Journal of
Water & Climate Change

 Journal of Water and Climate Change Vol 14 No 6, 1847 doi: 10.2166/wcc.2023.457

Evaluation of a multi-staged bias correction approach on CHIRP and CHIRPS rainfall product: a case study of the Lake Hawassa watershed

Tsegamlak Diriba Beyene [a,b,*], Fasikaw Atanaw Zimale [a,c], Sirak Tekleab Gebrekristos [d] and Dessie Nedaw [a,e]

[a] Africa Center of Excellence for Water Management, Addis Ababa University, Addis Ababa, Ethiopia
[b] School of Hydraulic and Water Resources Engineering, Dilla University, Dilla, Ethiopia
[c] Faculty of Civil and Water Resources Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia
[d] Ethiopian Institute of Water Resources, Addis Ababa University, Addis Ababa, Ethiopia
[e] School of Earth Sciences, Addis Ababa University, Addis Ababa, Ethiopia
*Corresponding author. E-mail: tsegamlak.diriba@aau.edu.et

TDB, 0000-0002-8500-0725; FAZ, 0000-0001-9778-2712; STG, 0000-0002-7934-8506; DN, 0000-0003-1663-078X

ABSTRACT

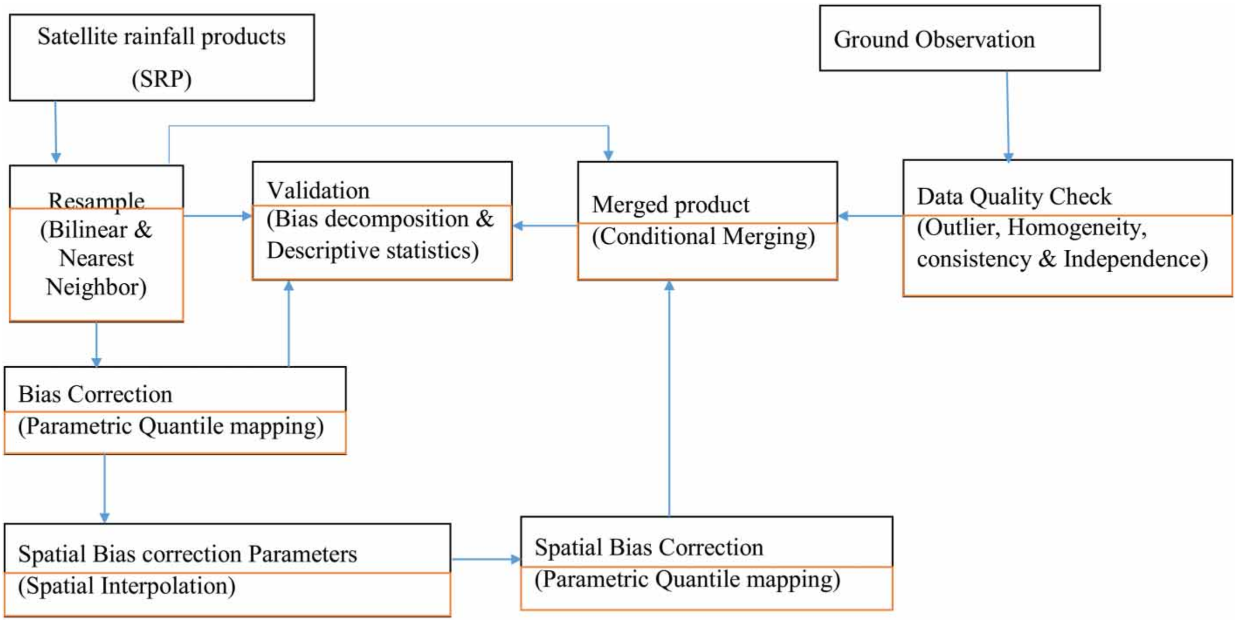
A promising future development area to improve the accuracy of satellite rainfall estimates (SREs) is accessing merits from different sources of data through combining algorithms. The main objective of this study is to assess the accuracy and importance of the fused multistage approach of bias correction. Accordingly, two versions of resampled and spatially bias-corrected Climate Hazards Group Infrared Precipitation (CHIRP) estimates were merged with ground measurements using a conditional merging procedure. Results of applied performance measures (i.e. seven) on corrected and merged CHIRP SREs show that the Percent of Detection (POD) and Percent Volume Error (PVE) have improved. Depending on the combination of coupled stations for validation, up to 70 and 50% PVE improvement was achieved at some stations for wet and dry periods, respectively. Moreover, the bias-corrected and conditionally merged CHIRP SREs have outperformed the estimates by resampling CHIRP with station dataset (CHIRPS) over the sparsely populated western part of the watershed. However, the devised method was limited in considering dry-day events during bias correction, which in turn has affected the performance of the bias correction of the CHIRPS product. Finally, future research should concentrate on such methods of fusing to understand the benefits of various approaches and produce more precise rainfall records.

Key words: bias correction, conditional merging, quantile mapping

HIGHLIGHTS

- The research provides a fused multi-staged approach for reducing errors in CHIRP and CHIRPS satellite rainfall estimates.
- Application of parametric QM for spatial bias correction followed by conditional merging improves the quality of CHIRP SREs.
- Bias-corrected and conditionally merged CHIRP estimate outperforms the estimates by CHIRPS.
- Incorporating additional ground station records improves the estimates of SREs.

GRAPHICAL ABSTRACT



1. INTRODUCTION

There are a large number of satellite precipitation data sources (Ziarh *et al.* 2021), which are used in the assessment of hydrological extremes and processes modeling, including climate change impact studies (Ebrahimi *et al.* 2017). Nevertheless, each source has strengths and gaps in representing the records collected by ground stations (Wang & Zhao 2022). Satellite rainfall products are acclaimed for their wide coverage and cost minimization related to large area coverage compared to *in situ* measurements (Suliman *et al.* 2020; Wang & Zhao 2022). On the other hand, the spatial coverage, accessibility, and density of ground stations are limiting factors when considered for use (Katiraie-Boroujerdy *et al.* 2020).

Systematic errors and random errors are common in satellite rainfall products (Goshime 2020). Error sources are mostly related to imperfection of retrieval algorithm, data source, and postprocessing procedures (Dubovik *et al.* 2021; Zhang *et al.* 2021). Compared to some meteorological variables like temperature, which has a steadier geographical and temporal pattern, bias correction of satellite rainfall data is thought to be the most difficult (Soo *et al.* 2020). Almost all available methods focus on correcting systematic errors while intrinsically trying to correct random errors as well (Dinku *et al.* 2011). One can argue that it is mandatory to correct patterns (i.e. both spatial and temporal) as much as it is important to correct magnitude (Iqbal *et al.* 2022). From available methods distribution mapping (DM) tends to address bias by correlating patterns of different rainfall magnitudes (i.e. by relating cumulative distribution function (CDF) of control data to CDF of satellite rainfall products using transfer function) (Valdés-Pineda *et al.* 2016; Katiraie-Boroujerdy *et al.* 2020; Soo *et al.* 2020).

The imperfection of conventional quantile mapping (QM) is that it was proven to perform poorly on a daily time scale. Other modified versions like censored-shifted Gamma DM have shown improvement over the standard QM technique, but still lack the ability to capture extreme events (low and/or high) (Lafon *et al.* 2013; Ma *et al.* 2018; Ma *et al.* 2019). Significant results were also gained by fitting fused distribution functions with different capabilities (Ma *et al.* 2019), for specific local cases. One alternative that can be applied for DM is the option of non-parametric QM, where CDFs are created only for available magnitudes. However, such methods will be impossible to apply for magnitudes greater than training/calibration data (Lehner *et al.* 2020). Additionally, transferring information from grid points with ground observation to unrepresented grid points is difficult using such methods.

Two well-known approaches to compare satellite products with rain gauge readings are the point to grid and grind to grid (Ebrahimi *et al.* 2017). Ebrahimi *et al.* (2017) assessed the accuracy of Tropical Rainfall Measuring Mission (TRMM) 3B42 v7 rainfall product by initially implementing the nearest neighbor (NN) and weighted bilinear interpolation (WBL) techniques to create co-located pairs of observations, where the study concluded the large errors occur due to spatial mismatch and WBL performed better to reduce such mismatch. A possible future development area to improve the accuracy of satellite rainfall forecasts is accessing good attributes from different sources of data through merging techniques (Kimani *et al.* 2018). Xie & Xiong (2011) discussed how two-stage correction, or first applying statistical bias correction method and then integrating satellite rainfall outputs with ground observations, can significantly enhance performance. On the contrary, application of merging techniques alone was preferred by some scholars (Jongjin *et al.* 2016; Gebremedhin *et al.* 2021) and still managed to decrease errors compared to the raw product.

Accurate depth of point rainfall is collected with a dense network of rain gauges, where continuous calibration and follow up are provided (Xie & Xiong 2011). Utilizing the complementary strengths of each source through the use of a merging process allows for the reduction of bias in satellite rainfall products (Jongjin *et al.* 2016). Merging techniques are valuable in reducing the biases from multiple factors (factors such as rainfall formation type, topography, and others) (Dinku *et al.* 2011; Xie & Xiong 2011). Based on the assumption that rain gauge data over China are bias-free and Climate Prediction Center morphing method (CMORPH) satellite rainfall products are capable of simulating spatial patterns of rainfall, Xie & Xiong (2011) developed a two-stage conceptual model to produce an enhanced rainfall map. The process involved bias correction (using statistical method) followed by a merging technique (i.e. optimal interpolation technique) (Xie & Xiong 2011). The result indicates that the conceptual method has improved the pattern agreement with independent observation and potentially reduced bias in the satellite product (Xie & Xiong 2011).

Another work by Jongjin *et al.* (2016) proved that merging techniques of Conditional Merging (CM), Geographical Differential Analysis and Geographical Ratio Analysis (GRA) are capable of improving the accuracy of satellite-based rainfall products without applying bias correction techniques. The research finding revealed that for a sparse rain gauge network, the CM technique performed better compared to the other two methods (Jongjin *et al.* 2016). Gebremedhin *et al.* (2021) illustrated a promising result gained by implementing the geographical weighting regression method of merging for meso-scale catchments of the Upper Tekeze basin. The method relies on accessing the freely available topographical information (explanatory variable) and incorporating them into the merging procedure (Gebremedhin *et al.* 2021). The main assumption of the method is a correlation of non-stationary in space of the dependent and independent variable in addition to spatial variability of the regression coefficient (Hu *et al.* 2019).

Working with multiple data sources with higher spatial resolution and longer records is a challenge to conduct data merging as well as perform bias correction. Such problems can be alleviated with the use of simple algorithms. Being an interpreted language, Python enables the easy development of algorithms with multiple open-source packages (Grus 2019). One such work was developed by Gupta *et al.* (2019), for climate data bias correction. Another work by Guenzi *et al.* (2017) developed an open source algorithm for the conventional CM technique, to ease the merging process of radar observation with gauge records. However, there are few available open-source algorithms that can support multiple tasks.

Multiple studies (Gebere *et al.* 2015; Gella 2019; Goshime 2020; Adane *et al.* 2021) have been done to assess the accuracy of satellite products in different parts of Ethiopia. These studies have demonstrated that raw satellite rainfall outputs overestimate or underestimate rainfall events compared to rain gauge observations. Whereas, other researchers (Valdés-Pineda *et al.* 2016; Omondi 2017; Gella 2019; Adane *et al.* 2021) investigated the accuracy of bias correction approaches employing two or more bias correction methods. The difficulty in correcting satellite-based precipitation data with ground data, however, is still a significant issue in the deployment of such data for various uses (Ziarh *et al.* 2021). Additionally, given scanty records in ground observation, integrating several sources could be useful in addressing propagated errors.

Therefore, the main objective of this study is to assess the accuracy and importance of the fused multistage approach of bias correction. Specific objectives assigned to answer the main objectives include the following: (1) assess the significance of conventionally applied NN and bilinear (BL) resampling techniques, (2) evaluate simple merging method of CM, (3) evaluate DM-based bias correction method followed by CM technique, and (4) develop simple algorithm of spatial bias correction and CM using python programming language.

2. METHODOLOGY

2.1. Study area

At the center of the Main Ethiopian Rift (MER) lies a hydrologically closed watershed of Lake Hawassa (Figure 1). Sourced from watershed delineation using a Digital Elevation Model (DEM) of resolution 30 m, the watershed covers a 1,473.3 km^2

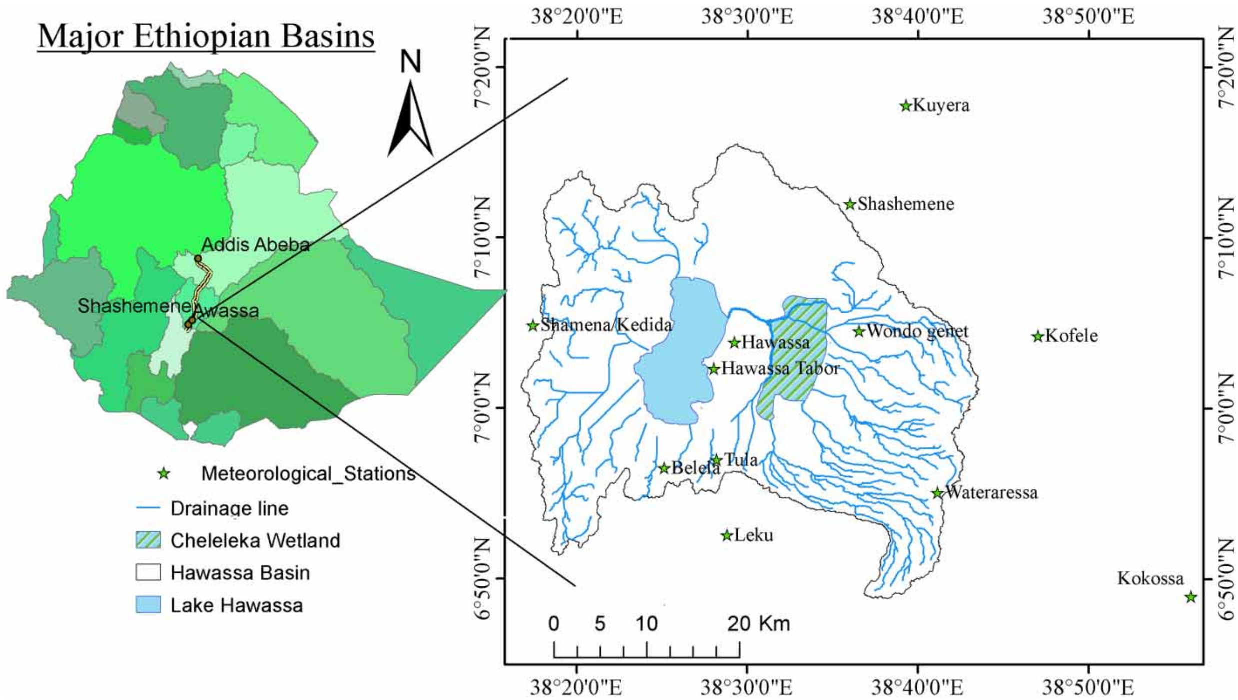


Figure 1 | Location map of the study area.

area. All surface water drains toward the center of the watershed (i.e. Lake Hawassa). The watershed has one perennial river (i.e. known as Tikur Wuha) draining 50% of the watershed area (Belete *et al.* 2017). The rest of the watershed part is drained by flush flood and incised gully formations. Eastern and western parts of the watershed possess 'moist Woina Dega' and 'dry Woina Dega' climate conditions, respectively (Degife *et al.* 2019), with 70% of rainfall occurring from March to September (Legesse *et al.* 2004). Such distinct climate conditions help to assess the impact of spatial rainfall variability and related uncertainty.

2.2. Dataset

The ground observed rainfall data were obtained from the National Ethiopian Meteorological Agency (EMA); whereas the Climate Hazards Group Infrared Precipitation (CHIRP) satellite rainfall product (SRP) and its modified version with ground station data, Climate Hazards Group Infrared Precipitation with station dataset version 2.0 (CHIRPS 2.0), for the time period from January 2008 to August 2018 on a daily time scale, were obtained from https://data.chc.ucsb.edu. However, it should be noted that the ground observation is available for different lengths of time since the establishment of each station; and that the specified length of period was selected to assure maximum overlapping length of observation from available stations with minimum missing data (on average up to 20%). In the study watershed, the formation of snow is less likely to occur, hence the satellite precipitation product is referred to as SRP.

Freely available CHIRPS 2.0 and CHIRP SRP, with a resolution of 0.05° for a time scale of monthly, decadal, pentadal, and daily (Funk *et al.* 2015), were downloaded from https://data.chc.ucsb.edu at a daily time scale. CHIRPS is a quasi-global product dependent on the algorithm which fuses estimates from infrared Cold Cloud Duration (CCD) observations and ground observation (Funk *et al.* 2015).

According to the station location information gathered from the EMA only six stations are found within the watershed. To complement missing data and represent the western part of the watershed as opposed to densely populated station distribution in the eastern, an additional six stations surrounding the watershed were considered. Further analysis of missing data rate and seasonal contribution (Table 1) was conducted to indicate the status of collected data.

The observed data for most stations have fairly minimum missing data (Table 1), which is depicted below. Analysis of the seasonal rainfall patterns was done separately for the dry period (November–March), when only 18% of the rainfall typically

Table 1 | Preliminary assessment of missing data rainfall record and rainfall contribution, in dry (November–March) and wet (April–October) periods

Station	Station code	Elevation (m)	Rainfall percentage		Missing data	
			Wet period (%)	Dry period (%)	Wet period (%)	Dry period (%)
Kuyera	KU	1,936	85.2	14.8	10.5	11.1
Belela	BE	1,866	88.5	11.5	7.9	16.3
Leku	LE	1,879	80.8	19.2	6.5	7.1
Kokossa	KOK	2,400	77.3	22.7	32.8	36.5
Shamena_Kedida	SHK	2,072	82.1	17.9	6.7	11.2
Wondo_genet	WO	1,880	81.6	18.4	7.8	1.7
Tula	TU	1,873	78.1	21.9	4.0	3.5
Hawassa	HA	1,750	83.7	16.3	0.1	3.1
Hawassa_Tabor	HW	1,750	82.8	17.2	5.2	1.9
Wateraressa	WT	2,631	83.3	16.7	16.1	19.8
Shashemene	SHA	1,927	80.7	19.3	17.0	23.7
Kofele	KOF	2,620	78.1	21.9	5.4	6.7

falls, and the wet period (April–October), when 82% of the rainfall typically falls. The observation from Shashemene, Wateraresa, and Kokossa stations have significant missing data, especially for dry periods. This connotes the importance of an efficient method to fill the missing observation.

Further data quality assessment was conducted on each station using outlier Grubbs and Beck (1972) test (G–B)), homogeneity and stationary Mann–Whitney (1947), independence & stationary test Wald-Wolfowitz (1943), and consistency test (double mass curve). The check for consistency indicated that the Wateraresa station had significant deviation. On the other hand, the consistency check for Hawassa and Shashemene has a slight deviation; hence, correction was applied to each station. Other quality tests were conducted for sample data (i.e. daily maximum, and annual rainfall), from each station. The test for homogeneity and stationarity indicated that data from all stations are homogeneous. Similarly, the data were independent for all stations. Furthermore, whenever required the missing data were filled using the arithmetic mean method. To enhance the accuracy of filling a total of 28 stations (Figure 2) were used including those used for this study.

2.3. General method

Ground measurements of rainfall are usually susceptible to errors, which creates additional burden on the modeling efforts. Furthermore, the ground observations are usually set in accessible areas (i.e. close to urban areas, and road side), which does not represent the distribution in remote and inaccessible areas (Dinku *et al.* 2014; Dile *et al.* 2018). This is where the remote sensing data will be important. Unfortunately, these data types also have errors. Therefore, we need separate data for bias correction and validation of satellite rainfall products (Gebremedhin *et al.* 2021). The ground observations are thoroughly assessed for any temporal or spatial inconsistency.

The satellite data were bias-corrected (BC) and spatially disaggregated. Then, the available ground rainfall records are used for validation by dropping some stations at a time and re-iterating each process. All stations in the watershed including those surrounding the watershed are used. Bias correction was performed using a parametric empirical QM procedure. While downscaling was performed for 1 km resolution by applying NN and bilinear (BL) interpolation methods. Different methods have been introduced to merge satellite rainfall products with *in situ* observation. These include geographically weighted regression (GWR), CM, geographic differential analysis (GDA), and GRA.

Based on the objective of the study, CM was implemented in this study. When applying the CM merging technique, both satellite data and *in situ* measurements at intersecting available grids were interpolated using the Ordinary Kriging (OK) method, then the residual was added to the original satellite rainfall map to create the merged product (Jongjin *et al.* 2016). In general, the study devised to apply a two-stage method (Figure 3); the first stage consists of resampling the original

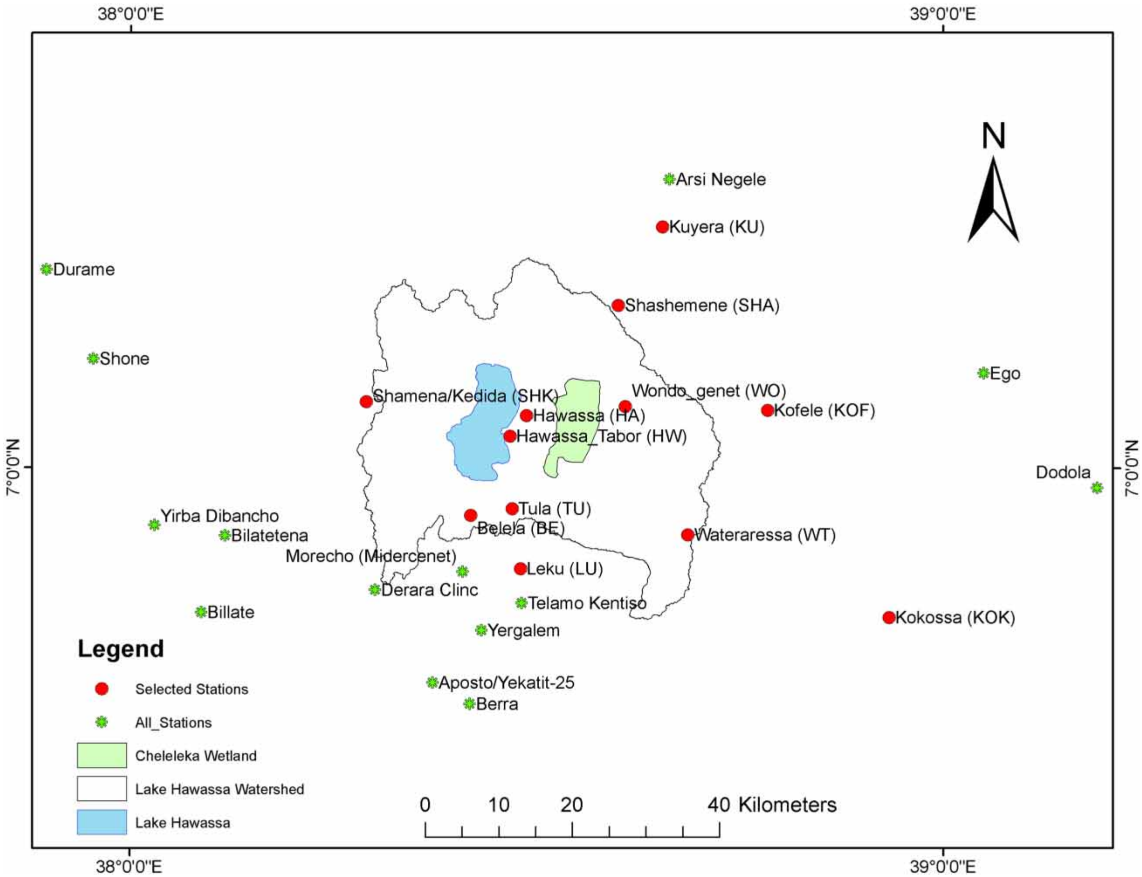


Figure 2 | All rain gauge stations accessed for the study.

SRP product to a 1 km resolution followed by bias correction, while the second stage uses interpolated bias correction parameters at available grid locations to correct all grids and then merge with ground measurement using a CM procedure.

One of the biggest challenges working with large data, both spatially and temporally, is the availability of comprehensive tools. There are multiple tools like Arc GIS, statistical softwares, and others to execute individual and/or multiple processes mentioned below (Figure 3). However, it was a challenge to find one tool that can execute all the tasks with minimum computation time and storage requirement. Hence, for this study simple algorithms dependent on existing different modules written in Python language were developed. Some of the open access libraries applied in these process include; PYKRIGES: a Kriging Toolkit for interpolation tasks (Murphy 2014); Python Data Analysis (PANDAS) Library: a data analysis and manipulation including statistical analysis (McKinney 2011); Geospatial Data Abstraction Library (GDAL/OGR), a translator library for raster and vector geospatial data formats (GDAL/OGR Contributors 2020); Numerical Python (NUMPY): for accessing, manipulating and operating on data in vectors, and arrays (Harris *et al.* 2020); RASTERIO: for reading writing raster files (Gillies 2019); XLRD: a library for reading data and formatting information from Excel files; SHAPELY: manipulation and analysis of planar geometric objects (Gillies 2013); and Scientific Python (SCIPY): a library for optimization, mathematical calculations, statistics and others (Jones *et al.* 2001).

2.4. Bias correction and CM algorithm

The developed algorithm is composed of four steps. The main stages of the program include preprocessing of input data, at station bias correction and fitting distribution parameters, spatial bias correction, and CM. The main inputs to the program are raw daily satellite rainfall estimates (SREs) in '.tif' file format and labeled with the date of the observation, and the daily gauge record prepared in '.csv' file format; furthermore the length of the gauge record should be equal to the length of SRP. Additionally, a simple Graphical User Interface (GUI) is prepared using the PyQt widget toolkit (Riverbank Computing 2016) and includes four separate tabs for each stage.

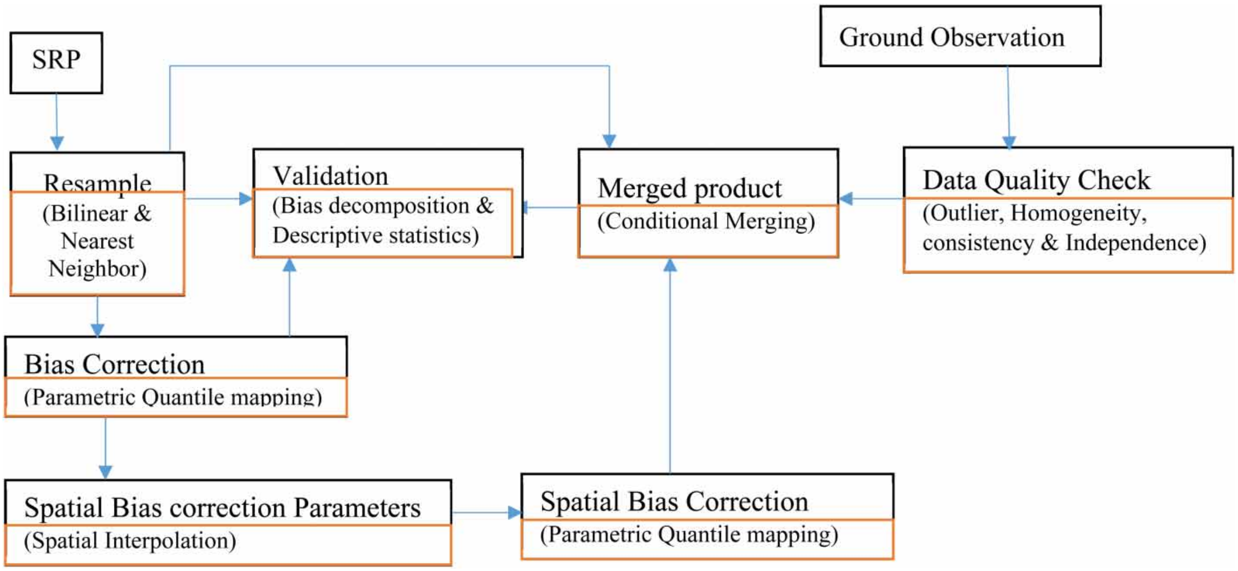


Figure 3 | Flow chart of the general method; method implemented in each stage is stated in brackets.

- **Input data preprocess:** at this stage the program helps to prepare the SRP. This includes projection, creating boundary shape file, clipping, and resampling. Since most of the procedures require interpolation the spatial reference system should be projected to metric systems, such as Transverse Mercator (UTM). Whereas the clipping and boundary file creation are required to reduce the number of grid points to be processed.
- **Bias correction and distribution fitting:** the pre-processed raster SRP files and ground observation data in '.csv' file format are used as input for this stage, where a distribution is fitted to the monthly non-zero events at collocated rain gauge locations. There are two widely used distribution options, Gamma and Exponential, where it is possible to choose both or either one. The gauge record is fitted to the selected distribution by excluding missing and non-zero events for all months separately. At this stage, the program is equipped to run bias correction (Equations (1)–(4)) either including performance measures at each rain gauge location or without performance measures (Equations (6)–(13)). The output at this stage includes fitted distribution parameters (i.e. to rain gauge records) on monthly bases, performance measure results, and BC rainfall data at all gauge locations.
- **All Grid Bias Correction:** Using the boundary extent, fitted distribution parameters are interpolated by applying the OK technique for all months separately. The mapped/gridded parameter is then used to conduct the bias correction at all grid points for the projected raster SREs. A K-fold validation process is also optional at this stage. Hence, working extent, resolution, number of validation stations at each iteration, and projection method need to be specified in this stage. Otherwise, one can also choose to only perform spatial bias correction without the performance measures. The outputs include corrected raster file, and performance indicators in '.csv' file of each iteration.
- **CM:** The corrected raster file and rain gauge records are inputs in this stage. With the specified procedure (i.e. in the following section) the CM is conducted using the inputs. As a cross validation mechanism, the program allows us to randomly drop a specified number of stations and conduct the merging process, while performing validation for the hidden stations. The process can be repeated for a specified number of iterations. For practical application, the program also allows the CM to be conducted without performance check. Based on the choice of simulation the output includes merged raster files and performance measures for each iteration.

2.5. Resampling

Recent studies recommended the use of information gathered from multiple sensors and/or satellites to increase the chance of retrieving the best representation from different sources (Dinku *et al.* 2011; Dubovik *et al.* 2021). However, incompatible scale and resolution of different satellite observations is a major challenge in mostly used retrieval algorithms (Omondi 2017). Based on the method applied, usually resampled raster data suffer from the inability of maintaining the originally stored information (Omondi 2017). Hence, one should expect that different methods might have different error propagation patterns.

One of the widely utilized methods of resampling is the NN method. This method is advantageous for its less computational time requirement, ease of application, and ability to maintain original stored information (Brandsma & Können 2006). However, Omondi (2017) indicated that this method is the least accurate for it distorts results and sometimes omits or duplicates multiple cell values. With their ability to smoothen output raster and for up-sampling, BL resampling methods are used for averaging values in the nearest four grids (Baboo & Devi 2010).

Rain gauge observations are usually perceived as a true value of rainfall and used for correcting satellite rainfall products. On the contrary, one can argue that most ground observations are full of errors, especially using manual rain gauges in developing countries (Beyene *et al.* 2018; Dile *et al.* 2018). Aside from these modest assumptions spatial mismatch (i.e. between rain gauge point observation and pixel value of SRP) is another eminent source of error, especially for low spatial resolution (large pixel size) products (Dinku *et al.* 2011; Gebere *et al.* 2015). As a solution to this challenge, many studies (Jongjin *et al.* 2016; Gebremedhin *et al.* 2021) have depended on interpolating ground observations to create matching gridded rainfall maps. Readers are recommended to review Hu *et al.* (2019) and Li & Heap (2008), for an elaborate discussion on available methods of interpolation including the pros and cons of the methods. Conducting the merging process requires creating a similar resolution between SRP and interpolated ground station measurements. Hence, both NN and BL methods of resampling were tested for performance and use in this study.

2.6. Bias correction

Bias correction based on the sampling window and length of sampling window is a well-practiced technique to improve bias correction (Bhatti *et al.* 2016). Aiming to decrease the percentage of wet spill of CHIRP SRE, this study has considered a 3- and 1-day central moving window for the dry and wet periods, respectively. After extracting the estimated rainfall at specified grid point (j) using pixel to point approach, the algorithm calculates the air distance from the center of the grid to all available rain gauge stations' locations. Thereafter, assuming rain gauge density requirements of 1 in 100 km^2, if the minimum distance to any of the stations falls below 10 km that station (x) is selected for comparison (Equation (1)). On the other hand, if the distance exceeds 10 km the closest four stations are selected for comparison.

$$\left\{\begin{matrix} P^i_{sj} = 0 & if & P^i_{gx} = 0 \\ P^i_{sj} = P^i_{sj} & if & P^i_{gx} > 0 \end{matrix}\right\} \tag{1}$$

where $P^i_{gx} = [P^i_{g1}, P^i_{g2}, \ldots, P^i_{gn}]$, for a given day (i), if any of the rain gauge stations (x) record rainfall (P^i_{gx}) the estimated SRE (P^i_{sj}) will be kept, if no rainfall is detected by any of the selected stations the estimated SRE is replaced by zero. The resulting time series of rainfall are used for bias correction. However, this procedure was not performed on the CHIRPS product, realizing that the estimates are already corrected using a few ground rain gauge records.

DM techniques were originally developed to bias correct regional and global climate models (McGinnis *et al.* 2015; Switanek *et al.* 2017). One of these DM techniques, QM, has been proven to perform in correcting both regional and global climate circulation modes (Teutschbein & Seibert 2012; Heo *et al.* 2019). Additionally, QM techniques are chosen to bias correct satellite rainfall products for their ability to adjust the standard deviation of daily satellite rainfall products while preserving other moments as well (Katiraie-Boroujerdy *et al.* 2020).

QM based on a parametric empirical DM (Equations (2)–(4)), a method of bias correction, was applied in this study. As presented by Themeßl *et al.* (2012), corrected rainfall ($R^{cor}_{t,i}$) at grid (i) and day (t) is given as:

$$R^{cor}_{t,i} = R^{raw}_{t,i} + Cf_{t,i} \tag{2}$$

$$Cf_{t,i} = ecdf^{obs,cal^{-1}}_{m,i}(P_{t,i}) - ecdf^{mod,cal^{-1}}_{m,i}(P_{t,i}) \tag{3}$$

$$P_{t,i} = ecdf^{mod,cal}_{m,i}(R^{raw}_{t,i}) \tag{4}$$

where $Cf_{t,i}$ is the transfer function; $ecdf^{obs,cal^{-1}}_{m,i}$ and $ecdf^{mod,cal^{-1}}_{m,i}$: are observed ($obs$) and modelled ($mod$) inverse empirical month (m), respectively; whereas P: is probability at the calibration (cal) period.

For a given station and month the ground observation (without missing data) and resampled SRP at coinciding grid point was fitted to an exponential distribution, using the Maximum Likelihood method. Additionally, the distribution was fitted to non-zero events and excluded missing data. Furthermore, a standard check of the χ^2 test and Kolmogorov–Smirnov test was conducted to confirm a good fit of the distribution to each dataset. Once the distribution parameters at collocated grid points were determined, parameters were transferred to grid points with no ground observation through interpolation (OK). Hence, the bias correction was conducted for all grid points with and without ground control observation. This would not have been possible with the non-parametric QM technique. Additionally, parametric QM methods allow us to correct maximum events (not recorded by ground stations) that may require extrapolation while transferring distribution parameters.

2.7. Conditional merging

Implementing a merging technique using two sources enables the capturing of complementary merits of each source, which could reduce bias in satellite rainfall products (Jongjin *et al.* 2016). Based on OK interpolation, CM is a conventionally used merging technique, which is used to merge radar data and meteorological data (Guenzi *et al.* 2017). OK is a regression technique that allows for value interpolation while minimizing mean squared error (Guenzi *et al.* 2017). While performing CM, it is anticipated that a theoretical semivariogram based on datasets of regularly distributed precipitation is fitted to the experimental semivariogram for the operation of OK (Jongjin *et al.* 2016).

The experimental semivariogram was calculated, and it was then fitted to the theoretical semivariogram using weighted least squares to determine the appropriate variogram parameters like nugget, sill, and range (Jongjin *et al.* 2016). In this study, CM was applied following the procedure applied by Jongjin *et al.* (2016). The initial step was the interpolation of the ground station record to a 1 km resolution field using OK; the second step requires extracting the collocated (i.e. with ground observation stations) resampled (1 km resolution) SRP grid cell observations; thirdly, the extracted SRP grid cells will be interpolated to 1 km resolution field using OK; in the fourth step error is calculated by deducting rainfall map created in the third step from resampled SRP observation; lastly this error field is added to the rainfall map created in the first step.

The merging procedure was conducted in two stages. The first stage was before applying the bias correction technique to the original SRP dataset. Next, it was applied after applying the bias correction described in the previous stage. Both products were tested for performance, in which the necessity of the procedure devised in this study is signified.

2.8. Performance evaluation methods

The validation was conducted at several stages, which enables it to signify the importance of each process (Figure 3). The objective function measures the goodness of fit between the computed and observed at a selected element. The choice of the objective function depends upon the need. Both bias decomposition and statistical methods were applied for two sets of data periods (i.e. wet period and dry period). The following statistical measures were used to quantify the performance accuracy of bias correction outputs. These are Percent Error in volume (PVE), Coefficients of Determination (R^2), and Root Mean Squared Error (RMSE) which are widely applicable in hydrologic modeling.

R^2 describe (Equation (5)) the degree of co-linearity between BC and gauged data (Moriasi *et al.* 2007).

$$r^2 = \frac{\left(\sum (P_{gi} - \overline{P_g}) \ \sum (P_{si} - \overline{P_s})\right)^2}{\sum (P_{gi} - \overline{P_g})^2 \sum (P_{si} - \overline{P_s})^2} \tag{5}$$

where R^2 is the coefficient of determination (where values can range from -1 to $+1$), P_{gi} is the gauged rainfall value at the ith time interval, P_{si} is SREs at collocated grid at the ith time interval, $\overline{P_g}$ is the mean of gauged rainfall and $\overline{P_s}$ is the mean of SREs at collocated grid. Perfect positive correlation of rainfall estimates is indicated with a value of 1, while a 0 result implies no correlation.

PVE ranges from –ve ∞ to +ve ∞, where results within -10 to $+10\%$ indicate good agreement and results out of the indicted range show poor performance, and 0 is the perfect match. PVE (Equation (6)) considers computed volume, and does not account for the magnitude of the peak:

$$\mathrm{PEV} = 100\left|\frac{V_{\mathrm{o}} - V_{\mathrm{s}}}{V_{\mathrm{o}}}\right| \tag{6}$$

where PEV refers to the percentage error in volume, V_{o} refers to the observed volume and V_{s} is the simulated volume.

Error variation between raw and BC products against ground observation can be indicated by RMSE (i.e. ranging from 0 to ∞), where close to zero results indicated very good performance and vice versa. The mathematical representation of RMSE (Equation (7)) is

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{si} - P_{gi})^2} \tag{7}$$

where P_{si} & P_{gi} are SRP and ground-based rainfall observations at day, i; whereas n is the total number of available pairs of data. Performance checks on each output were conducted by using descriptive statistics and bias decomposition. Classification of error components in satellite rainfall products include Hit Bias (HB), Miss Bias (MB), and False Alarm/precipitation (FA). HB is the sum of the difference for over-/under-estimation of detected ground observation by satellites, whereas MB is the sum of all undetected gauge events by satellite and FA is a number of all events estimated by satellite but not by gauging station (Omondi 2017).

For a better description of results from satellite observations compared to ground observations bias decomposition methods such as, Hit Bias (HB), Miss Bias (MB), Percent of Detection (POD), False Alarm Ratio (FAR), and Frequency Bias Index (FBI) were applied.

$$\mathrm{HB} = \sum_{i=1}^{n}(P_{si} - P_{gi}):(P_{si} > 0\ \&\ P_{gi} > 0) \tag{8}$$

$$\mathrm{MB} = -\sum_{i=1}^{n}P_{gi}:(P_{si} = 0\ \&\ P_{gi} > 0) \tag{9}$$

MB and HB measure the total depth of missed and correctly detected rainfall events. For these measures results close to zero indicate very good agreement, and the result (i.e. HB) can range from –ve ∞ to +ve ∞, whereas MB values range from –ve ∞ to 0.

$$\mathrm{POD} = \frac{H}{H + M} \tag{10}$$

$$\mathrm{FAR} = \frac{\mathrm{FA}}{H + \mathrm{FA}} \tag{11}$$

$$\mathrm{FBI} = \frac{H + \mathrm{FA}}{H + M} \tag{12}$$

These indicators depend on the number of SREs, successfully detected events (H), missing SRE events that are measured by rain gauges (M), and rainfall detected by SRE but not measured by a rain gauge (FA). The POD calculates the ratio of observed rainfall rates that are correctly detected by SREs (H), whereas the FAR calculates the ratio of wrongly detected rainfall rates (FA) by SREs. The ideal scores for the POD and FAR indices, which have a range of 0 to 1, are 1 and 0, respectively. The FBI calculates the ratio of the total number of rainfall rates observed to the total number of rainfall rates detected by the SREs, with a value range of 0 to ∞. Greater than 1 values of FBI suggest overestimation by SREs, whereas less than 1 values of FBI imply underestimation, with 1 being the perfect value.

3. RESULTS AND DISCUSSION

The raw CHIRP and CHIRPS products underwent analysis, and the results showed that the older/prior version had a higher percentage of wet spills or erroneous detections. As a result, the CHIRP version's detection percentage is higher. CHIRP rainfall estimates have greater than 50% of FAR.

As an illustration, the result gained (Figure 4) at the Hawassa rain gauge station (i.e. using point to pixel comparison) shows that 95% of the rainfall recorded by the ground station was detected by the CHIRP, whereas more than 50% of the wet spills were false alarms. However, looking at the false bias index value the deviation from the ground station is 31 and 100% for CHIRPS and CHIRP, respectively. However, due to the index's inability to assess correctly detected

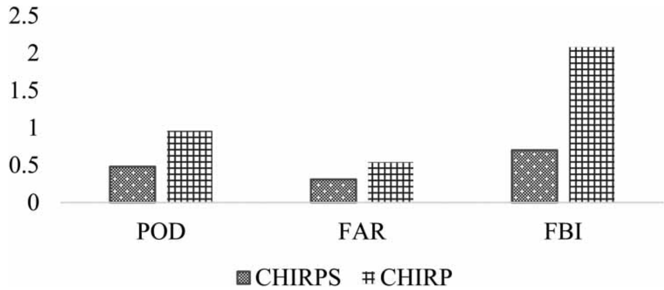


Figure 4 | Comparison of raw CHIRP and CHIRPS dataset at the Hawassa gauging station.

SREs that are not measured by ground stations, one cannot draw any firm conclusions about the performance based only on FAR.

3.1. Resampling

The results from the applied spatial disaggregation methods (BL and NN) were compared to corresponding available grid points with ground observations. Examining the NN resampling/downscaling method for the dry period (Table 2), where the raw SREs are preserved, the percent of CHIRPS (V2) detection falls below 30% for most locations. Looking at this in line with FAR as high as 69% (Supplementary material, Table S1) indicates that the performance is poor. Similar conclusions were made by Goshime (2020), where the false detections are more than 50%. However, the descriptive statistical results show that very good agreement was shown at most stations, especially at Leku, Hawassa, and Hawassa_Tabor stations.

On the other hand, CHIRP (V0) performance exhibited very good, i.e. referencing RMSE (Supplementary material, Table S1) and POD (Table 2). Nevertheless, as illustrated in the above section this is due to the higher rate of wet spill in V0 SREs compared to the V2 SREs. Very high FAR results are recorded in V0 in comparison to V2, which is due to the fact that V2 is sourced from ground observation and SRP (Funk *et al.* 2015). Negative percentage error in volume indicates over estimation, while near zero results imply very good agreements. The overestimation in volume reaches as high as 89% (at Belela station) using the NN resampling technique, whereas the BL method reduced the over estimation by 2% (Table 3).

Comparing the two resampling methods, slight improvement in performance (i.e. descriptive statistics) was shown by BL SREs. However, examining the bias decomposition result the NN showed slight performance improvement. This is consistent

Table 2 | Performance evaluation of NN resampling (i.e. to 1 km resolution) techniques for dry and wet periods, for CHIRP (V0) and CHIRPS (V2)

	Wet period				Dry period			
	PEV (%)		POD		PEV (%)		POD	
Station	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	−21.1	−19.7	0.98	0.45	−89.1	−60.3	0.77	0.29
HA	−3.38	1.81	0.98	0.41	−16.8	9.49	0.86	0.3
HT	−15.3	−9.8	0.98	0.44	−28.5	−5.9	0.85	0.31
KOF	−7.8	−5.3	0.99	0.38	−0.26	14.3	0.9	0.26
KOK	35.6	38.5	0.99	0.33	50.7	57.3	0.83	0.21
KU	−21	−19.3	0.93	0.36	−56.2	−18.7	0.89	0.28
LE	−25.9	−24.2	0.98	0.46	−13.2	3.02	0.83	0.35
SHK	−42.7	−37.1	0.98	0.4	−50.9	− **17.9**	0.81	0.29
SHA	−97.2	−87.2	0.98	0.37	−95	− **46.2**	0.83	0.25
TU	−20.3	−27.7	0.97	0.44	3.12	11.2	0.82	0.29
WT	−16.1	−13.4	0.97	0.38	−39.9	−21.2	0.86	0.23
WO	1	4.9	0.98	0.41	−4.23	17.7	0.86	0.26

Table 3 | Performance evaluation of BL resampling (i.e. to 1 km resolution) techniques for the wet period, for CHIRP (V0) and CHIRPS (V2)

	Wet period				Dry period			
	PEV (%)		POD		PEV (%)		POD	
Station	Vo	V2	Vo	V2	Vo	V2	Vo	V2
BE	−18.5	−17.4	0.98	0.51	−87.3	−58.8	0.77	0.31
HA	−5.94	−0.19	0.98	0.51	−19.1	6.54	0.86	0.4
HT	−12.7	−8.38	0.98	0.51	−25.4	−4.2	0.84	0.37
KOF	−8.01	−5.29	0.99	0.46	0.49	15.01	0.9	0.35
KOK	35.6	37.48	0.99	0.41	50.76	57.3	0.83	0.26
KU	−20.4	−17	0.93	0.46	−56.1	−10.9	0.89	0.33
LE	−25.2	−23.8	0.98	0.51	−13.4	2.88	0.83	0.38
SHK	−41.6	−35.9	0.98	0.47	−50.6	−18.8	0.81	0.38
SHA	−95.1	−91.7	0.99	0.44	−95.7	−54.5	0.88	0.31
TU	−24	−23.7	0.97	0.49	0.54	14.08	0.82	0.34
WT	−16.8	−13.8	0.97	0.47	−40	−21.5	0.9	0.3
WO	1.56	5.1	0.99	0.5	−2.72	18.54	0.86	0.31

with the core principle of the resampling techniques where new and undetected values are assigned by the BL technique from neighboring grids (Omondi 2017). Whereas the NN method preserves/does not introduce new observations at focus grids. For the study period (i.e. 11 years) the total depth of missed and correctly estimated SRP at the collocated ground station is determined by MB and HB (Supplementary material, Tables S3 and S4). Higher positive HB value (wet period of Shashemene, Kofele, and Leku station) indicates the total overestimated intensity of rainfall for correctly detected events by the CHIRPS, while the negative value (wet period of Belela, Kokossa, and Wondo_genet stations) implies underestimated intensities by CHIRP product. Whereas higher negative values of MB in V2 estimates compared to the V0 SRP indicates that the CHIRP product has fewer missed events compared to the CHIRPS product.

At this stage of the procedure, it is sufficient to check the percent of volume error (PVE) and RMSE in order to signify the importance of resampling. Overall, the assessment using descriptive statistics (Table 3) illustrates that BL interpolation is a more efficient method of resampling. Nevertheless, looking at the PVE performance we can see that in some stations (i.e. bolded in Table 2) NN method has shown to outperform the BL method. The assessment of bias decomposition methods (Supplementary material, Tables S3 and S4) illustrates that the BL interpolation method is a more efficient method of spatial disaggregation. In conclusion, the performance before and after resampling did not show exaggerated change. This is accredited to the fact that the CHIRPS algorithm also utilizes the information recorded at ground stations. Nevertheless, Gebremedhin *et al.* (2021) referenced that CHIRPS products can be corrected using ground station records not used in the original product. Refer to Supplementary Information for details on the results of bias decomposition performance measures.

The comparison of the resampling output clearly shows that the BL resampling technique has better performance on the CHIRPS product. The BL resampling technique has proven to perform better in upper Takeze watersheds as well (Gebremedhin *et al.* 2021). Total depth of missed events in wet period is greater than dry period depth for CHIRPS product, while the reverse is true for the CHIRP product.

Resampling procedures are important in reducing the spatial miss match between grid observation and point observation of ground stations. However, resampling methods need to be selected properly by taking into consideration the pros and cons of available methods in line with the objective of the study. In this particular study, the BL resampling has performed better than the NN resampling technique, which is similar to the study conducted by Gebremedhin *et al.* (2021).

3.2. Bias correction

The bias correction technique utilized in this study was parametric empirical QM, which requires determining best known distribution to fit gauge data and SRP. For similar SRP and control observation (gauge observation), the CDF might vary depending on temporal resolution, geographical location, season, rainfall formation, and other factors (Ma *et al.* 2019). Accordingly, one should consider the applicability of known distribution functions for the selected co-located data pairs (Soo *et al.* 2020), meaning different functions for different datasets. Hence, acknowledging the variability of rainfall temporally, distributions were fitted on monthly bases.

Accordingly, distribution good fit tests, and Kolmogorov–Smirnov showed that exponential distribution was a good fit both for resampled CHIRPS product and ground observation. The Kolmogorov–Smirnov test was assessed against critical value at a 5% significance level. The bias correction (Equations (1)–(3)) at available grid points was also tested for performance.

Compared to the result acquired after the resampling stage (Supplementary material, Table S3) some stations (Shahemene, Kokossa, Belela, Hawassa, and Wondo_genet) have exhibited poor performance, while the overall result suggests better performance and most definitely improved the result compared to the raw CHIRPS dataset. Looking at the results of the RMSE performance test (Figure 5) we can clearly observe that the performance has improved for all stations except Belela, Hawassa, and Wondo_genet stations.

The limitation of this process (i.e. specific to this study) is that no bias correction was done to correct no rainfall event missmach. Hence, one can clearly see that the bias correction has resulted in poor performance of HB and Miss Bias in comparison to the resampled product. With the aim to correct grid points with no ground observations, bias correction using interpolated distribution parameters was tested using three hidden stations (Shashemene, Wateraressa, and Tula stations); while the result indicates (Table 4) that the pattern has been captured well, but the volume was underestimated except for dry period at Tula station.

Moreover, the comparison against bias correction using parameters generated by using the ground station (Figure 6) indicates that information can be transferred to unrepresented grid points. Hence, future works should consider transferring bias correction parameters to ungauged grid points rather than using the Thiessen polygon to represent grid points without ground observation.

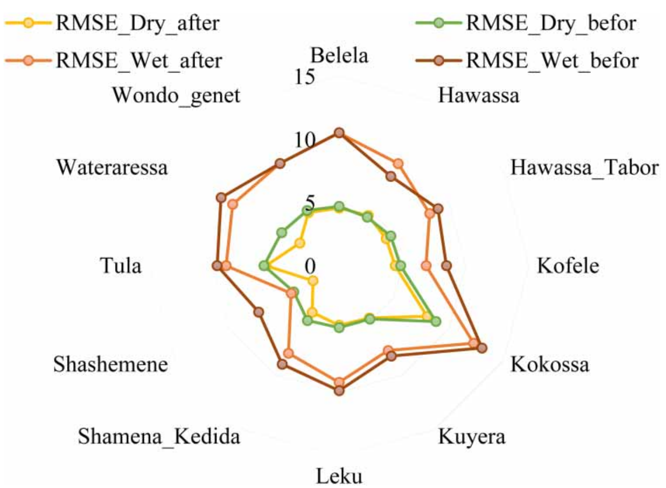


Figure 5 | CHIRPS RMSE test result for dry and wet periods separately; before indicates the result after resampling and after indicates the result after bias correction.

Table 4 | Performance result of CHIRPS spatial bias correction using interpolated distribution parameters

	RMSE(mm)		PVE (%)		HB (mm)		MB (mm)	
Station	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period	Dry period	Wet period
ShA	2.87	6.59	−75.09	−111.24	−863.50	−5,010.20	704.30	2,551.90
TU	4.02	7.80	5.57	−37.39	144.60	−3,278.70	1,717.60	4,486.20
WT	3.53	8.24	−30.11	−21.17	−599.70	−1,966.50	1,439.90	5,415.50

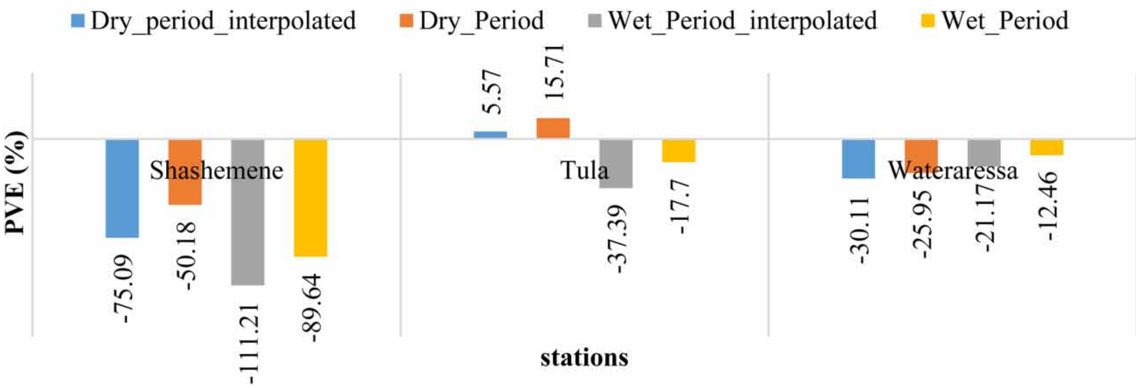


Figure 6 | Performance test for CHIRPS using interpolated distribution parameters.

Negative results (i.e. except RMSE) indicate that the estimated value has underestimated the ground observation at that point, while the positive results indicate over estimation. From the result (Supplementary material, Figure S1) it has been shown that most results suggest that pattern information can be transferred from grid points with ground observation to neighboring grid points with no ground observation. However, the result (except RMSE) at Tula station is not in agreement with the result of the other two stations. These might be due to the limitation of the study, that no rainfall events were corrected.

Two conventionally used distributions were tested for bias correcting the resampled CHIRP product. The result indicates that the exponential distribution (Table 5) best represents the wet period, while the dry period is well captured by a gamma distribution (Table 5). The result has also signified that false satellite rainfall detections can be eliminated or reduced by implementing moving window sampling techniques. The positive near zero results of FBI and PVE (Table 5) show very good agreement and a slight underestimation. However, compared to the NN performance (Table 2), the dry period is not well captured by the exponential distribution.

On the contrary, comparing the HB and MB (Supplementary material, Tables S6 and S7) implies that the missed events have not been changed, rather the correctly captured events have increased in intensity. As an illustration, the HB at Shamena_Kedida station has increased by 388.2 mm, while that Miss Bias only increased by 62.9 mm. The Gamma distribution estimate shows mostly underestimation of events except for Kuyera and Shamena_Kedida stations in the dry period.

Table 5 | At station bias correction result performance of exponential and Gamma distribution

	Exponential distribution				Exponential distribution			
	PVE (%)		POD		PVE (%)		POD	
Station	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
BE	−42.76	3.03	0.77	0.98	0.01	11.94	0.56	0.89
HA	−47.69	3.16	0.86	0.98	2.48	11.92	0.63	0.89
HT	−39.39	0.91	0.85	0.98	1.72	11.73	0.62	0.88
KOF	−30.33	0.14	0.9	0.99	10.92	11.96	0.67	0.89
KOK	−18.53	0.83	0.83	0.99	30.75	13.85	0.54	0.88
KU	−61.87	0.14	0.85	0.99	−30.65	10.34	0.71	0.9
LE	−26.66	1.14	0.83	0.98	10.12	10.98	0.62	0.87
SHK	−61.55	0.13	0.86	0.99	−16.45	12.99	0.6	0.87
SHA	−40.96	7.02	0.83	0.94	14.42	17.54	0.54	0.84
TU	−33.89	0.98	0.82	0.97	7.48	10.56	0.57	0.89
WT	−33.5	2.87	0.86	0.97	7.33	14.27	0.57	0.86
WO	−56.04	1.92	0.85	0.98	−1.95	12.37	0.58	0.88

The RMSE performance measure has improved slightly for bias corrections by Gamma distribution over the dry period, while wet period performance was improved using the exponential distribution (for details refer to Supplementary Information). Hence combining these distributions, i.e. applying gamma distribution for dry period and exponential distribution for wet period provides enhanced BC SREs.

Existing bias correction approaches are commented on for their inability to correct rainfall events (Lehner *et al.* 2020). Similarly, the underperformance was achieved by applying the parametric QM method without considering non rainfall events. However, significant estimation improvement was achieved at some stations and comparing the dry versus wet season, dry period performance was better. Furthermore, a potential method of spatial bias correction for grid points where no *in situ* records are available was tested. Encouraging results were achieved through this process, which will signify the importance of transferring bias correction parameters to grid points where ground measurements are not available.

3.3. Merged products

The study has assessed two merged products, which helps to understand the advantages of each stage in the devised procedure. The first product was a merged rainfall product gained by using available ground observations and resampled product. The validation procedure includes cross validation with 20 iterations, where in each iteration random two stations were dropped, and the merging was done using the remaining 10 stations. After merging the performance test was done at the hidden/dropped stations. The main reason to select two stations at a time was that it allows the remaining training data to be sufficient and to avoid completely unrepresented or distant observation to conduct interpolation.

The comparison was done against the test values of BL resampling product. Considering the RMSE performance test (Figure 7) on average 20% improvement in RMSE value was achieved. All other tests (not presented here) have shown significant improvement. This suggests that simple CM procedures are capable of reducing bias even for scanty ground records. Studies by Pignone *et al.* (2015) presented that this method can be effective in capturing the rain gauge records and correcting the SREs.

The output of merging using BC rainfall map and ground observation was tested using a similar procedure of cross validation over 20 iterations. The result indicates that all iterations show significantly improved performance test values. Comparing the dry period to wet period, the performance of the dry period is much improved. However, this does not entirely indicate the capability of the procedure to correct the dry period event better than the events in the wet period. Part of this discrepancy could be a result of the number of wet events being higher in wet period, while higher no rainfall events in dry period and vice versa.

Additionally, the study has highlighted the significance of combining two or more observational sources to improve rainfall spatial estimates. Even though the CHIRPS product is well-known for its retrieval algorithm to consider ground rainfall measurements, significant performance improvement was achieved by combining the raw product with ground measurement. Similar deductions were made by different studies (Navas *et al.* 2019; Gebremedhin *et al.* 2021). Furthermore, the performance of the simple CM technique has outperformed the devised fused bias correction method, where the parametric empirical

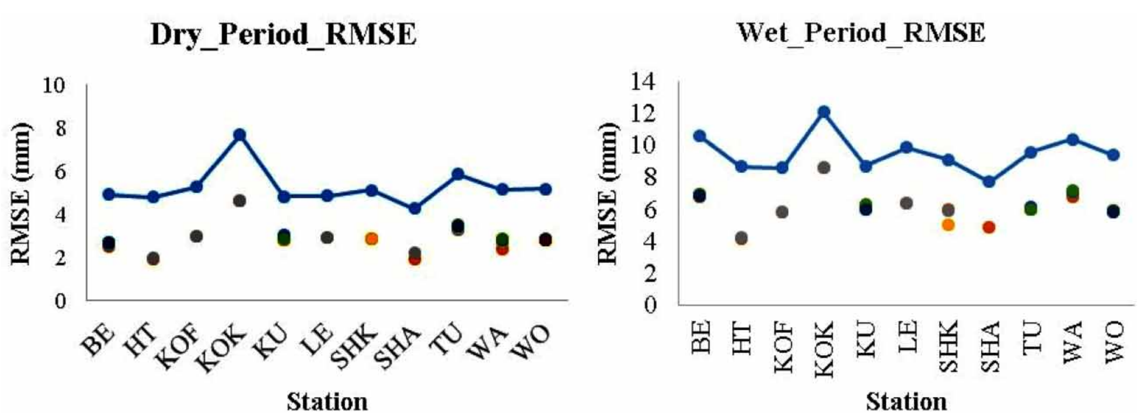


Figure 7 | Conditionally merged product RMSE result for all iterations (dots) compared to RMSE results after resampling (solid line) for dry and wet periods.

QM method was followed by the CM technique. However, this is due to the inability of the study to consider rainfall events in the bias correction process.

The after BC merged product of CHIRP SRE has shown significant improvement compared to the raw product. Especially, the wet period statistics result was better compared to the dry period statistics. Depending on the combination of coupled stations for validation, up to 70% PVE improvement for wet period and up to 50% (Figure 8) improvement was achieved at some stations (Shamena_Kedida, Wateraressa, and Hawassa_Tabor stations).

From the HB performance (Supplementary material, Table S8) one can observe most correctly detected events are underestimated (i.e. negative values) except Hawassa, and Kokossa in the wet season. The above result is the most efficient at a station performance out of 30 iterations by randomly choosing two hidden stations at a time. Comparing the BC and merged CHIRP output (Table 6) with the resampled CHIRPS estimates (Table 2) result indicates some performance measures (i.e. MB, HB, RMSE and POD) (Figure 9) have significantly improved.

However, looking at the error in volume of merged CHIRP products the estimation was poor at Kuyera station, and Belela station for dry period. Additionally, even though the error in volume at Kokossa and Shashemene stations has shown improvement when compared to CHIRPS estimate, the overall performance is not satisfactory. In general, the study signifies that the devised procedure to correct CHIRP SREs can be effective in reducing errors when compared to the CHIRPS product.

Spatial distribution/variability of an event on 16 August 2017 (i.e. by assuming that gauge records (Figure 10(a)) are the truth) illustrates that estimates by the resampled CHIRPS Version 2.0 (Figure 10(b)) have totally missed the event. However, the estimates by spatially BC and merged product CHIRP (Figure 10(c)) have captured the spatial coverage as well as the magnitudes at some stations. On the other hand from the estimates of the resampled CHIRP (Figure 10(d)) one can see that the northern part is wetter than the rest of the map, where the gauge records suggest heavier rainfall events in the western and central part of the map. Hence, the devised method is capable of correcting spatial pattern as well as the magnitude of the event.

3.4. Monthly SREs performance

Monthly estimates by both SRPs (i.e. CHIRP and CHIRPS) were assessed to investigate the impact of bias correction at different temporal resolution. Aggregating daily raw SRP and BC SREs at the monthly temporal resolution, the monthly estimates were compared (Table 7) using three statistical performance measures (Equations (5)–(7)). Performance tests were not conducted seasonally at this stage. The upgraded CHIRPS product has fairly good estimates at originally incorporated gauging station (i.e. such as Hawassa gauging station). Examining the error in volume over the study period BC CHIRP estimates have outperformed the remaining estimates at all stations except at Hawassa, Kokossa, and Wondo_genet stations. Additionally, BC CHIRP estimates have underestimated observations at Kuyera, Wondo_genet, and Leku stations.

When compared to resampled CHIRPS estimates, Spatially Bias-Corrected and Conditionally Merged (SBCM) estimates showed slightly better correlation at all stations except for estimates at Watararessa, Shashemene, and Kuyera stations

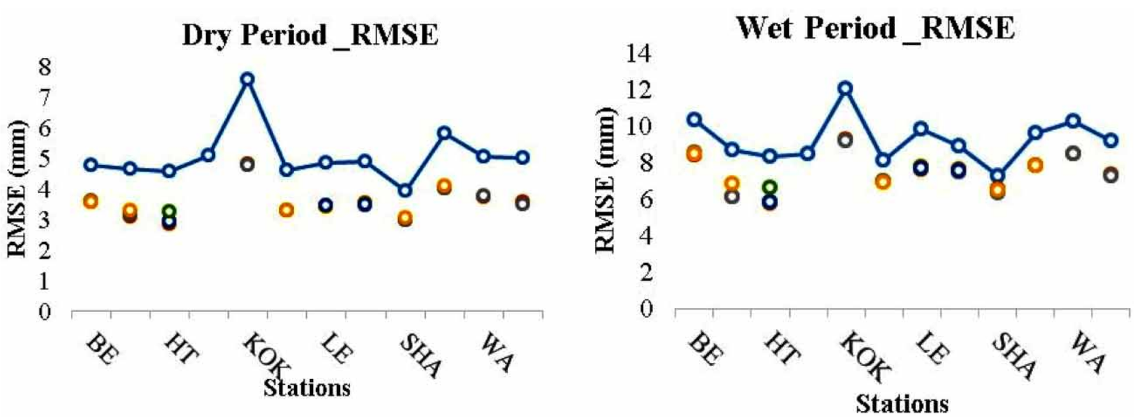


Figure 8 | Bias-corrected and merged product RMSE results for all iterations (dots) compared to RMSE results after resampling (solid line) for dry and wet periods.

Table 6 | Performance result for conditionally merged after bias correction of CHIRP SREs

	Dry period		Wet period	
Stations	**PVE (%)**	**POD**	**PVE (%)**	**POD**
BE	−53.36	0.93	− 13.04	0.99
HA	0.68	0.86	17.42	0.89
HT	−16.3	0.94	4.79	0.98
KOF	−10.17	0.91	− 8.75	0.95
KOK	45.89	0.78	34.33	0.92
KU	−52.92	0.87	− 17.77	0.8
LE	−17.64	0.95	− 19.8	0.98
SHK	−17.64	0.87	− 17.77	0.8
SHA	−81.45	0.9	− 72.97	0.81
TU	9.78	0.9	− 15.77	0.98
WT	−6.51	0.81	− 5.84	0.94
WO	15.7	0.79	16.29	0.92

(Table 7). Hence, the study has signified the importance of multistage bias correction methods to improve bias corrections for ungauged gridded SREs. This is supported by the comparably better result of SBCM estimates at the under-represented western part of the watershed which is only represented by the Shamena_Kedida station.

In conclusion, the study has proved that the SBCM approach is capable of improving the quality of SREs. The method is expected to improve SREs for other areas, given that all assumptions and requirements of the method are fulfilled. Especially, considering the decline of supplied ground station data to archives such as Climate Hazards Group (CHG)

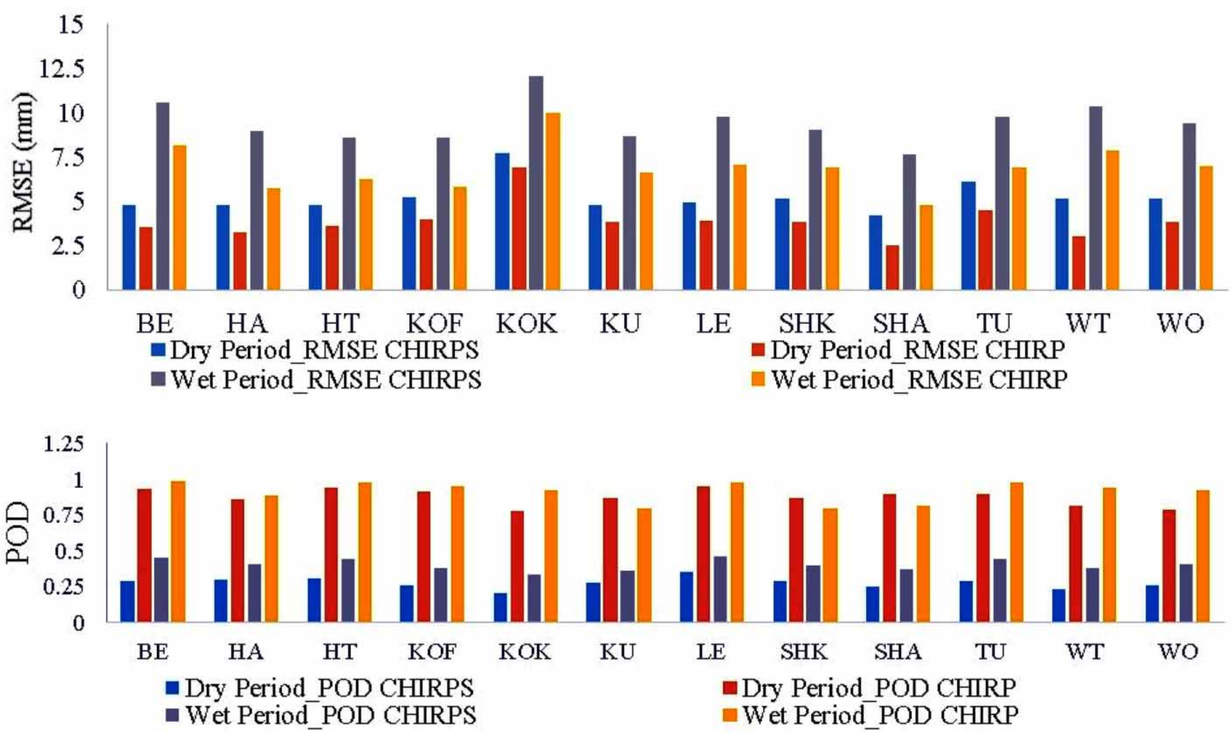


Figure 9 | Comparison of BC and merged CHIRP with resampled CHIRPS product.

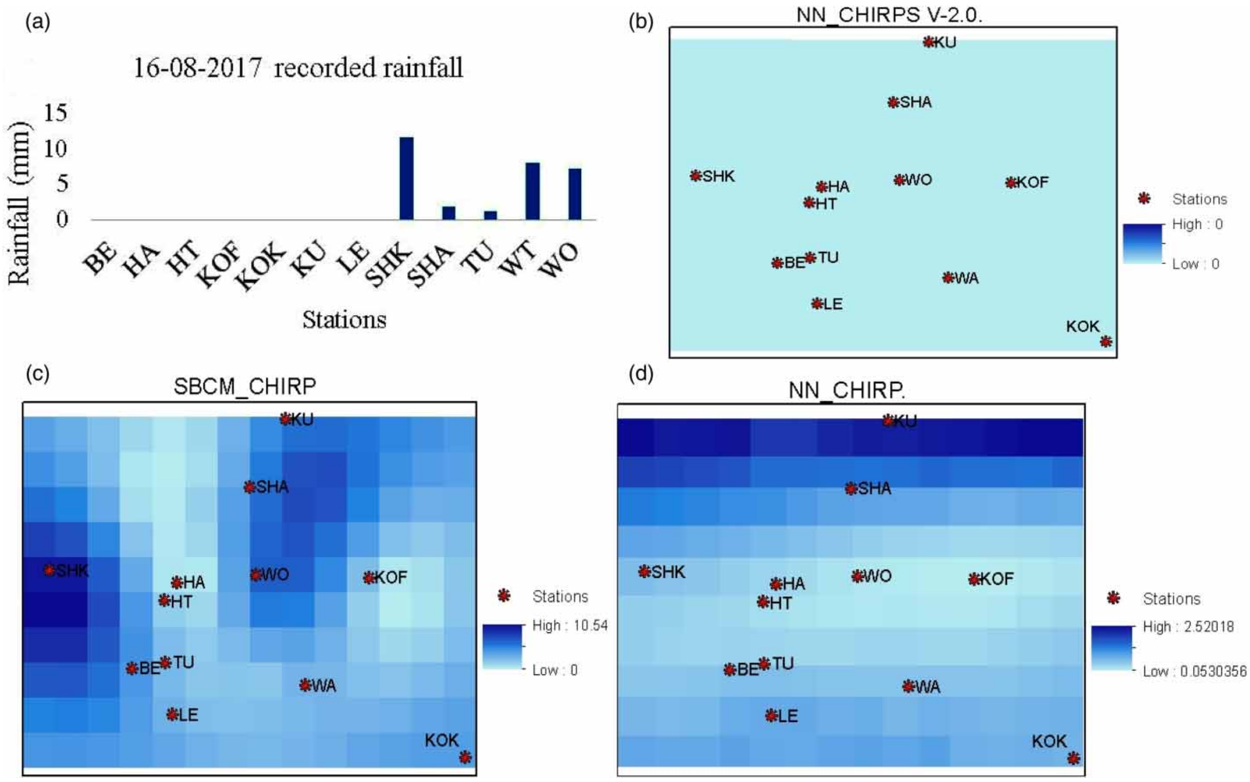


Figure 10 | Spatial distribution of rainfall estimates recorded on 16 August 2017 for (a) rainfall records at stations; (b) resampled (NN) CHIRPS version 2.0; (c) spatially BC and merged CHIRP; and (d) resampled (NN) CHIRP.

Table 7 | Comparison of raw CHIRP and CHIRPS product at collocated ground stations with BC, and SBCM

	RMSE (mm)				R^2				PVE (%)			
Station	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP	BC CHIRP	Raw CHIRP	Raw CHIRPS	SBCM CHIRP
WA	96.28	105.63	100.34	100.59	0.3	0.16	0.23	0.22	−3.55	−16.94	−14.77	−6.08
TU	45.13	47.23	41.26	40.22	0.58	0.52	0.63	0.67	−6.96	−11.57	−10.38	−9.31
HA	38.26	39.74	33.75	32.23	0.67	0.62	0.73	0.83	−5.86	−1.78	3.17	17.12
HT	43.51	46.06	44.8	42.79	0.52	0.49	0.5	0.53	−6.99	−14.53	−9.65	−5.6
SHA	21.09	53.48	49.77	59.9	0.66	0.46	0.51	0.31	−2.74	−89.41	−78.87	−74.95
KO	33.37	43.81	36.42	39.24	0.7	0.5	0.65	0.65	−6.84	−2.9	−0.86	−9.37
WO	43.04	45.55	41.87	42.66	0.66	0.58	0.66	0.69	9.66	3.54	7.34	16.18
SHK	46.68	51.81	48.93	42.35	0.57	0.53	0.56	0.63	−11.63	−39.79	−33.45	−22.83
LE	41.69	49.93	47.13	42.35	0.61	0.55	0.61	0.7	11.48	−20.07	−18.81	−19.37
KU	37.39	42.66	41.28	53.79	0.59	0.5	0.53	0.43	20.89	−20.95	−19.23	−23.2
KO	76.97	105.39	103.16	93.75	0.48	0.36	0.4	0.49	−3.66	43.38	42.9	37.59
BE	54.5	71.38	71.36	64.01	0.68	0.45	0.44	0.53	−2.95	−30	−25.04	−18.3

(Funk *et al.* 2015), the SBCM technique is useful for including unused datasets available from different sources. According to Funk *et al.* (2015), the number of ground station records available for the development of CHIRPS 2.0 has declined from 2004 stations in the year 2004–500 stations in the year after 2010 in Africa. Hence, the method can be implemented for areas having supplementary rainfall records which are not considered in the development of CHIRPS version 2.0. Furthermore, this study suggests that the performance of the SBCM improves for areas with scarce rain gauge stations. Thus implementation of the method for under-represented areas is highly encouraged in comparison to using readily available BC or merged SREs.

4. CONCLUSION

Nowadays different sources of rainfall records are flourishing, as the requirement for more accurate high-resolution records is high. However, each source is limited in providing the demanded quality data, in different aspects of quality. Hence, the importance of developing new and improved techniques for fusing multiple data sources is increasing. Similarly, in light of the systematic error and random error found in satellite rainfall products the demand for a more accurate bias correction technique is increasing. However, each method has different limitations: as an illustration, some bias correction techniques are better at correcting patterns rather than extreme events. Other methods used to merge multiple sources suffer from preserving the pattern of *in situ* observation.

Hence, this study has provided the potential to produce accurate rainfall maps, on a daily time scale, by using multistage procedure to correct CHIRP and CHIRPS satellite rainfall products. Additionally, by evaluating the performance of each stage, the significance of each method devised in each stage was illustrated. However, the performance achieved is limited to the study area, and does not limit the procedure to be tested to other study areas. Furthermore, the study is limited in bias correcting no rainfall frequency.

From the result, it has been proven that the study area BL resampling technique performs better. Additionally, it has been signified that the objective resampling technique helps to reduce the bias due to spatial mismatch between ground observation and selected grid estimates of SREs. The resampling outputs were BC using the parametric empirical QM method. The result indicates that, when good fit distribution equations specific to a given data are used, bias correction of raw satellite rainfall products can improve their performance when compared to *in situ* observation. Further assessment on transferring bias correction parameters from grid points with *in situ* records to grid points with no ground observation was performed. Also, the result has illustrated that one can effectively transfer bias correction parameters from available grid points with control ground stations to other grids through interpolation.

Finally, merging satellite rainfall products with ground observation is highly capable of preserving extreme events as well as preserving patterns of rainfall. However, the study is limited to correct dry-day frequency, which has significantly affected the performance bias correction. Hence, it is recommended that future studies focus on such fusing techniques to access the merits of different methods and provide improved bias correction procedures.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Adane, G. B., Hirpa, B. A., Lim, C.-H. & Lee, W.-K. 2021 Evaluation and comparison of satellite-Derived estimates of rainfall in the diverse climate and terrain of central and Northeastern Ethiopia. *Remote Sensing* **13**, 1275.

Baboo, S. S. & Devi, M. R. 2010 An analysis of different resampling methods in Coimbatore, District. *Global Journal of Computer Science and Technology* **10**, 61–66..

Belete, M. D., Diekkrüger, B. & Roehrig, J. 2017 Linkage between water level dynamics and climate variability: the case of Lake Hawassa hydrology and ENSO phenomena. *Climate* **5**, 21.

Beyene, T. D., Moges, M. A. & Tilahun, S. A. 2018 Development of rainfall disaggregation model in the Awash River Basin, Ethiopia. In *International Conference on Advances of Science and Technology*. Springer, pp. 50–64.

Bhatti, H. A., Rientjes, T., Haile, A. T., Habib, E. & Verhoef, W. 2016 Evaluation of bias correction method for satellite-based rainfall data. *Sensors* **16**, 884.

Brandsma, T. & Können, G. 2006 Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level. *International Journal of Climatology: A Journal of the Royal Meteorological Society* **26**, 75–89.

Degife, A., Worku, H., Gizaw, S. & Legesse, A. 2019 Land use land cover dynamics, its drivers and environmental implications in Lake Hawassa Watershed of Ethiopia. *Remote Sensing Applications: Society and Environment* **14**, 178–190.

Dile, Y. T., Tekleab, S., Ayana, E. K., Gebrehiwot, S. G., Worqlul, A. W., Bayabil, H. K., Yimam, Y. T., Tilahun, S. A., Daggupati, P. & Karlberg, L. 2018 Advances in water resources research in the Upper Blue Nile basin and the way forward: a review. *Journal of Hydrology* **560**, 407–423.

Dinku, T., Ceccato, P. & Connor, S. J. 2011 Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *International Journal of Remote Sensing* **32**, 5965–5979.

Dinku, T., Hailemariam, 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. *International Journal of Climatology* **34**, 2489–2504.

Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J. & Li, Z. 2021 Grand challenges in satellite remote sensing. *Frontiers in Remote Sensing* **2**, 619818.

Ebrahimi, S., Chen, C., Chen, Q., Zhang, Y., Ma, N. & Zaman, Q. 2017 Effects of temporal scales and space mismatches on the TRMM 3B42 v7 precipitation product in a remote mountainous area. *Hydrological Processes* **31**, 4315–4327.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L. & Hoell, A. 2015 The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. *Scientific Data* **2**, 1–21.

GDAL/OGR Contributors, W. 2020 *GDAL/OGR Geospatial Data Abstraction Software Library*. Open Source Geospatial Foundation, Chicago, IL.

Gebere, S. B., Alamirew, T., Merkel, B. J. & Melesse, A. M. 2015 Performance of high resolution satellite rainfall products over data scarce parts of Eastern Ethiopia. *Remote Sensing* **7**, 11639–11663.

Gebremedhin, M. A., Lubczynski, M. W., Maathuis, B. H. & Teka, D. 2021 Novel approach to integrate daily satellite rainfall with in-situ rainfall, Upper Tekeze Basin, Ethiopia. *Atmospheric Research* **248**, 105135.

Gella, G. W. 2019 Statistical evaluation of High Resolution satellite precipitation products in arid and semi-arid parts of Ethiopia: a note for hydro-meteorological applications. *Water and Environment Journal* **33**, 86–97.

Gillies, S. 2013 *The Shapely User Manual*. Available from: https://pypi. org/project/Shapely.

Gillies, S. 2019 rasterio Documentation. *MapBox, July*, 23.

Goshime, D. 2020 *Integration of Satellite and Ground-Based Rainfall Data for Water Resources Assessment in Central Rift Valley Lakes Basin, Ethiopia*. CY Cergy Paris Université, Paris.

Grus, J. 2019 *Data Science From Scratch: First Principles with Python*. O'Reilly Media. Inc., Sebastopol, CA.

Guenzi, D., Fratianni, S., Boraso, R. & Cremonini, R. 2017 Condmerg: an open source implementation in R language of conditional merging for weather radars and rain gauges observations. *Earth Science Informatics* **10**, 127–135.

Gupta, R., Bhattarai, R. & Mishra, A. 2019 Development of climate data bias corrector (CDBC) tool and its application over the agro-ecological zones of India. *Water* **11**, 1102.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S. & Smith, N. J. 2020 Array programming with NumPy. *Nature* **585**, 357–362.

Heo, J.-H., Ahn, H., Shin, J.-Y., Kjeldsen, T. R. & Jeong, C. 2019 Probability distributions for a quantile mapping technique for a bias correction of precipitation data: a case study to precipitation data under climate change. *Water* **11**, 1475.

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y. & Li, L. 2019 Rainfall spatial estimations: a review from spatial interpolation to multi-source data merging. *Water* **11**, 579.

Iqbal, Z., Shahid, S., Ahmed, K., Wang, X., Ismail, T. & Gabriel, H. F. 2022 Bias correction method of high-resolution satellite-based precipitation product for Peninsular Malaysia. *Theoretical and Applied Climatology* **148**, 1429–1446.

Jones, E., Oliphant, T. & Peterson, P. 2001 *SciPy: Open Source Scientific Tools for Python*.

Jongjin, B., Jongmin, P., Dongryeol, R. & Minha, C. 2016 Geospatial blending to improve spatial mapping of precipitation with high spatial resolution by merging satellite-based and ground-based data. *Hydrological Processes* **30**, 2789–2803.

Katiraie-Boroujerdy, P.-S., Rahnamay Naeini, M., Akbari Asanjan, A., Chavoshian, A., Hsu, K.-L. & Sorooshian, S. 2020 Bias correction of satellite-based precipitation estimations using quantile mapping approach in different climate regions of Iran. *Remote Sensing* **12**, 2102.

Kimani, M. W., Hoedjes, J. C. & Su, Z. 2018 Bayesian bias correction of satellite rainfall estimates for climate studies. *Remote Sensing* **10**, 1074.

Lafon, T., Dadson, S., Buys, G. & Prudhomme, C. 2013 Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology* **33**, 1367–1381.

Legesse, D., Vallet-Coulomb, C. & Gasse, F. 2004 Analysis of the hydrological response of a tropical terminal lake, Lake Abiyata (Main Ethiopian Rift Valley) to changes in climate and human activities. *Hydrological Processes* **18**, 487–504.

Lehner, F., Nadeem, I. & Formayer, H. 2020 An improved statistical bias correction method that also corrects dry climate models. *Hydrology and Earth System Sciences Discussions* 1–23.

Li, J. & Heap, A. D. 2008 *A Review of Spatial Interpolation Methods for Environmental Scientists*. Geoscience Australia, Canberra.

Ma, Q., Xiong, L., Xu, C.-Y. & Guo, S. 2018 Assessing the adequacy of bias corrected IMERG satellite precipitation estimates using extended mixture distribution mapping method over Yangtze River basin. In *MATEC Web of Conferences*. EDP Sciences, p. 01096.

Ma, Q., Xiong, L., Xia, J., Xiong, B., Yang, H. & Xu, C.-Y. 2019 A censored shifted mixture distribution mapping method to correct the bias of daily IMERG satellite precipitation estimates. *Remote Sensing* **11**, 1345.

Mcginnis, S., Nychka, D. & Mearns, L. O. 2015 A new distribution mapping technique for climate model bias correction. In *Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the 4th International Workshop on Climate Informatics*. Springer, pp. 91–99.

Mckinney, W. 2011 . pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing* **14**, 1–9.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50**, 885–900.

Murphy, B. S. 2014 PyKrige: development of a kriging toolkit for Python. AGU fall meeting abstracts, H51K-0753.

Navas, R., Sapriza, G., Vervoort, W. & Baethgen, W. 2019 On the assimilation of hourly data to scale CHIRPS daily precipitation in Uruguay. In *Geophysical Research Abstracts*.

Omondi, C. K. 2017 *Assessment of Bias Corrected Satellite Rainfall Products for Streamflow Simulation: A TOPMODEL Application in the Kabompo River Basin, Zambia*. University of Twente, Enschede.

Pignone, F., Rebora, N. & Silvestro, F. 2015 Modified Conditional Merging technique: a new method to estimate a rainfall field combining remote sensed data and raingauge observations. In *EGU General Assembly Conference Abstracts*. p. 3013.

Riverbank Computing. 2016 PyQt5 reference guide. PyQt. https://doc. bccnsoft.com/docs/PyQt5.

Soo, E. Z. X., Jaafar, W. Z. W., Lai, S. H., Othman, F., Elshafie, A., Islam, T., Srivastava, P. & Hadi, H. S. O. 2020 Evaluation of bias-adjusted satellite precipitation estimations for extreme flood events in Langat river basin, Malaysia. *Hydrology Research* **51**, 105–126.

Suliman, A. H. A., Awchi, T. A., Al-Mola, M. & Shahid, S. 2020 Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq. *Atmospheric Research* **242**, 105007.

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R. & Demaria, E. 2017 Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences* **21**, 2649–2666.

Teutschbein, C. & Seibert, J. 2012 Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *Journal of Hydrology* **456**, 12–29.

Themeßl, M. J., Gobiet, A. & Heinrich, G. 2012 Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change* **112**, 449–468.

Valdés-Pineda, R., Demaría, E., Valdés, J. B., Wi, S. & Serrat-Capdevilla, A. 2016 Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa. *Hydrology and Earth System Sciences Discussions* 1–28.

Wang, Y. & Zhao, N. 2022 Evaluation of eight high-Resolution gridded precipitation products in the Heihe River Basin, Northwest China. *Remote Sensing* **14**, 1458.

Xie, P. & Xiong, A. Y. 2011 A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *Journal of Geophysical Research: Atmospheres* **116**, D21106.

Zhang, Y., Ye, A., Nguyen, P., Analui, B., Sorooshian, S. & Hsu, K. 2021 Error characteristics and scale dependence of current satellite precipitation estimates products in hydrological modeling. *Remote Sensing* **13**, 3061.

Ziarh, G. F., Shahid, S., Ismail, T. B., Asaduzzaman, M. & Dewan, A. 2021 Correcting bias of satellite rainfall data using physical empirical model. *Atmospheric Research* **251**, 105430.

First received 22 November 2022; accepted in revised form 21 May 2023. Available online 31 May 2023