - gauge mean rainfall [mm day⁻¹].

| Stations  | PT  | ME   | MAE  | RMSE | ME   | MAE  | RMSE | N_MAE | N_RMSE | N_MAE | N_RMSE |
|-----------|-----|------|------|------|------|------|------|-------|--------|-------|--------|
| Adigudem  | 0.31| 0.01 | 0.01 | 0.05 | 0.00 | 0.01 | 0.04 | 0.04  | 0.15   | 0.03  | 0.13   |
| Debub     | 0.74| 0.02 | 0.03 | 0.16 | 0.01 | 0.01 | 0.11 | 0.04  | 0.22   | 0.02  | 0.14   |
| Dengolat  | 0.91| −0.01| 0.06 | 0.23 | −0.01| 0.04 | 0.17 | 0.07  | 0.25   | 0.05  | 0.19   |
| Finarawa  | 0.52| 0.03 | 0.03 | 0.18 | 0.02 | 0.02 | 0.16 | 0.06  | 0.35   | 0.03  | 0.30   |
| Gijet     | 0.61| 0.10 | 0.15 | 0.50 | 0.07 | 0.11 | 0.41 | 0.24  | 0.82   | 0.19  | 0.67   |
| Hashenge  | 1.34| 0.00 | 0.02 | 0.06 | −0.01| 0.02 | 0.07 | 0.01  | 0.05   | 0.01  | 0.05   |
| Maichew   | 1.18| 0.05 | 0.17 | 0.72 | 0.05 | 0.06 | 0.20 | 0.14  | 0.61   | 0.05  | 0.17   |
| Wedisemero| 0.89| 0.08 | 0.09 | 0.30 | 0.10 | 0.10 | 0.36 | 0.11  | 0.34   | 0.12  | 0.40   |
| Hiwane    | 0.30| 0.02 | 0.05 | 0.39 | 0.01 | 0.01 | 0.07 | 0.17  | 1.32   | 0.05  | 0.23   |
| Borra     | 0.39| 0.12 | 0.19 | 0.58 | 0.12 | 0.18 | 0.54 | 0.49  | 1.49   | 0.48  | 1.39   |
| AVG       | 0.72| 0.04 | 0.08 | 0.32 | 0.04 | 0.06 | 0.21 | 0.14  | 0.56   | 0.10  | 0.37   |

(Tables 5 and 6) and after merging decreased to 0.02 mm day⁻¹ and 0.06 mm day⁻¹ for CHIRPS, and to 0.02 mm day⁻¹ and 0.07 mm day⁻¹ for MPEG (Table 9), respectively. The lowering of dry season MAE and RMSE indicate that GWR-approach improved MPEG better than CHIRPS. The dry season MAE and RMSE are smaller than in wet season because of lower PT. However, the lower N_MAE and N_RMSE in wet than in dry season indicate better performance of both products in wet than in dry season but lower N_MAE and N_RMSE for MPEG than for CHIRPS in wet and dry seasons, indicate generally better performance of MPEG than of CHIRPS.

The wet season bias estimates of GWR-merged CHIRPS and MPEG are presented in Table 10. They indicate that: i) all the bias estimates (HB, AHB and MB) show substantial improvement after GWR merging approach in both CHIRPS and MPEG; ii) the HB in CHIRPS and MPEG have positive and negative values, so no specific trend can be observed; iii) the AHB of MPEG in most of the stations is lower than in CHIRPS; iv) the MB in all gauges is lower in MPEG than in CHIRPS; v) N_AHB and N_MB are generally very low but lower for MPEG than for CHIRPS.

The dry season bias estimates of GWR-merged CHIRPS and MPEG are presented in Table 11. They indicate that: i) all the bias estimates (HB, AHB and MB) show substantial improvement after GWR merging approach in both CHIRPS and MPEG; ii) all the biases are smaller in dry than in wet season because of smaller PT; iii) the HB in CHIRPS and MPEG have positive and negative values, so like in case of wet season no specific trend can be observed; iv) the AHB in most of the stations, is lower in MPEG than in CHIRPS; v) the MB in CHIRPS is small, < 2 mm, but in MPEG is negligible at all stations; v) N_AHB and N_MB are generally very low but lower for MPEG than for CHIRPS.

5. Discussion

The relationship among the in-situ rainfall station measurements (Fig. 1) in UTB is not dependent on distances between the stations. This indicates that spatial rainfall variability can differ from region to region due to other influencing factors (e.g. altitude). Although the in-situ observations in UTB are better represented by altitudinal variation than by distances, they are too sparsely distributed to properly represent the spatial rainfall variability for example by interpolation. Hence, satellite-derived rainfall products with high temporal and relatively high spatial resolution, like CHIRPS (~5 km) and MPEG (3 km), offer sources of complementary rainfall data. However, CHIRPS and MPEG, as any other rainfall product, are prone to errors, in UTB mainly because of the high daily rainfall variability and the scale difference between the pixel-wise satellite-derived rainfall versus the point-wise in-situ rainfall observation. Downscaling methods can narrow down the scale difference, particularly in topographically complex areas, like UTB. The NN downscaling method does not change the original pixel rainfall in both satellite-derived rainfall products because the new pixel value is determined by matching to the corresponding position of the original pixel, which helps to preserve original values. Conversely, BL method changes the original pixel rainfall values in both products because it uses a weighted average of four nearest pixels to assign rainfall value to the new pixel. This study indicated that by considering in-situ observations as a reference, the BL method, downscaled slightly better the CHIRPS and MPEG products than NN. This result disagreed with the study by (Kimani et al., 2017) where NN method outperformed BL method with high RMSE difference at monthly time scale. This disagreement might be attributed to the temporal resolution difference.

Table 10
Bias decomposition of the daily GWR-merged rainfall (in mm), totalized over wet seasons (1 June – 30 September) within 1 January 2015 – 31 December 2018: PT - gauge accumulated total rainfall (mm); HB - hit bias (mm); AHB - absolute hit bias (mm); MB - miss bias (mm); N_AHB - normalized AHB and N_MB - normalized MB.

| Stations     | PT  | HB   | AHB  | MB   | CHIRPS | N_AHB | N_MB  | MPEG  | N_AHB | N_MB  |
|--------------|-----|------|------|------|--------|-------|-------|--------|-------|-------|
| Adigudem     | 1515.2| −5.2 | 31.4 | 0.0  | −6.6   | 38.3  | 0.0   | 0.02   | 0.00  | 0.03  |
| Debub        | 1666.1| 7.6  | 76.7 | −0.1 | 4.6    | 18.5  | 0.0   | 0.05   | 0.00  | 0.01  |
| Dengolat     | 2635.6| 1.4  | 215.4| −5.3 | −54.2  | 103.1 | −0.9  | 0.08   | 0.00  | 0.04  |
| Finarawa     | 735.9 | 4.4  | 30.6 | −10.7| 0.0    | 14.0  | 0.0   | 0.04   | −0.01 | 0.02  |
| Gijet        | 1912.1| 12.4 | 217.5| −13.7| 115.9  | 235.6 | −4.3  | 0.11   | −0.01 | 0.12  |
| Hashtenge    | 2170.8| −9.4 | 28.6 | 0.0  | −11.4  | 24.2  | −17.5 | 0.01   | 0.00  | 0.01  |
| Maichew      | 1542.3| 50.9 | 131.7| −16.5| 38.9   | 85.2  | −0.6  | 0.09   | −0.01 | 0.06  |
| Wedisemero   | 2739.4| 47.4 | 171.7| −11.1| 101.7  | 171.6 | −2.7  | 0.06   | 0.00  | 0.06  |
| Hiwane       | 1223.3| −0.5 | 68.8 | 0.0  | −7.8   | 22.2  | 0.0   | 0.06   | 0.00  | 0.02  |
| Borra        | 1953.0| −64.7| 252.8| −2.5 | −34.1  | 207.5 | 0.0   | 0.13   | 0.00  | 0.11  |
| AVG          | 1809.4| 4.4  | 122.5| −6.0  | 14.7   | 92.0  | −2.6  | 0.07   | 0.00  | 0.05  |
scale of study area and rainfall variability. Other studies by Din et al. (2008) and Ulloa et al. (2017) demonstrated, that in-situ observed rainfall better correlated with BL downscaled satellite-derived rainfall than with the original rainfall product, which was also the case in this study.

In descriptive statistics, the mean error (ME) is a good measure of the systematic error to characterize the underestimation or overestimation performance of satellite-derived rainfall products; for example, in wet season analysis, it indicated that the MPEG rainfall was highly underestimated, even more than in CHIRPS. However, considering quantitative assessment of error, the ME is not conclusive because in time series analysis, underestimation and overestimation errors can cancel out. The MAE and RMSE are more reliable quantitative error measures. The MAE and RMSE were slightly smaller in the BL than in NN and smaller in MPEG than in CHIRPS, with the smallest BL-MPEG combination suggesting the best rainfall detection. Remarkable however was the range of the MAE and RMSE errors, as in all the in-situ gauges, they exceeded the mean daily rainfall ($P_T$).

The N$_{NRMSE}$ provides higher difference in the dry season than in the wet season, because of being prone to bias, erratic rainfalls in dry season that are better captured by RMSE than by MAE as the former is more sensitive to outliers.

According to the categorical statistics, MPEG provided higher frequency of hit rainfall events ($H$) and lower missed rainfall events ($M$) compared to CHIRPS in both, wet and dry seasons. The low $M$ is more reliable than high $H$ to evaluate the performance of satellite rainfall products as it refers to the low failures of satellite algorithm to detect rainfall recorded at in-situ observations. The $H$ indicator is also valuable to evaluate the ability of satellite products to detect rainfall events recorded at the ground; however, there is a risk that different showers can be detected by the satellite than those measured at the ground (Rahmawati and Lubczynski, 2017; Lekula et al., 2018). Accordingly, the MPEG satellite-derived product performed better than CHIRPS during both seasons, which might be attributed to its higher spatial and temporal resolution of data acquisition. The fraction of miss (FM), new indicator introduced in this study, also suggests that MPEG outperformed CHIRPS in detecting the true rainfall occurrence at the ground in both, wet and dry seasons (Fig. 4). The lower FM in MPEG is because of the lower $M$ and higher $H$ compared to CHIRPS product. The higher FD in MPEG than in CHIRPS in both seasons, indicates that the former better captures the rainfall than the latter. However, compared to FM, the FD is less reliable indicator (Eq. 8) because of its low sensitivity to the satellite performance when correct negative (CN) is high, such as in long dry season in UTB with only a few scattered rainfall events but lots of CN. The categorical indicators do not consider the rainfall magnitude (in

![Fig. 6.](image-url) Box plots of N$_{NMAE}$ and N$_{NRMSE}$ for CHIRPS and MPEG during the wet (1 June – 30 September) and dry (1 October – 31 May) seasons within 1 January 2015 – 31 December 2018, before (a, b) and after (c, d) GWR-merging. Note the scale differences.

### Table 11

Bias decomposition of the daily GWR-merged rainfall (in mm), totalized over dry seasons (1 October – 31 May) within 1 January 2015 – 31 December 2018: $P_T$– gauge accumulated total rainfall (mm); $HB$ - hit bias (mm); $AHB$ - absolute hit bias (mm); $MB$ - miss bias (mm); $N_{AHB}$ - normalized $AHB$ and $N_{MB}$ - normalized $MB$.

| Stations  | $P_T = \sum_{i=1}^{N} P_i$ | CHIRPS | MPEG | CHIRPS | MPEG |
|-----------|-----------------------------|--------|------|--------|------|
|           |                             | HB     | AHB  | MB     | N$_{AHB}$ | N$_{MB}$ | N$_{AHB}$ | N$_{MB}$ |
| Adigudem  | 287.5                       | −0.9   | 5.9  | 0.0    | 0.4   | 4.7   | 0.0    | 0.2    | 0.0    |
| Debub     | 692.4                       | 6.1    | 15.3 | −0.7   | 6.2   | 7.2   | 0.0    | 0.02   | 0.00   |
| Dengolat  | 774.7                       | −18.3  | 40.9 | −0.3   | −12.1 | 31.2  | 0.0    | 0.05   | 0.00   |
| Finarawa  | 309.4                       | 2.9    | 7.2  | 0.0    | 3.9   | 4.9   | 0.0    | 0.02   | 0.02   |
| Gijet     | 555.3                       | 8.8    | 54.1 | 0.0    | 5.3   | 44.0  | 0.0    | 0.10   | 0.00   |
| Hashenge  | 1257.8                      | −7.9   | 10.5 | 0.0    | −10.5 | 12.4  | 0.0    | 0.01   | 0.00   |
| Maichew   | 1106.4                      | 15.4   | 115.9| −1.8   | 34.1  | 45.3  | 0.0    | 0.10   | 0.00   |
| Wedisemero| 835.6                       | −12.0  | 54.7 | −0.8   | 51.4  | 56.3  | 0.0    | 0.07   | 0.00   |
| Hiwane    | 276.6                       | −5.5   | 31.4 | 0.0    | 0.3   | 3.6   | 0.0    | 0.08   | 0.00   |
| Borra     | 299.7                       | −19.9  | 37.2 | 0.0    | −15.4 | 34.8  | 0.0    | 0.12   | 0.00   |
| AVG       | 639.5                       | 2.8    | 36.3 | −0.4   | 6.3   | 24.4  | 0.0    | 0.06   | 0.00   |

In descriptive statistics, the mean error ($ME$) is a good measure of the systematic error to characterize the underestimation or overestimation performance of satellite-derived rainfall products; for example, in wet season analysis, it indicated that the MPEG rainfall was highly underestimated, even more than in CHIRPS. However, considering quantitative assessment of error, the $ME$ is not conclusive because in time series analysis, underestimation and overestimation errors can cancel out. The $MAE$ and $RMSE$ are more reliable quantitative error measures. The $MAE$ and $RMSE$ were slightly smaller in the BL than in NN and smaller in MPEG than in CHIRPS, with the smallest BL-MPEG combination suggesting the best rainfall detection. Remarkable however was the range of the $MAE$ and $RMSE$ errors, as in all the in-situ gauges, they exceeded the mean daily rainfall ($P_T$).
than by a gauge observation (Lekula et al., 2018). Besides, in MB rainfall quantity, this study introduced normalized bias indicators i.e. dry seasons, is related with the high frequency and variability of rainfall CHIRPS and MPEG towards the high elevated stations for the wet and magnitude of satellite rainfall product biases. The increasing AHB by CHIRPS and MPEG towards the high elevated stations for the wet and dry seasons, is related with the high frequency and variability of rainfall occurrence. To account in the analysis for bias dependence on the rainfall quantity, this study introduced normalized bias indicators i.e. \( N_{AHB} \) and \( N_{MB} \), both computing ratio of accumulated AHB and MB to the total gauge accumulated rainfall, respectively (Fig. 7). These are more reliable bias indicators than standard AHB and MB, as they account rainfall biases relative to the rainfall quantity. The higher \( N_{AHB} \) and the lower \( N_{MB} \) in CHIRPS and MPEG in the wet season than in the dry season (Fig. 7 a,b) indicate better performance of both satellite rainfall products in the wet season, because the bias increases with increased rainfall intensity of individual showers (Hu et al., 2019), which is typically lower in wet season, also in UTB.

The overestimation of CHIRPS as compared to in-situ rainfall can be attributed to the limitation of TIR sensor to identify cirrus clouds from rain clouds (Thiemig et al., 2013; Young et al., 2014; Dinku et al., 2014; Toté et al., 2015). Unlike the finding by Kimani et al. (2017) that CHIRPS underestimated the in-situ observation at the altitude < 2500 m a.s.l. during wet season, CHIRPS at UTB overestimated in-situ observation during the wet season. This supports the opinions of Zambrano-Bigiarini et al. (2017) and Fenta et al. (2018) who remarked that satellite-derived products might show different performances from region to region due to regional and/or local factors, therefore always require validation and bias correction before using them in any hydrological studies.

The performance of MPEG in this study is in line with studies by Derin and Yilmaz (2014); Dhib et al. (2017) and Worqlul et al. (2018), i.e. is also substantially underestimated as compared to in-situ observations. That underestimation, especially high at the highlands, is likely attributed to the limitation of a product to detect warm orographic rainfall expected at mountainous areas (Worqlul et al., 2014; Dhib et al., 2017). Moreover, in highland parts of Ethiopia, including UTB, localized convective light rainfalls might not have been captured because the PMW sensors have tendencies to underestimate light rainfall (Rahmawati and Lubczynski, 2017).

The performance of both satellite-derived rainfall products at UTB, even with relatively high spatio-temporal resolution (downscaled to 1 km), indicated large biases compared to gauge measured rainfall. This required improvement, i.e. bias correction of both products. There are different bias correction methods. The mean bias correction (MBC) is a commonly used approach by deriving additive or multiplicative factors from the difference or ratio of in-situ observations and satellite-derived rainfall (Habib et al., 2014; Lekula et al., 2018; Worqlul et al., 2018). However, the MBC cannot reflect the non-stationary, spatial relationship between satellite-derived rainfall and in-situ rainfall influenced by land surface conditions; such approaches assume simplistic uniform bias distribution over the spatial domain (Nerini et al., 2015; Lv and Zhou, 2016). The applied in this study GWR method was used to merge CHIRPS and MPEG with in-situ observations considering altitude as explanatory variable because the spatial rainfall variability at UTB is correlated with altitudinal variation.

The Fig. 6 shows that after GWR merging (c, d) the \( N_{MAE} \) and \( N_{RMSE} \) were substantially improved in both, wet and dry seasons as compared to original, before GWR-merging (a, b) state of the rainfall. It also shows better performance of MPEG than of CHIRPS and better performance of both products in wet than in dry season, which is consistent with the \( N_{MAE} \) and \( N_{RMSE} \) from before GWR-merging. Remarkable is relatively large GWR-merged \( N_{RMSE} \) (with its variability range) of CHIRPS in dry season, much larger than corresponding \( N_{RMSE} \) of MPEG, which is attributed to better response of MPEG than CHIRPS to scattered, intense (outliers), UTB rainfall in the dry season.

The Fig. 7 shows great improvement of \( N_{MAE} \) and \( N_{MB} \) after GWR-merging, for both CHIRPS and MPEG, but also indicating similar, normalized bias magnitudes (in case of \( N_{MB} \) close to zero) in both, wet and dry seasons (c, d), unlike before merging (a, b), when the differences and the normalized biases themselves, were much larger. The very small \( N_{MB} \) after GWR-merging of MPEG, but also of CHIRPS, indicate that the biases, at least in locations close to stations, are small relative to the rainfall quantities in these locations for most seasons.

The time series example graphs, comparing the in-situ Bora daily rainfall with corresponding daily rainfall in CHIRPS (Appendix 2) and MPEG (Appendix 3) before and after merging indicate that GWR approach substantially improved time series projection of rainfall as compared to original, non-corrected products, better for MPEG than for CHIRPS rainfall. The general GWR rainfall improvement agrees well with the study by Chao et al. (2018) who stated that merging in-situ observation with satellite rainfall using GWR is an effective tool for producing improved rainfall data, particularly for the areas with sparse gauge observations and characterized with high rainfall variability. The improvement of CHIRPS rainfall data after the GWR merging approach confirms also the observation by Navas et al. (2019), who stated that CHIRPS rainfall product can be still improved by merging new in-situ
observations, despite the inherent blending procedure. In contrast, the MPEG is a pure satellite product with no inherent ground-based calibration, but it provided better rainfall accuracy than CHIRPS before and after the GWR merging, which can be attributed to the higher spatio-temporal resolution, so better capability of the product to detect the frequency of rainfall events. The GWR results demonstrated the effectiveness of the merging method and its high accuracy, tested over 10 gauges available at the UTB, but there is still uncertainty away of the gauges that can be mitigated by installation of additional gauges. Considering the GWR-method implementation in the UTB, further improvement considering other explanatory variables can also be tested.

6. Conclusion

This study validated and bias corrected, the downscaled to 1 km spatial resolution CHIRPS and MPEG satellite-derived rainfall products at daily resolution. The validation was carried out over downscaled products, using descriptive statistics, categorical statistics and bias decomposition methods, separately for wet and dry seasons, introducing new bias indicators for each of the evaluation methods. To improve the accuracy of the rainfall products, the bias correction was carried out, applying GWR method of merging satellite with in-situ observations and using rainfall dependence on altitude as explanatory variable. From this study, the following main conclusions can be drawn:

1. The satellite rainfall products are sources of complementary to ground-based rainfall data, particularly in areas with sparse rain gauges and high spatio-temporal rainfall variability, provided they are downscaled (if necessary), validated, and bias corrected, but applying effective methods, appropriate for the study area of concern (such as GWR for UTB).

2. The bilinear downscaling method worked in UTB slightly better than nearest-neighbor for both satellite products.

3. The ME is useful indicator of a systematic error to characterize the underestimation or overestimation performance of satellite but quantitatively not conclusive, as the underestimation and overestimation errors can cancel out when analysing time series rainfall data. The MAE and RMSE are reliable quantitative measures of rainfall performance, but they are dependent on quantity of rainfall analyzed, so are not suitable for comparison of different rainfall data sets, such as for example rainfalls of the wet and dry seasons. That role was optimally performed by normalized measures of MAE and RMSE, i.e., $N_{\text{MAE}}$ and $N_{\text{RMSE}}$, introduced in this study.

4. Considering rainfall frequency measures, the ‘miss’ is more reliable than the ‘hit’, while the ‘false alarm’ is not reliable so it was not used in this study; to avoid ‘false alarm’ in commonly used indices, ‘fraction of miss’ and ‘fraction of detection’ modified indices were introduced. All the frequency measures, indicated better MPEG performance in the UTB than CHIRPS.

5. The bias decomposition is a reliable method to evaluate the performance of satellite rainfall products as it quantifies the accumulated magnitude of biases. The ‘absolute hit bias’ and the ‘miss bias’ are reliable measures of rainfall performance, but they are dependent on quantity of rainfall analyzed, so are not suitable for comparison of different rainfall data sets, such as rainfalls of the wet and dry seasons. That role was optimally performed by normalized measures of $A_{\text{HB}}$ and $M_{\text{B}}$, i.e. $N_{\text{A_{HB}}}$ and $N_{\text{M_{B}}}$, introduced in this study.

6. The bias correction of the MPEG and CHIRPS rainfall products by GWR-merging algorithm with rainfall dependence on altitude as explanatory variable, showed substantial improvement, decreasing the bias between satellite-derived rainfall and in-situ observations in both products, although with slightly better final accuracy of MPEG than of CHIRPS.

7. The GWR method has inherent assessment of uncertainty, which typically propagates away of rainfall observation points. The way to reduce uncertainty of a satellite product, is to increase the number of merged ground stations and eventually to condition the GWR method with more explanatory variables, if relevant.

Declaration of Competing Interest

None.

Acknowledgement

This study was supported by the Dutch organisation for internationalisation in education (Nuffic), University of Twente, Faculty of Geo-information Science and Earth Observation (ITC) and Ministry of Science and Higher Education of Ethiopia (MoSHE) under the Ethiopian Educational Network to Support Agricultural Transformation (EENSAT) project (CF13198, 2016). We would like to gratefully acknowledge the Ethiopian National Meteorological Agency (NMA) for providing daily rainfall data at multiple stations in the study area.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2020.105135.

References

Din, S., Al-Dousari, A., Ramdan, A., Al Ghabban, A., 2008. Site-specific precipitation estimate from TRMM data using bilinear weighted interpolation technique: an example from Kuwait. J. Arid Environ. 72, 1520–1528. https://doi.org/10.1016/j.jaridenv.2007.12.013.

AghaKouchak, A., Mehran, A., Norouzi, H., Behranghi, A., 2012. Systematic and random error components in satellite precipitation data sets. Geophys. Res. Lett. 39, 3–6. https://doi.org/10.1029/2012GL051592.

Alazzy, A.A., Liu, H., Chen, R., Ali, A.B., Zhu, Y., Su, J., 2017. Evaluation of satellite precipitation products and their potential influence on hydrological modeling over the Ganzi River Basin of the Tibetan Plateau. Adv. Meteorol. https://doi.org/10.1155/2017/369285.

Ayeulu, G.T., Tadesse, T., Gesesse, B., Dinku, T., 2018. Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. Atmos. Meas. Tech. 11, 1921–1936. https://doi.org/10.5194/amt-11-1921-2018.

Bärdossy, A., Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water Resour. Res. 49, 4545–4565. https://doi.org/10.1002/wrcr.20307.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically Weighted Regression: a Method for Exploring Spatial Nonstationarity. Geogr. Anal. 28, 281–298.

Chen, S., Zhang, L., She, D., Chen, J., 2019. Spatial Downscaling of Tropical Rainfall Measuring Mission (TRMM) Annual and Monthly Precipitation Data over the Middle and Lower Reaches of the Yangtze River Basin, China. Water 11568, 1–17. https://doi.org/10.3390/w110310568.

Chao, L., Zhang, K., Li, Z., Zhu, Y., Wang, J., Yu, Z., 2018. Geographically weighted regression based methods for merging satellite and gauge precipitation. J. Hydrol. 558, 275–289. https://doi.org/10.1016/j.jhydrol.2018.01.042.

Chen, F., Gao, Y., Wang, Y., Qin, F., Li, X., 2018. Downscaling satellite-derived daily precipitation products with an integrated framework. Int. J. Climatol. 1–18. https://doi.org/10.1002/joc.5879.

Chen, S., Zhang, L., She, D., Chen, J., 2019. Spatial Downscaling of Tropical Rainfall Measuring Mission (TRMM) and Multi-satellite Precipitation Data over Complex Terrain. J. Remote Sens. 29, 4097–4110. https://doi.org/10.1016/j.rse.2019.03.002.

Dinku, T., Ceccato, P., Connor, S.J., Ropelewski, C.F., 2008. Validation of high-resolution satellite rainfall products over complex terrain. Int. J. Remote Sens. 29, 4097–4110. https://doi.org/10.1016/j.rse.2019.03.002.

Dinku, T., Ceccato, P., Connor, S.J., 2011. Challenges of satellite rainfall estimation over complex terrain. Int. J. Remote Sens. 32, 5965–5979. https://doi.org/10.1080/01431161.2010.519381.

Dinku, T., Hallemann, K., Maidment, R., Tarnavsky, E., Connor, S., 2014. Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. Int. J. Climatol. 34, 2489–2504. https://doi.org/10.1002/joc.3855.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., Ceccato, P., 2018.
