Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

Type of the Paper (Article)

Bayesian Bias Correction of Satellite Rainfall Estimates for Climate Studies

Margaret Wambui Kimani [1], Joost C.B. Hoedjes [1], Zhongbo Su [1]

[1] Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, 217 7500 AE, Netherlands

Correspondence to: Margaret W. Kimani (m.w.kimani@utwente.nl)

Abstract: Advances in remote sensing have led to use of satellite-derived rainfall products to complement the sparse rain gauge data. Although globally derived and some regional bias corrected, these products often show large discrepancies with ground measurements attributed to local and external factors that require systematic consideration. Decreasing rain gauge network however inhibits continuous validation of these products. We propose to deal with this problem by the use of Bayesian approach to merge the existing historical rain gauge information to create a consistent satellite rainfall data that can be used for climate studies. Monthly Bayesian bias correction is applied to the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS v2) data to reduce systematic errors using a corresponding gridded (0.05°) rain gauge data over East Africa for a period of 33 (1981–2013) years of which 22 years are utilized to derive error fields which are then applied to an independent CHIRPS data for 11 years for validation. The bias correction is spatially and temporally assessed during the rainfall wet months of March-May (MAM), June-August (JJA) and October–December (OND) in East Africa. Results show significant reduction of systematic errors at both monthly and yearly scales and harmonization of their cumulative distributions. Monthly statistics showed a reduction of RMSD (29–56)% and MAE (28–60)% and an increase of correlations (2–32) %, while yearly ones showed reductions of RMSD (9-23)%, and MAE (7–27)% and increase of correlations (4–77)% for MAM months, reduction of RMSD (15–35)% and MAE (16–41)% and increase in correlations (5–16)% for JJA months, and reduction of RMSD (3–35)% and MAE (9–32)% and increase of correlations (3–65)% for OND months. Systematic errors of corrected data were influenced by local processes especially over Lake Victoria and high elevated areas. Large-scale circulations induced errors were mainly during JJA and OND rainfall seasons and were reduced by the separation of anomalous years during training. The proposed approach is recommended for generating long-term data for climate studies where consistencies of errors can be assumed.

Keywords: Bayesian bias correction; satellite rainfall; rain gauge; climate studies; East Africa

---

1. Introduction

High temporal and spatial rainfall distribution is vital for many applications such as climate studies, water resource management and agriculture. Rain gauges provide the most direct representations of rainfall, but their distribution over land is sparse, especially in mountainous areas [1], and being point observations, they lack spatial representativeness. Use of satellite rainfall products is increasing because of their high spatiotemporal coverage. However, these products often exhibit large discrepancies with ground measurements [2,3] and the errors need to be reduced to make the products more representative of the rainfall variability. This has been done at global scale (Krajewski et al., 2000; Huffman et al., 2007; Arkin and Xie, 1994) and some regional evaluations [4,5], but relatively few efforts have been made to reduce the often large errors that

 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

occur at local scales. Studies have found that satellite rainfall products have systematic errors that cause overestimations/ underestimations [6] [7,8], especially on high elevated areas.

Different bias correction approaches for improving the satellite rainfall estimates have been proposed. [7] applied bias correction using empirical cumulative distribution (CDF) maps on a seasonal basis for hydrological applications in the upper Blue Nile in Ethiopia. The choice of the seasonal scale was meant to reduce error related to temporal variability but in areas of high rainfall variabilities, seasonal scale may not capture such variabilities. It is worth noting that effectiveness of bias correction of rainfall products may differ from location to location and consideration of spatial scale is of great importance. This was observed by [9,10] who used of quantile mapping approach to bias correct model rainfall products and observed that the approach improves the estimates in some locations, while it degrades in others.

[9] assessed the performance of two bias correction methods; successive correction method (SCM) and optimal interpolation. Qualitative analysis and visual inspections showed better results by SCM [11]. However, the study noted the limitation of this approach in defining the optimal weight of the error distributions. [12] evaluated satellite rainfall estimates combined with high-resolution rain gauge data using different bias correction methods based on additive and multiplicative approach. The evaluation was carried out on monthly basis in different rainfall seasons and with different rain gauge network and revealed that the choice of the temporal and spatial scale of the rain gauge data is vital for effective bias correction.

[13] used probabilistic Bayesian approach which requires historical rain gauge and satellite data to create satellite estimates-rain gauge data relationship which is then applied in the absence of gauge data. The assumption of this approach is that error is consistent in time and the error weight derived from the climatology is, therefore, a representative of a given region. The study was carried on a high temporal resolution aimed at improving hydrological applications. One notable observation is the impact of rain gauge distribution used in training showed significant impact on the effectiveness of the approach.

Because rain gauge distributions are decreasing [14] especially over the African countries because of their cost of maintenance this means availabilities of the rain gauge data to validate the increasing satellite rainfall products may be affected by inconsistencies of the rain gauge network. To solve this problem we propose an approach that can be used with the existing historical rain gauge information to create a consistent satellite rainfall data for climate studies. A long term temporal scale bias correction is therefore applied on the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS v2) data to reduce systematic errors using a corresponding gridded (0.05^{o}) rain gauge data. The choice of CHIRPS v2 product is based on its high spatial resolution and long coverage period suitable for climate studies. Furthermore, a recent study [6] over East Africa showed a close correspondence of CHIRPS v2 with ground observations. This study further spatially evaluates how CHIRPS rainfall estimates compare with the gridded rain gauge data after bias correction on monthly and yearly timescale during the wet rainfall months (March-May, June-August, and October-December) over East Africa.

This paper is arranged as follows. Description of study areas and data are given in section 2. Bayesian approach and methods of evaluation are described in section 3. Results and discussion are given in section 4, followed by summary and conclusion in section 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

2. Study Region and Data

Study Region

Figure 1 shows the study area in East Africa that extends between 29°E and 42°E, and 12°S and 5°N and covers five countries: Kenya, Uganda, Tanzania, Burundi, and Rwanda. The region shows diverse topography delineated by the embedded elevation map. Two main rainy seasons are experienced during the months of March, April, and May (MAM) and October, November, and December (OND). The rainy seasons coincide with overlying of the low-pressure belt of the Inter-Tropical Convergence Zone (ITCZ). The ITCZ migrates from 15°S to 15°N between January and July and is characterized by convective activities that lead to increased precipitation. A third rainfall season is the JJA and affects a small part of western Kenya and Uganda but significantly affects water resources within the region and surroundings of Lake Victoria. Satellite-derived rainfall estimates are nowadays widely used over the region because of their good spatial coverage and consistency in time. Further, the rain gauge distributions are decreasing and none represented over mountainous areas but the present rain gauge distribution are still useful in validating the satellite rainfall products.

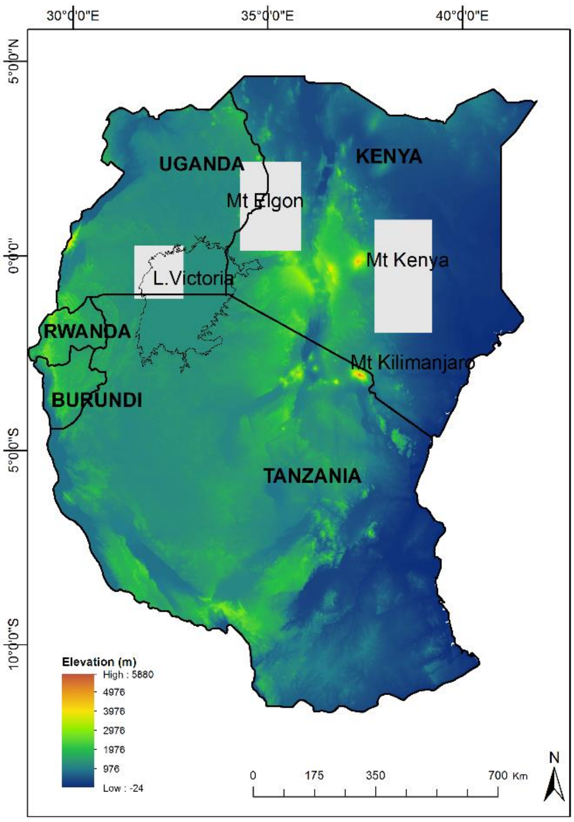

Figure 1: Map of East Africa, with Shuttle Radar Topography Mission (SRTM) 90 m digital elevation model. Highlighted are sections of areas of high rainfall amounts during March-May (Lake Victoria), June-August (Mt Elgon) and October-December (Mt Kenya) rainfall months.

Rainfall data

Two monthly rainfall data sets are used in this study and include CHIRPS v2.0 rainfall estimates and gridded (0.05°) rain gauge data.

CHIRPS is a quasi-global dataset developed by the United States Geological Survey (USGS) Earth Resources Observations and Science Centre and the University of California Santa Barbara Climate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

Hazards Group. It has a spatial resolution of 0.05°, and a daily/pentad/monthly temporal resolution. It uses TRMM multi-satellite precipitation analysis version 7 to calibrate the CCD rainfall estimates. The product covers the area between 50°N and 50°S, and data are available from January 1981 to the near present. CHIRPS v 2 data were used. Further details can be found in a study by [15,16], and an assessment of its performance relative to other products is provided in research by[5].

The gridded rain gauge data were provided by Intergovernmental Authority on Development (IGAD) Climate Prediction and Application Centre (ICPAC; available online at http://www.icpac.net**).** They applied interpolated, quality controlled available rain gauge measurements from 284 rainfall stations over East Africa. The GeoCLIM tool (http://wiki.chg.ucsb.edu/wiki/Geoclim) with the inverse distance weighting (IDW) (Zhang et al., 2014) was utilized. The GeoCLIM tool was developed by Tamuka Magadzire of the United States Geological Survey (USGS) Famine Early Warning Systems Network (FEWSNET) for rainfall, temperature, and evapotranspiration analysis. The data has been used for climate studies over East Africa and used for evaluation of satellite rainfall data [6].

Elevation data from the Shuttle Radar Topography Mission (SRTM) 90-m DEM (Digital Elevation Model) website (www. cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1) were used. The 5° spatial resolution tiles were mosaicked over East Africa through Geographical Information System (GIS) functionality. All the data were changed to 0.05° for compatibility (see below for details in methodology).

3 Methodology

We first describe the Bayesian method and then explain the training and testing procedures.

3.1 Bayesian method

Bayesian method is a probabilistic approach that merges data from different sources (Carlin and Louis, 1996) to get optimal representative values from the input datasets. It is based on spatial transformation, using the variances of the input datasets. In this study, it is used to adjust CHIRPS satellite rainfall estimates using the gridded rain gauge data for a period of 33 years in two steps. First, data from 22 years training period (1981- 2002) are used to calculate bias fields for the multi-annual monthly averages, yielding nine individual bias fields for each wet month. The monthly averaged bias fields are then used to correct an independent satellite rainfall estimates during an 11 year (2003-2013) validation period. The Bayesian approach is carried out at a 0.05° x 0.05° spatial scale for both data but for compatibility, the CHIRPS data are resampled using nearest neighbour [17] interpolation to match the georeference of the rain gauge data. The resampling is more robust in reprocessing algorithms according to this study.

3.1.1 Training period

The Bayes theorem [18] aims at getting the maximum likelihood of P(s|g), which is the conditional probability of the satellite estimates (s) given the gridded rain gauge data (g).

$$\boldsymbol{P}(\boldsymbol{s}|\boldsymbol{g}) = \frac{\boldsymbol{P}(\boldsymbol{s})\boldsymbol{P}(\boldsymbol{g}|\boldsymbol{s}))}{\boldsymbol{P}(\boldsymbol{g})} \quad (1)$$

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

Where, $P(s)$, $P(g|s)$ denotes the probability of satellite data and likelihood function of raingauge data given satellite estimates.

Since the gridded rainfall data distribution is known, $P(g)$ =1 then Eq. (1) reduces to Eq. (2):

$$\boldsymbol{P(s|g) = P(g|s)P(s)} \quad (2)$$

Following Talagrand [19], the least squares estimation can be used to simplify data assimilation problems to linear relationships. Equation (2) can, therefore, be changed from the probabilistic form into independent variables.

Assuming the monthly averaged errors (ε) of the satellite rainfall estimates and gridded rain gauge data to be unbiased and consistent in time and E the expected value as in Eq. (3).

$$\boldsymbol{E(\varepsilon_g) = E(\varepsilon_s) = 0} \quad (3)$$

The variances (σ^2) of each dataset can be related to the errors (ε), assuming the errors to be uncorrelated (Eq. (4) and (5).

$$\boldsymbol{E(\varepsilon_g^2) = \sigma_g^2} \quad (4)$$

$$\boldsymbol{E(\varepsilon_s^2) = \sigma_s^2} \quad (5)$$

Bias-corrected satellite estimates can then be represented as a linear combination of the gridded rainfall data and uncorrected satellite rainfall estimates. The weighing factors, $\boldsymbol{\alpha_g}$ and $\boldsymbol{\alpha_s}$, are dependent on the respective variances; the higher the variance of the respective dataset, the lower the corresponding weighting factor. This means that in areas where the variance of the reference gridded rain gauge dataset is high, the correction that is applied to the satellite data will be reduced. This is the case where large random errors from year to year are large during correction period and were not accounted for in the correction.

$$\boldsymbol{\bar{s}_c = \alpha_g \bar{g} + \alpha_s \bar{s}} \quad (6)$$

With the overbars denoting the averaged values for each month in the 22 years learning training dataset. Equation (6) again assumes the bias-corrected satellite estimates (s_c) to be unbiased as their errors are consistent during the training period. This may not be the case when the data is used for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

climate analysis because of the randomness of the errors arising from year to year. The training period is supposed to be long enough to include periodic external influences in error derivations. The sum of the satellite estimates' weighing factor, $\boldsymbol{\alpha_s}$, and the gridded rain gauge weighting factor, $\boldsymbol{\alpha_g}$, equals one.

$$\boldsymbol{\alpha_g + \alpha_s = 1} \tag{7}$$

s_c will best estimate g if the weighing factors α_g and α_s minimize the mean squared error of the variance of the corrected satellite estimates, $\boldsymbol{\sigma_c^2}$, with respect to α_g, following Eq. (8-10).

$$\boldsymbol{\frac{\partial \sigma_c^2}{\partial \alpha_g} \to 0} \tag{8}$$

$$\boldsymbol{\sigma_c^2 = (\bar{s}_c - \bar{g})^2} \tag{9}$$

$$\boldsymbol{\sigma_c^2 = \alpha_g^2 \sigma_g^2 + \left(1 - \alpha_g\right)^2 \sigma_s^2} \tag{10}$$

This leads to

$$\boldsymbol{\alpha_g = \frac{\sigma_s^2}{\sigma_g^2 + \sigma_s^2}} \tag{11}$$

$$\boldsymbol{\alpha_s = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_s^2}} \tag{12}$$

Equation (11 and 12) imply that the weights of the satellite estimates and the corresponding rain gauge data are related to the inverse of their variances. Using these weighting factors, the average satellite estimates for each month in the 22 years training dataset can then be corrected using the linear relationship shown in Eq.(6).which can be rewritten as shown eqn.(13)

$$\boldsymbol{\bar{s}_c = \bar{s} + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_g^2}(\bar{g} - \bar{s}) = \bar{s} + \alpha_g(\bar{g} - \bar{s})} \tag{13}$$

Equation (13) implies that when the variance of the reference data is very high, i.e. $\boldsymbol{\sigma_s \gg \sigma_g}$, then $\boldsymbol{\sigma_{g \to 0}}$ and s_c approaches s, and that when $\boldsymbol{\sigma_s \ll \sigma_g}$, $\boldsymbol{\sigma_{g \to 1}}$ and s_c approaches g.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

3.1.2 Testing period

In this section, the Bayesian approach is described using the error fields derived during training on monthly data. Evaluation of the corrected satellite estimates in relation to a corresponding rain gauge data on monthly and yearly timescale is used.

The bias fields were calculated from the satellite estimates for each wet month during MAM and OND rainfall season using Eq. (13).The subscript 'i' stand for the time step.

$$\boldsymbol{Bias} = \frac{1}{n}\sum_{i}^{N}(s_{ci} - g_{i}) \qquad (14)$$

The bias is then subtracted from satellite data of each corresponding month (subscript 'i') using Eq. (15)

$$s_{ci} = s_i - \boldsymbol{bias} \qquad (15)$$

3.2 An assessment of the quality of bias corrected rainfall

Validation of bias correction on CHIRPS satellite rainfall was carried out for a period of 11 years from 2003-2013. The corrected and raw data were compared with the gridded rain gauge data for the months of the rainy seasons (March-May, June-August and October-December). Continuous statistics of the correlation coefficient (cc), root mean square difference (RMSD), standard deviations (σ) (Eq. (16-17) and MAE (Eq. (18) were used to quantify their relationships and Taylor diagrams [20], spatial maps and plots used for visualization.

$$cc = \frac{\frac{1}{N}\sum_{i=1}^{N}(s_i - \bar{s})(g_i - \bar{g})}{\sigma_s \sigma_g} \qquad (16)$$

$$\mathbf{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - g_i)^2} \qquad (17)$$

where overbar stands for the respective mean satellite estimates (s) , gridded rain gauge datasets (g), and N the number of samples considered.

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{N}[s_i - g_i] \qquad (18)$$

3.3 Spatial distribution assessment of bias corrected rainfall

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

Because East Africa rainfall is influenced by external factors that occurs inter-annually and they influence the occurrences of the systematic errors in rainfall products, the bias-corrected CHIRPS estimates were assessed spatially on yearly timescale using equations (16-18). Cumulative distributions of monthly averages for each validation year (2003-2011) were utilized. Further, analysis were carried on raingauge gauge weight correction factor (equation 11) to establish the impact of largescale circulations on satellite estimates' systematic errors. The spatial distribution of bias corrected CHIRPS were assessed with respect to corresponding raingauge data during the validation period (2003-2013).

4. Results and Discussion

4.1 Evaluation of bias-corrected monthly CHIRPS

Bias correction was carried for all the months from January to December but only the wet month of March to May, June-August and October to December are discussed in this paper. The months of January, February and September are generally dry for most of East Africa region and were therefore excluded in the analysis**.** Figure 2 shows the Taylor diagrams displaying the error metrics before and after bias corrections during the wet months of March to May, June-August and October to December with respect to rain gauge data over East Africa. It is illustrated in this figure that Bayesian approach significantly improved the accuracy of the CHIRPS estimates. This is indicated by reduced RMSD and increased correlations for all the months except the month of August. The overcorrection in this month was attributed to erroneous inconsistencies caused by misrepresentation of the rainfall regime by rain gauge data (more details later). The reduction of systematic errors showed dependence on rainfall amounts and were, therefore, more during the months of increased rainfall of April, May and November. This is in line with what was recently documented in (Kimani et al., 2017), that satellite rainfall products underestimate high rainfall amounts over the region.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

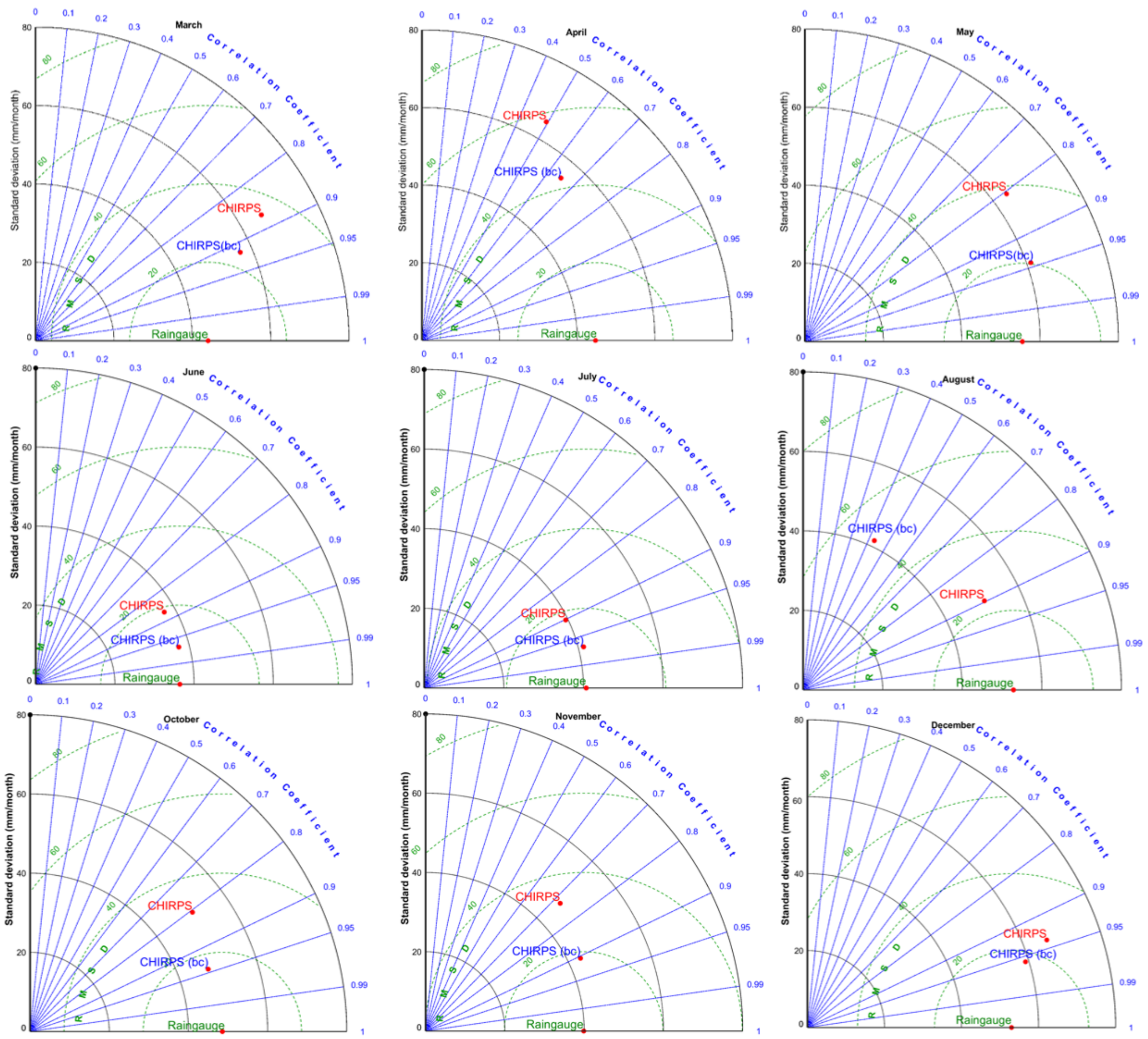

Figure 2: Monthly Taylor diagrams displaying statistical comparison between uncorrected (red) and corrected (blue) CHIRPS estimates with corresponding rain gauge data as the reference. Only the wet months of the rainfall season (March-May and June-August, October-December) over a period of 11 years (2003–2013) were utilized. The azimuthal angle represents correlation coefficient; radial distance the standard deviation (mm/month) and green contours represent RMSD (mm/month).

Figure 3 shows the spatial distribution of monthly averaged satellite rainfall estimates before and after bias corrections represented for each season and corresponding rain gauge data. The spatial patterns of bias-corrected (bc) display areas of improved rainfall estimates, and the change maps indicate the areas where correction of CHIRPS estimates was done.

It can be observed from rain gauge data high rainfall is over Mt Kenya, Lake Victoria and around Mt Elgon. CHIRPS estimates are able to capture those areas of highest rainfall but show overestimates over southern Tanzania. This overestimation is associated with the consistent high variance of the rain gauge, hence the corrected estimates approach the uncorrected ones. These findings are in line with (Tian et al., 2010) that used the probability distribution to adjust satellite rainfall estimates and associated with corrections to misrepresentation of rainfall variability by the rain gauge network. During JJA seasons represented by August month, the CHIRPS (bc) estimates overestimate rainfall amounts, especially over Mt Kenyan highlands. This again is attributed to

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

temporal inconsistencies of the patterns in rain gauge spatial rainfall that resulted in overcorrection. However, in November of OND season, the CHIRPS (bc) estimates are close to the rain gauge rainfall and (Kimani et al., 2017) reported CHIRPS underestimations on high elevated areas (over the same study areas) and it is credible that this approach adequately reduced these errors.

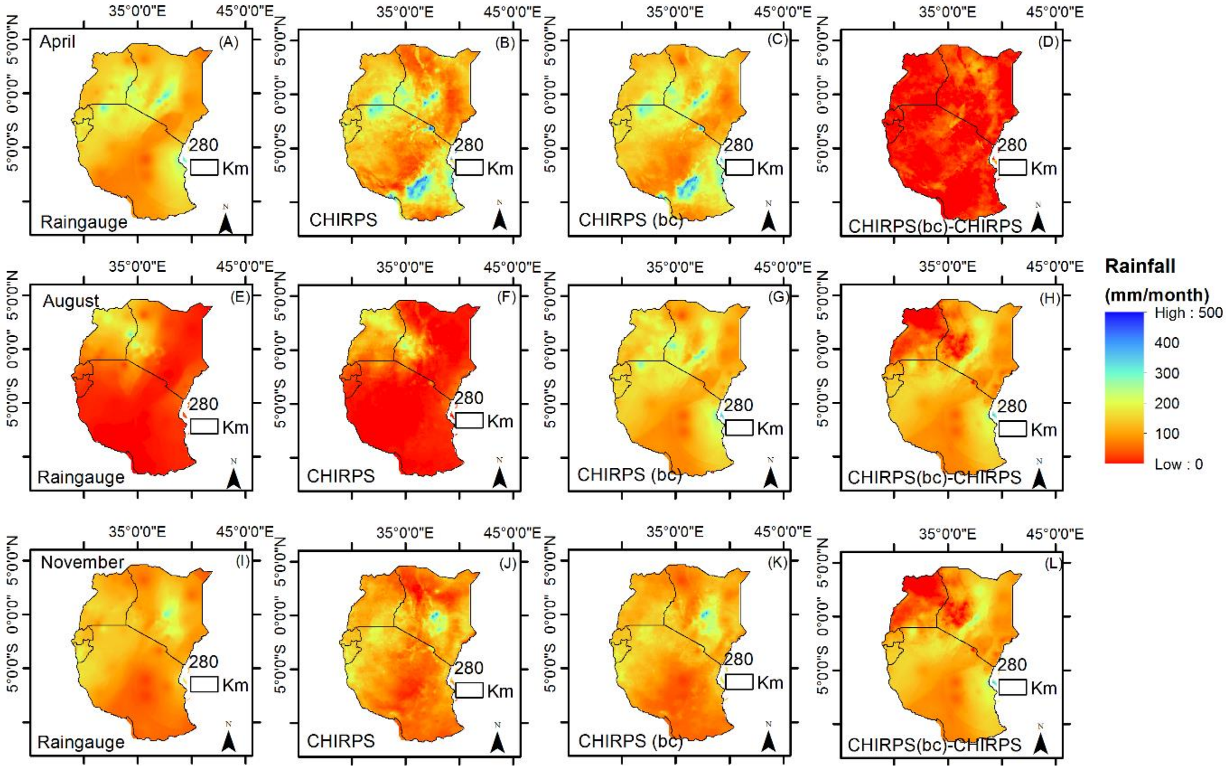

Figure 3: Monthly rainfall averages (2003-2013) of rain gauge data and satellite rainfall estimates before (CHIRPS), after bias corrections (CHIRPS (bc)) and the difference between CHIRPS (bc) and CHIRPS.

Figure 4 shows the CHIRPS systematic errors adjusted using empirical Cumulative Distribution Function (CDF) plots before and after bias corrections during MAM, JJA and OND rainfall seasons. Before corrections, CHIRPS overestimated the relatively low (<200 mm/month) rainfall amounts and underestimated high (>200 mm/month) amounts. This concurs with (Paredes-Trejo et al., 2017) that CHIRPS monthly estimates overestimate/underestimate low/high rainfall amounts. After bias correction, CHIRPS estimates in each of the nine months show significant change in spatial distribution close to the rain gauge data except in the month of August that show overcorrections.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

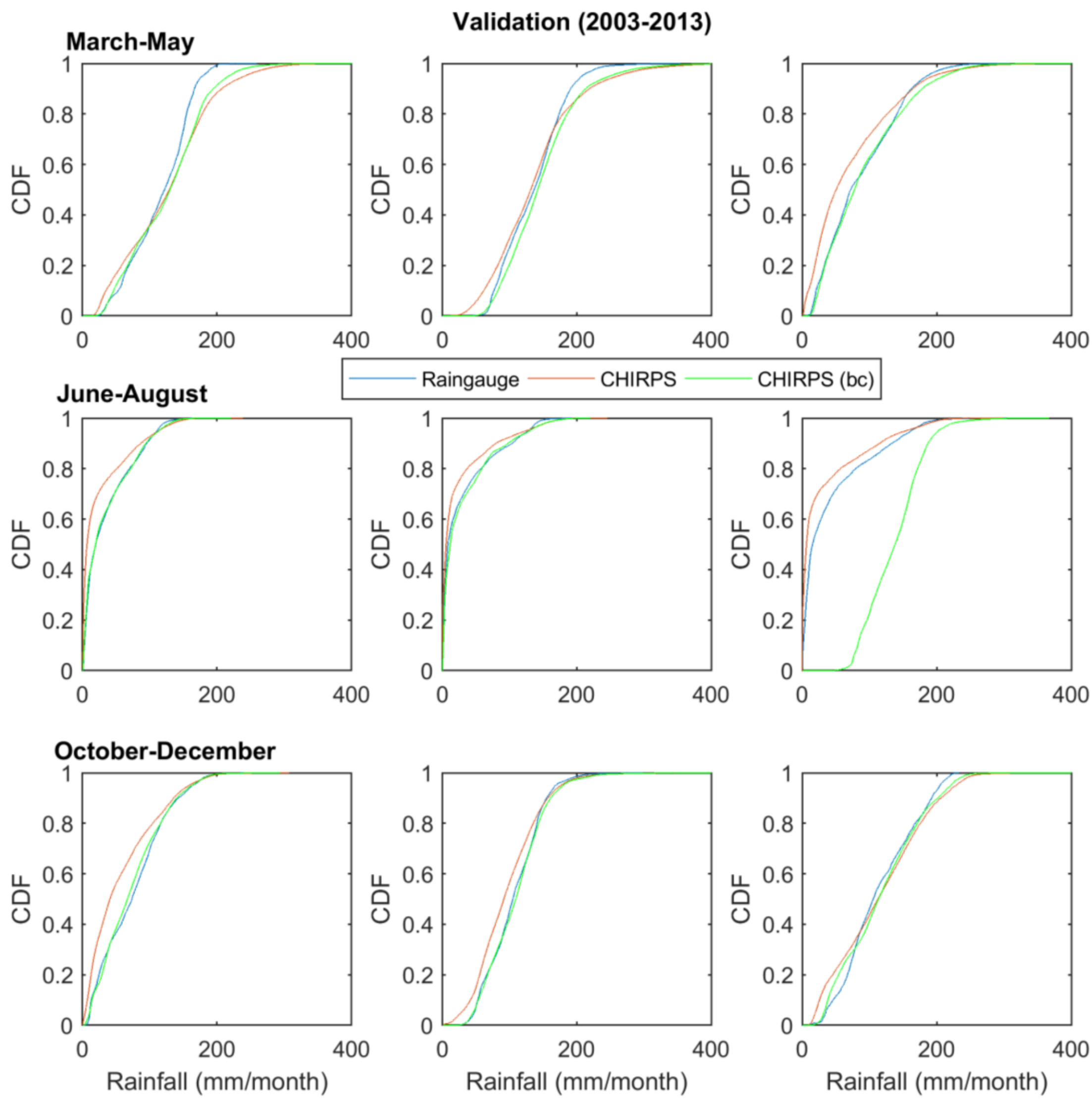

Figure 4: Empirical Cumulative Distribution Function (CDF) of monthly rain gauge data, raw satellite rainfall estimates (CHIRPS), and bias-corrected (CHIRPS (bc)) over the validation period (2003–2013).

Table 1 shows the monthly error statistics for all the wet months before and after bias corrections. The change of errors for each month is given in percentages. It is evident the bias corrected CHIRPS estimates show reductions in RMSD and MAE errors and increase of correlations. The overall monthly average reduction of RMSD (29-56) % and MAE (28-60) % and correlations increase (2-32) % is an indication of the high skill of the bias correction approach. It can be observed that the bias correction was successful for both high and low rainfall amounts. This is an indication that the dependence of corrected systematic errors not only on overall rainfall magnitudes but also on its distribution and regimes.

The corrected errors also showed dependence on seasons and this was indicated by RMSD and MAE reduction patterns that are high during JJA rainfall season. Consequently, a reduction of 50% and 60% of RMSD and MAE respectively were observed in the month of June. These changes are attributed to the onset of south-east monsoon in May that ends by November. This shows the corrected CHIRPS errors follow the rainfall systems affecting rainfall variabilities over East Africa.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

Table1: Statistics for the monthly spatial evaluation

Months	RMSD	RMSD (bc)	Change (%)	MAE	MAE (bc)	Change (%)	CC	CC (bc)	Change (%)	Rainfall (mm)
March	34.9	24.0	**-31**	24	16	**-32**	0.87	0.92	**5**	116
April	57.7	25.6	**-56**	38	24	**-37**	0.49	0.65	**32**	135
May	38.0	20.3	**-47**	29	14	**-51**	0.81	0.94	**17**	86
June	18.7	9.4	**-50**	14	6	**-60**	0.87	0.97	**11**	36
July	17.8	10.4	**-42**	11	6	**-44**	0.90	0.97	**7**	30
August	23.5	51.3	**118**	15	97	**545**	0.90	0.43	**-52**	42
October	31.0	16.1	**-48**	24	11	**-53**	0.81	0.94	**17**	75
November	32.7	18.4	**-44**	24	13	**-45**	0.73	0.91	**25**	106
December	24.5	17.4	**-29**	18	13	**-28**	0.94	0.96	**2**	114

Performances of CHIRPS and CHIRPS (bc) with respect to rain gauge data were further compared over Lake Victoria, Mt Elgon and Mt Kenya during the wettest (April and November) of MAM, OND and driest (JJA) rainfall seasons. These areas are significant in that they experience rainfall of different regimes from local effects. Figure 5 shows the CDFs of the rain gauge data, CHIRPS and CHIRPS (bc) over those areas. It is evident the CDFs of CHIRPS (bc) is aligned closer to those of the rain gauge data in all the months. However, it is also evident in highly elevated areas of Mt Kenya the cumulative distributions of the rain gauge data and CHIRP (bc) were not well aligned. This was more evident during April and November. It is worth noting these two seasons are influenced by ITZC and this large-scale circulation was associated with observed fluctuations in rainfall. These results concur with Sun et al., (2015) study that used model simulation to study the relationship between rainfall over Lake Victoria and surface temperature that showed the relationship is associated to influences of large-scale circulations and orography to rainfall variability. This suggests the systematic errors in the uncorrected CHIRPS may be linked to these external factors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

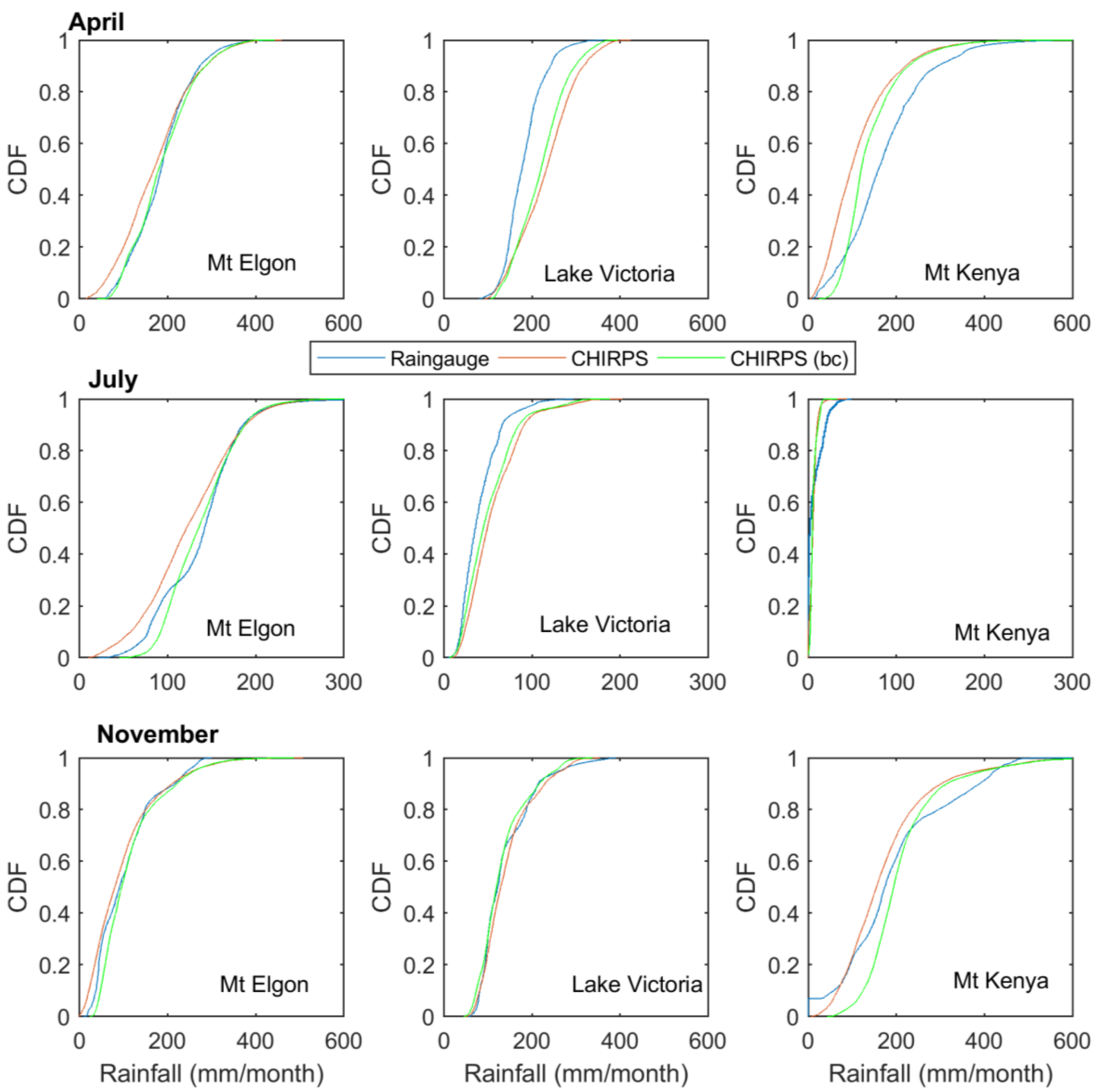

Figure 5: Empirical Cumulative Distribution Function (CDF) of monthly rain gauge data, satellite rainfall estimates (CHIRPS), and bias-corrected CHIRPS (bc) over the validation period (2003–2013). Shown are only areas within which high rainfall amounts are experienced during March-May (Lake Victoria), June-August (Mt Elgon) and October-December (Mt Kenya) as represented by the months of April, July and November respectively.

4.2 Yearly spatial distribution of bias corrected CHIRPS

In this section yearly CHIRPS and the CHIRPS (bc), estimates are described in cumulative distribution plots in Figures 6, 7, and 8 of the month of April, July and November respectively.

In April (Figure 6), it is evident before bias corrections CHIRPS underestimates/overestimates low/high rainfall amounts. Underestimation was more evident in years 2004, 2006 and 2013, which are the wettest years during MAM season and were adjusted to align with rain gauge data. In July (Figure 7), rainfall amount is generally low and general overestimation before correction is observed. The bias correction significantly adjusted the rainfall distribution to match the rain gauge data. However in years 2004, and 2009-2013 the distribution plots show low skills in adjusting the rainfall estimates. From the rainfall averages, 2004 and 2009 are years with anomalous wet rain gauge records which explains the source of over-corrections.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

Similar overcorrection is evident in Figure 8 in 2013 and this one of the driest year during the validation period in OND rainfall season. From these results, it is clear that during MAM more systematic errors were reduced that corresponds to years of high rainfall amounts. However, during JJA and OND changes in the frequency of extreme rainfall amounts relative to other years caused irregularities of errors and over corrections resulted. The seasonality of the anomalous years suggests external influences.

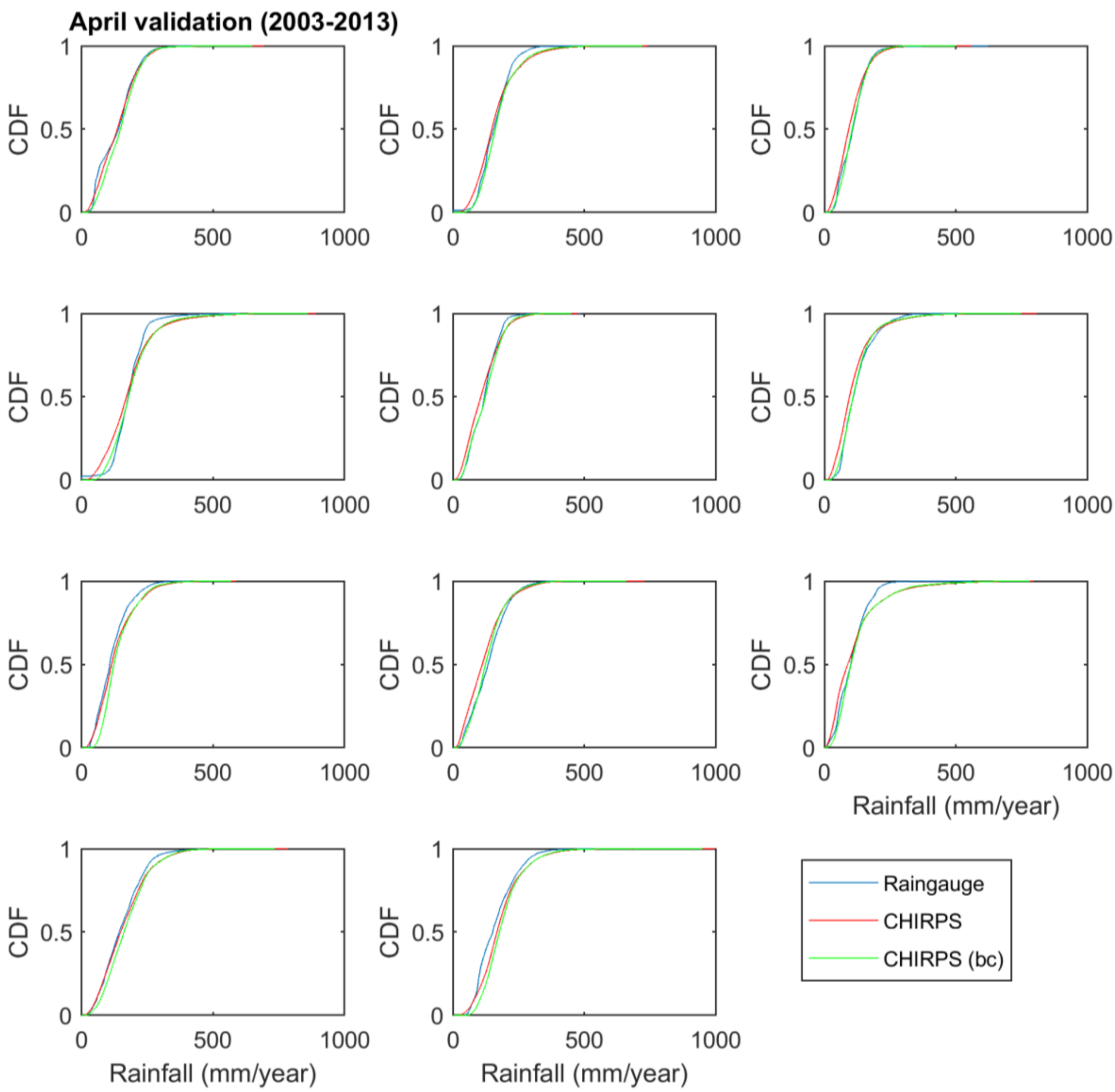

Figure 6: Empirical Distribution Function (CDF) of rain gauge data raw satellite rainfall estimates (CHIRPS), bias-corrected CHIRPS (bc) over the validation period (2003–2013) of the month of April.

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

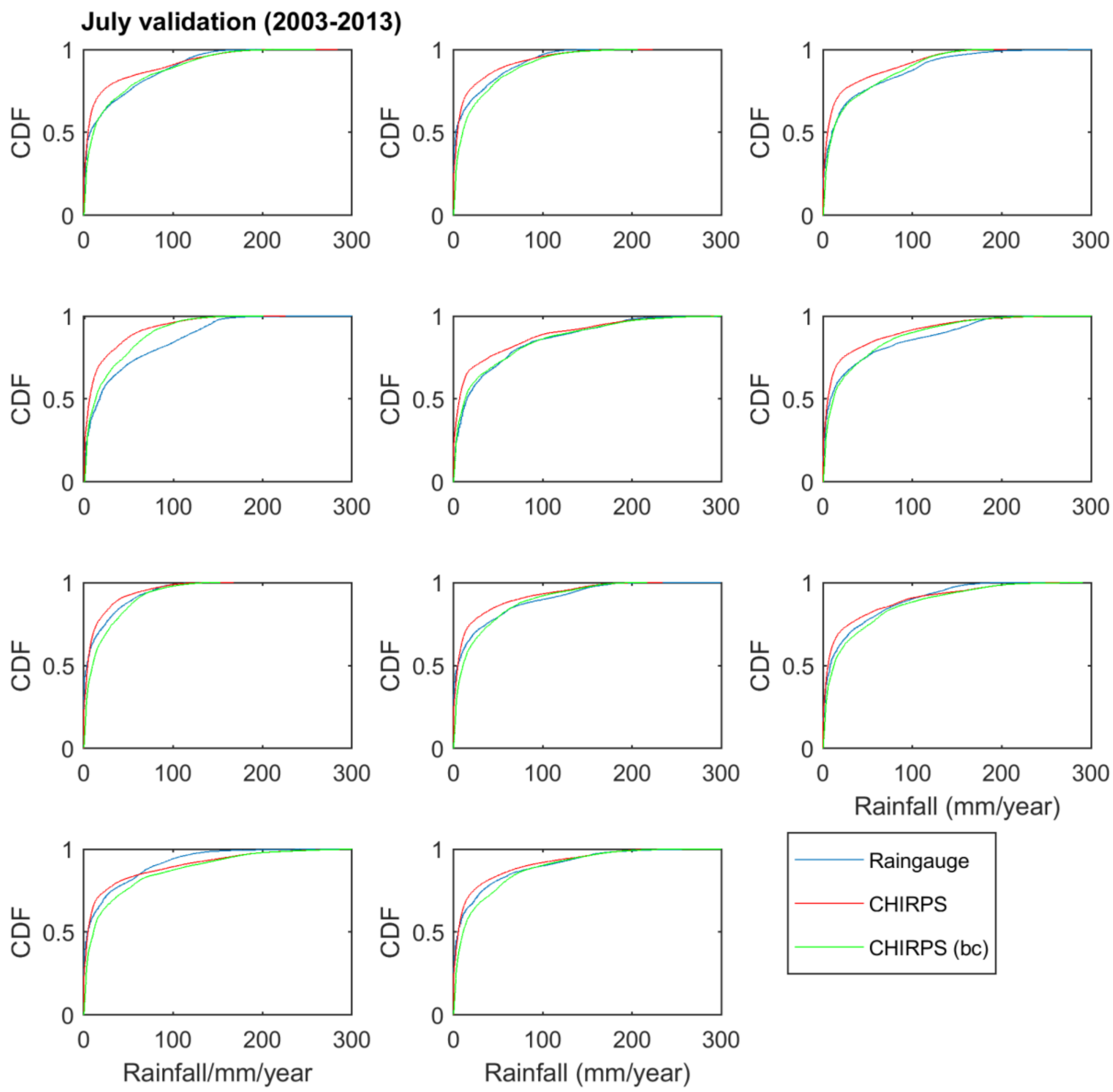

Figure 7: The same as Figure 6, except for the month of July.

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

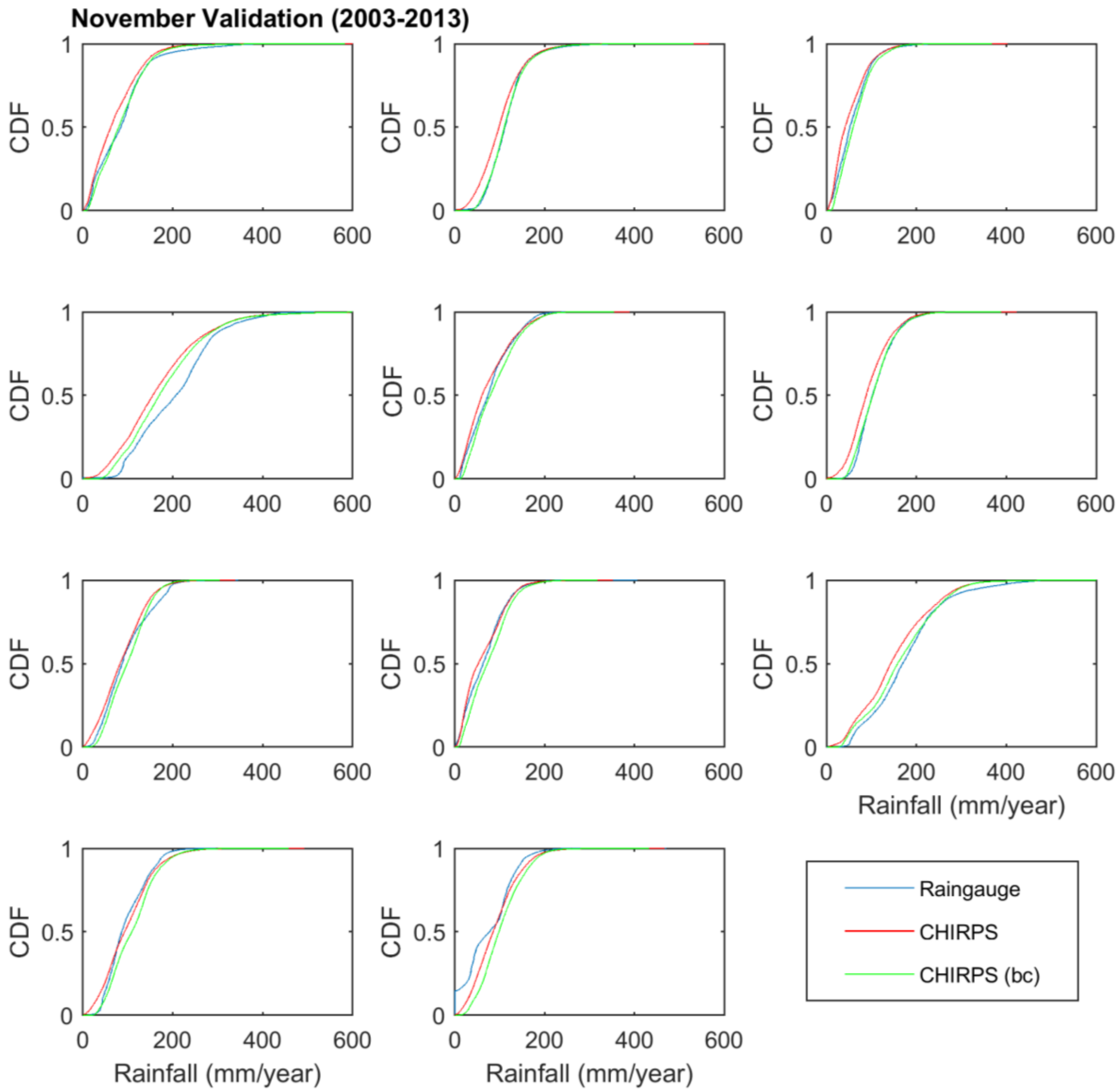

Figure 8: The same as Figure 6, except for the month of November.

The statistics of yearly evaluation are summarized in Table 2. The improvement in spatial pattern during the three rainfall seasons as represented by each month show reduced RMSD (mm/year) and MAE (mm/year), and the corresponding increase in correlations. It is evident, in April the corrected large systematic errors correspond with years of high rainfall and consequently, the dry years have fewer errors. The wettest year (2006) show the highest change in correlations (77%) and the driest year (2009) the least (<10%). In July, there is a significant reduction of errors but the approach showed low skills in eradicating errors of anomalous increased rainfall observed in the years 2004 and 2009. About four times rainfall magnitudes different from other validation years was observed during these two years leading to inconsistent errors. Similar to July, in November, significant errors were reduced but in the anomalous dry year in 2013, least errors were corrected and overcorrection was observed. RMSD and MAE increased, while correlations decreased meaning besides errors dependence on rainfall magnitudes the consistencies of the errors inter-annually affect the performance. The randomness of the errors is evident in the inter-annual analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

Table2: Statistics for the yearly spatial evaluation

April	**RMSD**	**RMSD (bc)**	**Change (%)**	**MAE**	**MAE (bc)**	**Change (%)**	**CC**	**CC (bc)**	**Change**	**Rainfall**
2003	66.7	55.1	**-17**	49	40	**-19**	0.56	0.70	**24**	**134**
2004	79.7	68.3	**-14**	56	44	**-21**	0.39	0.48	**24**	**158**
2005	50.8	40.5	**-20**	37	28	**-23**	0.57	0.68	**20**	**109**
2006	109.4	98.1	**-10**	77	68	**-11**	0.09	0.15	**77**	**179**
2007	51.5	40.9	**-21**	39	29	**-25**	0.63	0.74	**18**	**117**
2008	64.9	54.4	**-16**	45	36	**-21**	0.60	0.67	**12**	**123**
2009	61.6	56.2	**-9**	42	39	**-7**	0.61	0.66	**8**	**116**
2010	65.5	50.2	**-23**	47	35	**-27**	0.61	0.73	**19**	**135**
2011	80.3	73.5	**-9**	46	39	**-13**	0.60	0.63	**4**	**103**
2012	68.5	59.2	**-13**	49	41	**-16**	0.67	0.75	**13**	**152**
2013	85.4	74.1	**-13**	60	51	**-15**	0.44	0.58	**31**	**161**
July	RMSD	RMSD (bc)	Change (%)	MAE	MAE (bc)	Change (%)	CC	CC (bc)	change	Rainfall
2003	23.8	15.4	**-35**	14	8	**-41**	0.85	0.94	**10**	**30**
2004	17.1	15.6	**-9**	9	10	**2**	0.85	0.90	**6**	**203**
2005	26.4	20.0	**-24**	15	11	**-25**	0.87	0.92	**6**	**33**
2006	38.5	29.2	**-24**	22	18	**-20**	0.73	0.85	**16**	**39**
2007	26.2	20.1	**-23**	16	11	**-33**	0.90	0.94	**4**	**42**
2008	32.1	21.6	**-33**	17	12	**-31**	0.82	0.92	**11**	**36**
2009	14.0	16.3	**16**	8	10	**26**	0.84	0.82	**-2**	**164**
2010	22.1	14.1	**-36**	12	9	**-25**	0.88	0.95	**8**	**28**
2011	26.8	22.6	**-15**	15	12	**-19**	0.84	0.91	**8**	**30**
2012	29.8	30.5	**3**	14	15	**4**	0.84	0.88	**5**	**24**
2013	22.9	18.5	**-19**	12	10	**-16**	0.88	0.92	**5**	**28**
November	RMSD	RMSD (bc)	Change (%)	MAE	MAE (bc)	Change (%)	CC	CC (bc)	change	Rainfall
2003	50.6	32.8	**-35**	32	22	**-32**	0.67	0.86	**28**	**88**
2004	51.4	39.9	**-22**	39	30	**-23**	0.44	0.59	**35**	**115**

2005	29.2	23.1	**-21**	21	17	**-22**	0.72	0.82	**14**	**583**
2006	81.9	66.2	**-19**	62	51	**-19**	0.67	0.76	**12**	**205**
2007	34.6	27.9	**-19**	26	21	**-19**	0.77	0.85	**10**	**79**
2008	50.4	36.1	**-28**	36	27	**-25**	0.38	0.62	**65**	**108**
2009	36.5	32.5	**-11**	28	25	**-9**	0.76	0.79	**3**	**97**
2010	30.5	24.3	**-20**	21	18	**-16**	0.78	0.88	**13**	**66**
2011	75.4	59.1	**-22**	51	43	**-17**	0.63	0.75	**20**	**176**
2012	45.1	43.7	**-3**	33	30	**-9**	0.62	0.64	**4**	**97**
2013	58.2	63.9	**10**	43	45	**3**	0.41	0.35	**-16**	**77**

Further analysis was carried out to determine the influence of large-scale circulations to satellite rainfall systematic errors in CHIRPS. This was done by assessing the impact of excluding anomalous wet years (1997/1998) during El Niño. Figure 9 and 10 show the spatial and temporal distribution of bias correction weight when the El Niño years of 1997/1998 are included (w) and excluded (wc) and the difference (wc-w) respectively. From the weight distribution map in figure 9 during MAM and JJA rainfall months of April and July respectively, no significant difference occurs in bias correction weight. However, in November, increased performance after correction of systematic errors is evident particularly over Lake Victoria, southwestern Tanzania and Eastern Kenya. These changes are supported by temporal analysis in figure 10 where significant changes are observed from the month of August and including the OND rainfall months. It can be depicted that general circulations related to El Niño are associated with CHIRPS systematic errors observed inter-annually. These findings concur with the study (Indeje et al., 2000) that during this period influences of El Niño–Southern Oscillation (ENSO) are experienced over East Africa. The rainfall variabilities in Lake Victoria are also linked to largescale circulations (Sun et al., 2015). It is therefore evident that the interannual systematic errors of CHIRPS and any other satellite-derived rainfall estimates can be reduced more effectively by a prior knowledge of the external factors influencing the rainfall variabilities.

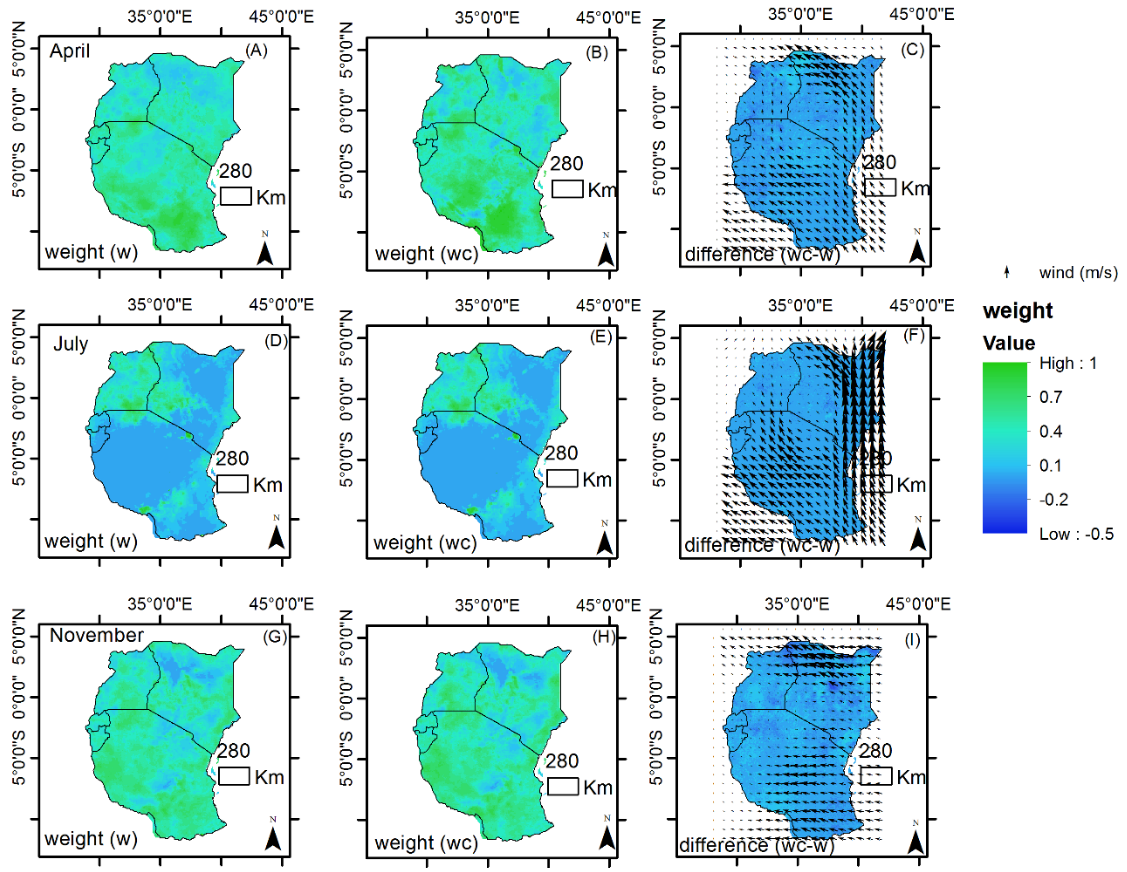

Figure 9: Spatial distributions of raingauge correction weights before (w) and after (wc) exclusion of El Niño years (1997/1998) during training in the three rainfall season months of April, July and November. The weights are derived from the variances of the rain gauge data and those of the corresponding satellite rainfall estimates. Mean (1981-2013) wind patterns are embedded on the maps and the arrows point in the direction of wind flow and the size of the arrow represents the wind speed.

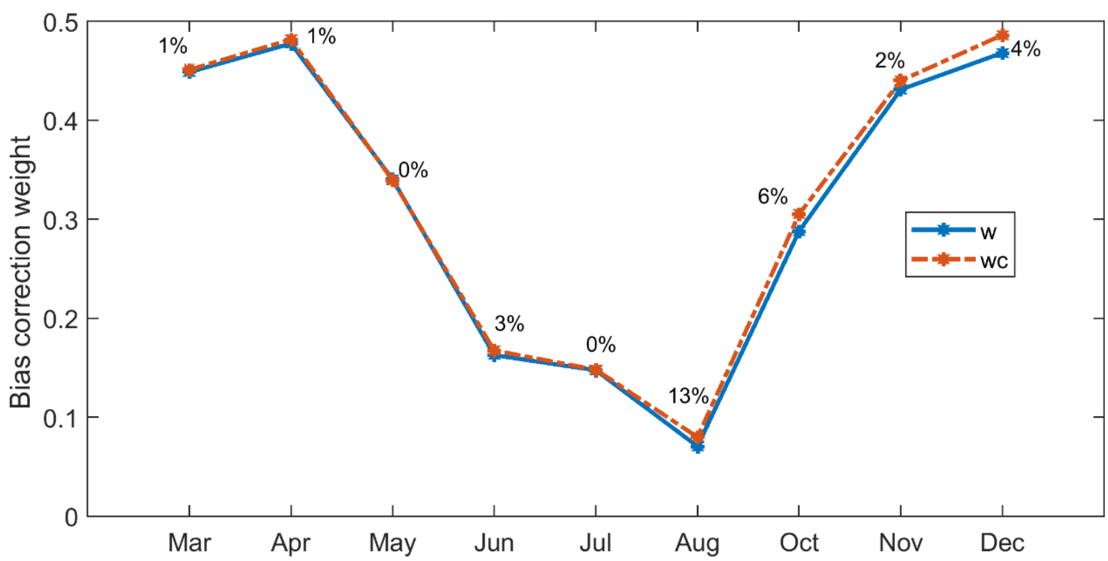

Figure 10: Temporal changes of bias correction weights before (w) and after (w c) exclusion of El Niño years (1997/1998) during training in the three rainfall season months of April, July and November. The weights are derived from the variances of the rain gauge data and those of the corresponding satellite rainfall estimates.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2018 doi:10.20944/preprints201804.0225.v1

Peer-reviewed version available at Remote Sens. 2018, 10, 1074; doi:10.3390/rs10071074

5. Conclusions

Rain gauge distributions are on the decline and satellite rainfall estimates are increasingly used to complement the sparse rain gauge data. However, rain gauge data is used as reference data to validate the incoming satellite products but with the decreasing trend, the quality of this validation may be compromised. To solve this problem Bayesian approach is hereby applied to the existing historical rain gauge information to create a consistent satellite rainfall data that can be used for climate studies. A long term temporal scale (monthly) bias correction is therefore applied on Climate Hazards Group Infrared Precipitation with Stations (CHIRPS v2) data to reduce systematic errors using a corresponding gridded (0.05°) rain gauge data over East Africa. The gridded rain gauge data was developed by ICPAC intergovernmental organization to safeguard the decreasing network. The choice of CHIRPS was based on its close correspondence with rain gauge data over the region [6] and its period of long coverage suitable for climate studies. Satellite rainfall products exhibit systematic errors and although CHIRPS show good performance over East Africa, reducing these errors would increase its performances for climate studies, and agricultural and water management. The study aimed at temporally and spatially evaluating how CHIRPS rainfall estimates compare with the gridded rain gauge data in magnitude and distributions after bias correction. Only the wet rainfall months of MAM, JJA and OND are utilized for a period of 33 years of which 22 years were for calibration and 11 years for validation.

Monthly analysis showed CHIRPS estimates have systematic errors mainly of underestimations and application of the Bayesian method adequately reduced such errors. The remaining errors after correction showed dependence on rainfall magnitudes and hence increased with increase in rainfall amounts. The highest change in correlation coefficients of 32% in April, and 25% in November which are the peak rainfall months of MAM and OND rainfall seasons were observed. Cumulative distributions plots revealed that in areas of low rainfall (<200mm/month) the corrected errors were associated with overcorrections beyond which underestimations were dominant. Remarkably the approach significantly reduced these errors except in August when the presence of a different rainfall regime produced irregular errors in rain gauge data that were observed as overcorrections.

The corrected CHIRPS estimates showed dependence on seasons and this was indicated by patterns of reduction in RMSD and MAE that were highest during JJA rainfall season. Consequently, a reduction of RMSD (50%), and MAE (60%) were observed in the month of June. These changes coincide with the onset of south-east monsoon that peak in the month of May of MAM season and includes JJA rainfall season. This shows the bias corrected CHIRPS follow the rainfall systems affecting rainfall variabilities over East Africa. The overall monthly RMSD and MAE are reduced between (26-48) % and (28-60) % respectively and correlations increase between 2-32 %. It can be concluded that the Bayesian approach reduced CHIRPS errors of monthly scale which were locally induced, like topographic and lake processes.

The areas of highest systematic error reductions include the high elevated areas especially Mt Kenya, Mt Elgon and Lake Victoria region. These areas are also significant because of the mixed rainfall regimes influenced by local effects. Performances of CHIRPS and CHIRPS (bc) with respect to rain gauge data were further compared over these areas (Figure1) during the wettest (April and November) and driest (July) months of MAM, OND and JJA rainfall seasons respectively. The results showed that even though the areas differ in the amount of rainfall in each season, the bias correction aligned the CHIRPS cumulative distribution closer to that of the rain gauge data. However, on highly elevated areas of Mt Kenya, the cumulative distributions of the rain gauge data and CHIRP (bc) were not well aligned. This was more evident during April and November, of MAM and OND rainfall seasons and this was associated with largescale influences as the two seasons are experienced during ITCZ overpass.

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

Interannual CHIRPS bias correction assessments were carried out with respect to rain gauge data for the months of April, July and November to represent MAM, JJA and OND rainfall seasons. Spatial rainfall patterns revealed that systematic errors exist in yearly estimates mainly of underestimation which increased with increase in rainfall magnitudes. However cumulative distribution analysis showed CHIRPS estimates bias correction adjusted underestimations/overestimations of low/high rainfall amounts. Underestimation was more during extreme wet years (2004, 2006 and 2011), which were shown in error statistics with highest rainfall amounts. However, the bias correction adjusted the CHIRPS estimates to align with rain gauge data. Similar to monthly analysis, MAM interannual systematic errors were mainly related to rainfall magnitudes and rainfall distributions and were effectively reduced.

During JJA and OND rainfall seasons CHIRPS systematic errors associated with rainfall magnitude and distribution were reduced. However, over corrections were observed in extreme wet and dry years which were attributed to irregularities of rainfall patterns that were not well captured by the rain gauge data. The overall monthly statistics measures a reduction of RMSD (29-56) % and MAE (28-60) % and an increase of correlations (2-32) %. For yearly analysis they are reductions of RMSD (9-23) %, and MAE (7-27) % and increase of correlations (4-77) % for MAM months, reduction of RMSD (15-35) % and MAE (16-41) % and increase in correlations (5-16)% for JJA months, and reduction of RMSD (3-35) % and MAE (9-32) % and increase of correlations (3-65)% for OND months.

The impacts of largescale phenomena on systematic errors in CHIRPS were assessed by exclusion of known anomalous wet years (El Niño) of 1997/1998 over East Africa. The result showed a significant spatial reduction of these errors more evident from the month of August (13%) and OND (between 2-6%) rainfall months. Minimal (1%) impacts were observed during MAM rainfall months and were observed mainly in areas around Lake Victoria, Eastern Kenya and southwest Tanzania. These are areas associated with large-scale circulation of ENSO (Lake Victoria) and low-level Turkana jet (Eastern Kenya). In conclusion, the bias corrected CHIRPS estimates are more representative of the rainfall magnitude and distribution. Further, these data and the associated errors are more informative of local and large-scale influences over East Africa and can, therefore, be used for climate studies. The approach is recommended for other areas and to other products. However, prior long-term analysis is advised to exclude anomalous years and correct them separately. The approach is suitable for long-term data correction where consistencies of errors can be assumed.

Supplementary Materials: Bayesian Bias Corrected Rainfall generated data in this study.

Acknowledgments: This work was made possible through the funding provided by the Netherlands Fellowship Programmes (NFP), and we greatly appreciate their support. The University provided funds for open access.

Author Contributions: All three authors contributed in concept building and manuscript preparation. Kimani analyzed the data and drafted the manuscript. Su and Hoedjes provided conceptual advice and contributed to the overall writing and approval of the final manuscript.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Kidd, C.; Bauer, P.; Turk, J.; Huffman, G.J.; Joyce, R.; Hsu, K.L.; Braithwaite, D. Intercomparison of high-resolution precipitation products over northwest europe. *J. Hydrometeorol.* **2012**, *13*, 67-83.

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 1074; doi:10.3390/rs10071074

2. Sorooshian, S.; Hsu, K.L.; Gao, X.; Gupta, H.V.; Imam, B.; Braithwaite, D. Evaluation of persiann system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteorol. Soc.* **2000**, *81*, 2035-2046.
3. Feidas, H.; Lagouvardos, K.; Kotroni, V.; Cartalis, C. Application of three satellite techniques in support of precipitation forecasts of a nwp model. *Int J Remote Sens* **2005**, *26*, 5393-5417.
4. Maidment, R.I.; Grimes, D.I.F.; Allan, R.P.; Greatrex, H.; Rojas, O.; Leo, O. Evaluation of satellite-based and model re-analysis rainfall estimates for uganda. *Meteorological Applications* **2013**, *20*, 308-317.
5. Tote, C.; Patricio, D.; Boogaard, H.; van der Wijngaart, R.; Tarnavsky, E.; Funk, C. Evaluation of satellite rainfall estimates for drought and flood monitoring in mozambique. *Remote Sensing* **2015**, *7*, 1758-1776.
6. Kimani, M.W.; Hoedjes, J.C.B.; Su, Z.B. An assessment of satellite-derived rainfall products relative to ground observations over east africa. *Remote Sensing* **2017**, *9*.
7. Abera, W.; Brocca, L.; Rigon, R. Comparative evaluation of different satellite rainfall estimation products and bias correction in the upper blue nile (ubn) basin. *Atmospheric Research* **2016**, *178*, 471-483.
8. AghaKouchak, A.; Mehran, A.; Norouzi, H.; Behrangi, A. Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters* **2012**, *39*.
9. Mateus, P.; Borma, L.S.; da Silva, R.D.; Nico, G.; Catalao, J. Assessment of two techniques to merge ground-based and trmm rainfall measurements: A case study about brazilian amazon rainforest. *Giscience & Remote Sensing* **2016**, *53*, 689-706.
10. Maurer, E.P.; Pierce, D.W. Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences* **2014**, *18*, 915-925.
11. Cressman, G.P. An operational objective analysis system. *Mon.Weather Rev* **1959**, 367 – 374.
12. Vila, D.A.; de Goncalves, L.G.G.; Toll, D.L.; Rozante, J.R. Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental south america. *J. Hydrometeorol.* **2009**, *10*, 533-543.
13. Tian, Y.D.; Peters-Lidard, C.D.; Eylander, J.B. Real-time bias reduction for satellite-based precipitation estimates. *J. Hydrometeorol.* **2010**, *11*, 1275-1285.
14. Stokstad, E. Hydrology - scarcity of rain, stream gauges threatens forecasts. *Science* **1999**, *285*, 1199-1200.
15. Funk, C.; Hoell, A.; Shukla, S.; Blade, I.; Liebmann, B.; Roberts, J.B.; Robertson, F.R.; Husak, G. Predicting east african spring droughts using pacific and indian ocean sea surface temperature indices. *Hydrology and Earth System Sciences* **2014**, *18*, 4965-4978.
16. Funk, C.C.; Peterson, P.J.; Landsfeld, M.F.; Pedreros, D.H.; Verdin, J.P.; Rowland, J.D.; Romero, B.E.; Husak, G.J.; Michaelsen, J.C.; Verdin, A.P. A quasi-global precipitation time series for drought monitoring: U.S. Geological survey data series 832, 4 p. **2014**.
17. Suwendi, A.; Allebach, J.P. Nearest-neighbor and bilinear resampling factor estimation to detect blockiness or blurriness of an image. *Journal of Electronic Imaging* **2008**, *17*.
18. Carlin, B.P.; Louis, T.A. Bayes and empirical bayes methods for data analysis, chapman and hall, london, uk. 1996; p 399 pp.

19. Talagrand, O. Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II* **1997**, *75*, 191-209.
20. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research-Atmospheres* **2001**, *106*, 7183-7192.