Chapter

Assessment of the CHIRPS-Based Satellite Precipitation Estimates

Franklin Paredes-Trejo, Humberto Alves Barbosa, Tumuluru Venkata Lakshmi Kumar, Manoj Kumar Thakur and Catarina de Oliveira Buriti

Abstract

At present, satellite rainfall products, such as the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) product, have become an alternative source of rainfall data for regions where rain gauge stations are sparse, e.g., Northeast Brazil (NEB). In this study, continuous scores (i.e., Pearson's correlation coefficient, R; percentage bias, PBIAS; and unbiased root mean square error, ubRMSE) and categorical scores (i.e., probability of detection, POD; false alarm ratio, FAR; and threat score, TS) were used to assess the CHIRPS rainfall estimates against ground-based observations on a pixel-to-station basis, during 01 January 1981 to 30 June 2019 over NEB. Results showed that CHIRPS exhibits better performance in inland regions (R, PBIAS, and ubRMSE median: 0.51, −3.71%, and 9.20 mm/day; POD, FAR, and TS median: 0.59, 0.44, and 0.40, respectively) than near the coast (R, PBIAS, and ubRMSE median: 0.36, −5.66%, and 12.43 mm/day; POD, FAR, and TS median: 0.32, 0.42, and 0.26, respectively). It shows better performance in the wettest months (i.e., DJF) than in the driest months (i.e., JJA) and is sensitive to both the warm-top stratiform cloud systems and the sub-cloud evaporation processes. Overall, the CHIRPS rainfall data set could be used for some operational purposes in NEB.

Keywords: CHIRPS, Northeast Brazil, satellite rainfall, rainfall, remote sensing, rain gauge, ground-based validation

1. Introduction

Rainfall is a key component of the global water cycle and is essential for a wide range of applications such as crop modeling, hydrometeorology, water resource management, flood and drought monitoring, and climatological applications [1–3]. Accurate and consistent rainfall estimates are also of remarkable importance for the drought-prone regions, such as the semiarid region of Northeast Brazil (NEB), which is at high risk of food insecurity due to the occurrence of prolonged droughts whose impacts affect adversely their water resources and crop production [4–6].

Nowadays, the measurement of precipitation is based on rain gauge stations, meteorological radars, and satellite retrievals [7, 8]. Rainfall data from ground stations provide high accuracy [9], but they are limited in spatial coverage [10]. Meteorological radars suffer from reduced data quality owing to signal blockage or

distortion [11]. Satellites can be used for sensing large regions with a high temporal and spatial resolution, though satellite retrieval approaches are prone to biases and systematic errors [12]. Consequently, satellite-based rainfall estimates must be validated against rain gauge data in order to assess their uncertainties before being used [13, 14].

In NEB, despite the efforts of the state climate agencies (e.g., National Center for Monitoring and Early Warning of Natural Disasters, CEMADEN; National Institute of Meteorology, INMET; Meteorology and Hydrologic Resources Foundation of Ceara, FUNCEME; Superintendence for the Development of the Northeast, SUDENE; and National Water Agency, ANA), most of the rain gauge networks currently available are inadequate to produce reliable rainfall analysis, because of their scarce spatial coverage, high proportion of missing data, and short-length records [15]. To overcome these limitations, there is a wide variety of satellite-based rainfall products, such as the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS).

CHIRPS is a quasi-global rainfall data set with relatively high spatial resolution (°0.05 × °0.05) and long-term temporal coverage (from 1981 to near real time), whose processing chain blends satellite and gauge rainfall estimates [16]. Since early 2014, CHIRPS rainfall estimations are disseminated with different temporal scales (monthly, 10-day, 5-day, and daily time steps) by the University of California at Santa Barbara (UCSB). It has been subjected to various assessments worldwide by comparing to gauge measurements. According to these studies, the CHIRPS rainfall data set performs relatively well at both a regional and global scale, mainly in terms of bias and the Pearson's correlation coefficient when compared to other state-of-the-art satellite rainfall products [1, 8, 17–21].

Unlike other natural regions, very few studies have been carried out to validate CHIRPS rainfall estimates in NEB. Overall, CHIRPS achieves better results during the rainy season (i.e., March to May), but its ability for the rain detection is poor [22]. Moreover, CHIRPS displays a rainfall pattern similar to the rain gauge data in the south-southeast subregion of the NEB, even though some performance scores are lower than the ones derived from the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) 3B42V7 product, particularly from 2012 to 2014 [23]. Interestingly, CHIRPS provides performance better in terms of rain amount than the Multi-Source Weighted-Ensemble Precipitation (MSWEP), SM2RAIN-CCI (Climate Change Initiative), and Climate Prediction Center Morphing Technique (CMORPH) rainfall products over the Cerrado biome of NEB [24]. These findings are promising for operational applications in NEB (e.g., remote drought monitoring). Nevertheless, to our knowledge, a study investigating the performance of the CHIRPS rainfall data set by using new available ground-based observations is still absent.

The purpose of this study is to evaluate the quality of the CHIRPS rainfall estimates in NEB by considering the newest in situ data from the INMET meteorological stations, which is used as a benchmark rainfall data set over a 39-year period (1981–2019).

2. Materials and methods

2.1 Study area

The study was carried out in NEB (~8,515,759 km^2), which is located between 5.2° N–33.7° S and 34.7°–48.7° W [25]. In this region, the annual precipitation decreases from the east and northeast coast (>1500 mm/year) to inland dry regions

(<500 mm/year) [22], due to the impact of the orography [26] and the influence of different meteorological systems, such as the intertropical convergence zone (ITCZ), squall lines (SL), easterly wave disturbances (EWD), upper tropospheric cyclonic vortices (UTCV), frontal systems (FS), mesoscale convective complexes (MCC), and the South Atlantic convergence zone (SACZ) [27]. The rainy season occurs at different times of the year: April to June in the eastern coast of the NEB; November to January in the southern part of the NEB; and March to May in the semiarid northwestern part of the NEB [27]. This region includes two main river basins, namely, the basins of the São Francisco River (where the Sobradinho reservoir is located) and the Parnaíba River. It also contains the Amazonia, Cerrado, Atlantic Forest, and Caatinga biomes, which are strongly related to the spatial distribution of rainfall regimes [6, 15].

2.2 Rainfall data sets

Daily rain gauge observations from rain gauge stations were provided by the INMET (www.inmet.gov.br). The higher values than daily mean ± 3.5 standard deviations (method for detection of outliers) were coded as missing data [20]. The daily rainfall time series with more than 25% missing data per month were omitted [22]. A number of 27 stations were selected with these criterions (temporal coverage: January 1981 to June 2019). It is worth mentioning that 77%, 62%, and 42% of these stations were used in the blending process of CHIRPS during 1981–1998, 1999–2013, and 2014–2019, respectively (see https://bit.ly/2ZZFAvA); therefore, this sample is not a completely independent data set [13]. As depicted in **Figure 1**, most stations are located in the northwest NEB or near the coast.

CHIRPS rainfall estimates were obtained from the UCSB-Climate Hazards Group (CHG) webpage (https://www.chc.ucsb.edu/data; version 2 released in February 2015) at a daily time scale and spatial resolution of 0.05°, starting 1 January 1981 to 30 June 2019. This rainfall product uses a three-step development process. First, infrared precipitation (IRP) pentad (5-day) rainfall estimates are created from satellite data using cold cloud durations (CCD) lower than 235 K as a threshold value and calibrated in relation to the TRMM 3B42-based precipitation pentads by local regression. Then, the IRP pentads are divided by its long-term IRP mean values to present a percent of normal. Second, the percent of normal IRP pentad is multiplied by the corresponding Climate Hazards Precipitation

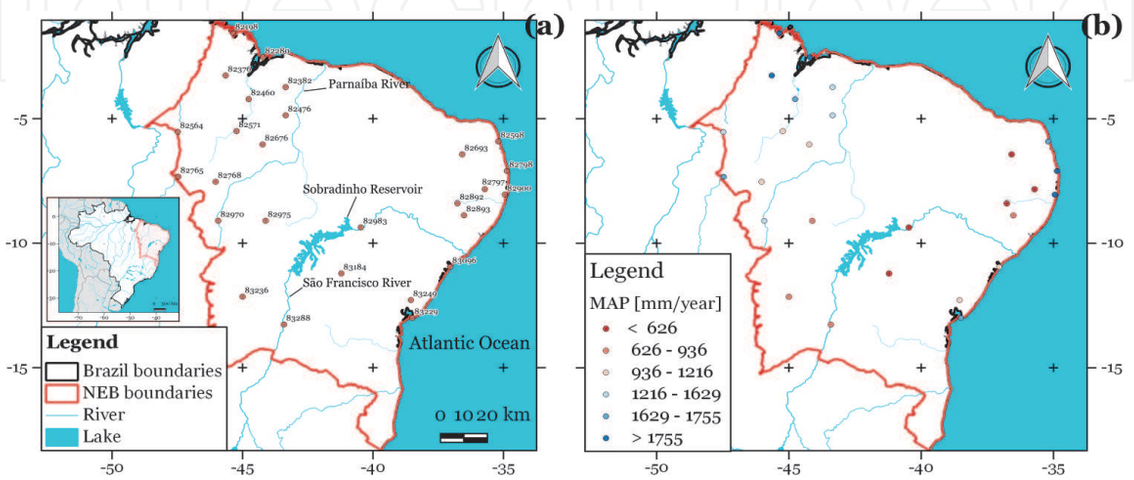

Figure 1.
Geographical location of the study area showing (a) selected stations. The numbers indicate the World Meteorological Organization (WMO) serial of each station; (b) annual mean precipitation for selected stations from 1 January 1981 to 30 June 2019.

Climatology (CHPclim) pentad to generate an unbiased rainfall estimate, with units of millimeters per pentad, called the CHG IR Precipitation (CHIRP). Third, pentadal CHIRP values are disaggregated to daily precipitation estimates based on daily NOAA Climate Forecast System (CFS) fields rescaled to 0.05° resolution. Finally, CHIRPS is produced through blending stations with the CHIRP data sets via a modified inverse distance-weighted algorithm [8]. For more details about the CHIRPS data set, the reader is referred to Funk et al. [16].

2.3 Auxiliary data sets

The land cover, annual rainfall, elevation, and type of climate were used as auxiliary information. The land cover was derived from the Land Cover-Climate Change Initiative (LC-CCI) product [28] (available online at http://maps.elie.ucl.ac.be). The average annual rainfall was estimated from the selected stations. The gauge elevation was obtained from the metadata information at each station. The slope and aspect of the terrain were derived from the Shuttle Radar Topographic Mission (SRTM) (available online at https://earthexplorer.usgs.gov). The type of climate was extracted from the Köppen-Geiger climate classification developed by Beck et al. [29] (available online at https://bit.ly/2Zt90Bu).

2.4 Methodology

The methodology applied in this study is summarized in **Figure 2**. The CHIRPS rainfall data set was chosen because of its low latency (about 3 weeks), high spatial resolution (0.05° × 0.05°), daily temporal resolution, and long-term temporal coverage (1981 to near real time), respectively, so it is potentially suitable for operational purposes in NEB. Firstly, the CHIRPS product was clipped using a shapefile of NEB as a mask. Then, CHIRPS rainfall estimates were extracted using the nearest neighbor (NN) method to generate a paired rainfall data from 1 January 1981 to 30 June 2019 (i.e., the common temporal coverage). The rationale behind the choice of

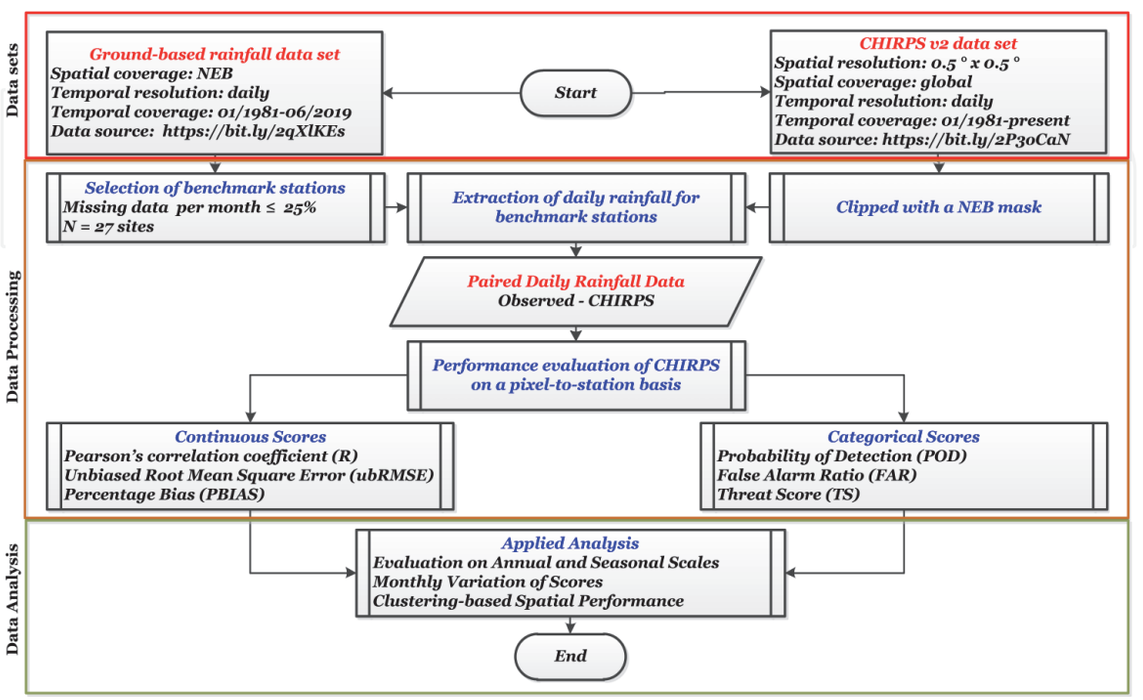

Figure 2.
Simplified flowchart of the methodology used in this study.

the NN method instead of gridded ground-based rainfall data (e.g., via spatial interpolation) is related to the fact that the latter would involve large uncertainties given the lack of a high-density rain gauge network to reproduce adequately the rainfall gradients in NEB [22]. Secondly, an intercomparison of both rainfall data sets was carried out in order to explore the performance of the CHIRPS product at the monthly, seasonal, and annual time scales during the common temporal coverage. Consequently, several metrics on a pixel-to-station basis were computed. The Pearson's correlation coefficient (R), unbiased root mean square error (ubRMSE), and percentage bias (PBIAS) were used as continuous scores. R measures the linear relationship strength between estimations and observations, while ubRMSE and B scores measure how the value of estimates differs from the observed values [20]. To examine the rain detection capability of the CHIRPS product, the probability of detection (POD), false alarm ratio (FAR), and threat score (TS) were used as categorical scores. POD and FAR indicate the fraction of the observed events that were correctly forecasted and the fraction of the predicted events did not occur, respectively. TS is the fraction between hits to all CHIRPS-based events. The categorical scores were derived from a contingency table using a rainfall threshold of 1 mm/day to discriminate between rain and no-rain event [29] (see **Table 1**). This rainfall threshold was chosen due to its previous use in semiarid regions [22, 23, 30]. Finally, in order to investigate the influence of the rainfall station spatial distribution on the performance scores, a cluster analysis based on the k-medoid algorithm was applied using the score values of all stations as cases. This unsupervised classification technique was implemented because it is not sensitive to outliers and reduces noise [31]. The equations, ranges, and optimal values of the performance scores are outlined in **Table 2**.

	Gauge ≥ threshold	**Gauge < threshold**
CHIRPS ≥ threshold	A	B
CHIRPS < threshold	C	D

Table 1.
Contingency table to estimate categorical scores. A, number of hits; B, number of false alarms; C, number of misses; D, number of correct negatives; threshold, rainfall threshold (1 mm/day).

Name	**Formula**	**Range**	**Perfect score**
Pearson's correlation coefficient	$R = \frac{\sum (G-\overline{G})(S-\overline{S})}{\sqrt{\sum (G-\overline{G})^2}\sqrt{\sum (S-\overline{S})^2}}$	[−1, 1]	1
Root mean square error	$RMSE = \sqrt{\frac{1}{N}(S-G)^2}$	[0, ∞)	0
Percentage bias	$B = 100\frac{\sum (S-G)}{N}$	(−∞, ∞)	0
Unbiased root mean square error	$ubRMSE = \sqrt{RMSE^2 - (B/100)^2}$	[0, ∞)	0
Probability of detection	$POD = \frac{A}{A+C}$	[0, 1]	1
False alarm ratio	$FAR = \frac{B}{A+B}$	[0, 1]	0
Threat score	$TS = \frac{A}{A+B+C}$	[0, 1]	1

Table 2.
Formulas of continuous and categorical scores. G, gauge-based rainfall measurement (mm/day); S, CHIRPS-based rainfall estimate (mm/day); $\overline{G}$ and $\overline{S}$, average for G and S, respectively (mm/day); N, number of data pairs; A, B, and C for POD, FAR, and TS, as per **Table 1**.

3. Results

For clarity, this section is split into three parts: (1) evaluation on annual and seasonal scales; (2) monthly variation of scores; and (3) clustering-based spatial performance.

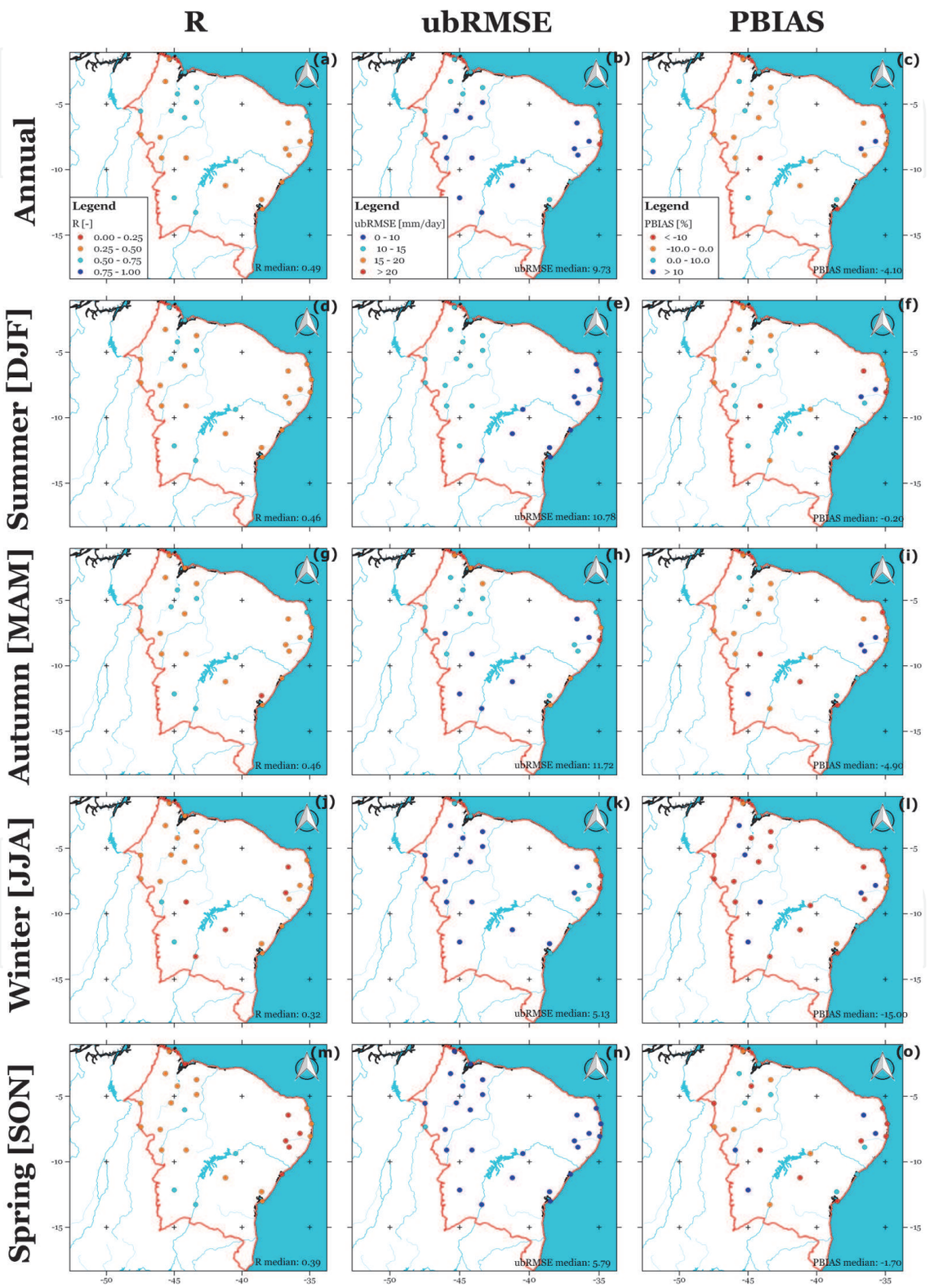

Figure 3.
Spatial distribution of R, ubRMSE, and PBIAS derived from the CHIRPS rainfall estimates against ground observations for (a–c) annual; (d–f) summer; (g–i) autumn; (j–l) winter; and (m–o) spring. The median value of each score is reported.

3.1 Evaluation on annual and seasonal scales

Figure 3 shows the spatial distribution of the continuous scores obtained after the pixel-to-station comparison of the CHIRPS rainfall estimates against the gauge-based data set during the study period. The seasons were defined as summer (Dec-Jan-Feb), autumn (Mar-Apr-May), winter (Jun-Jul-Aug), and spring

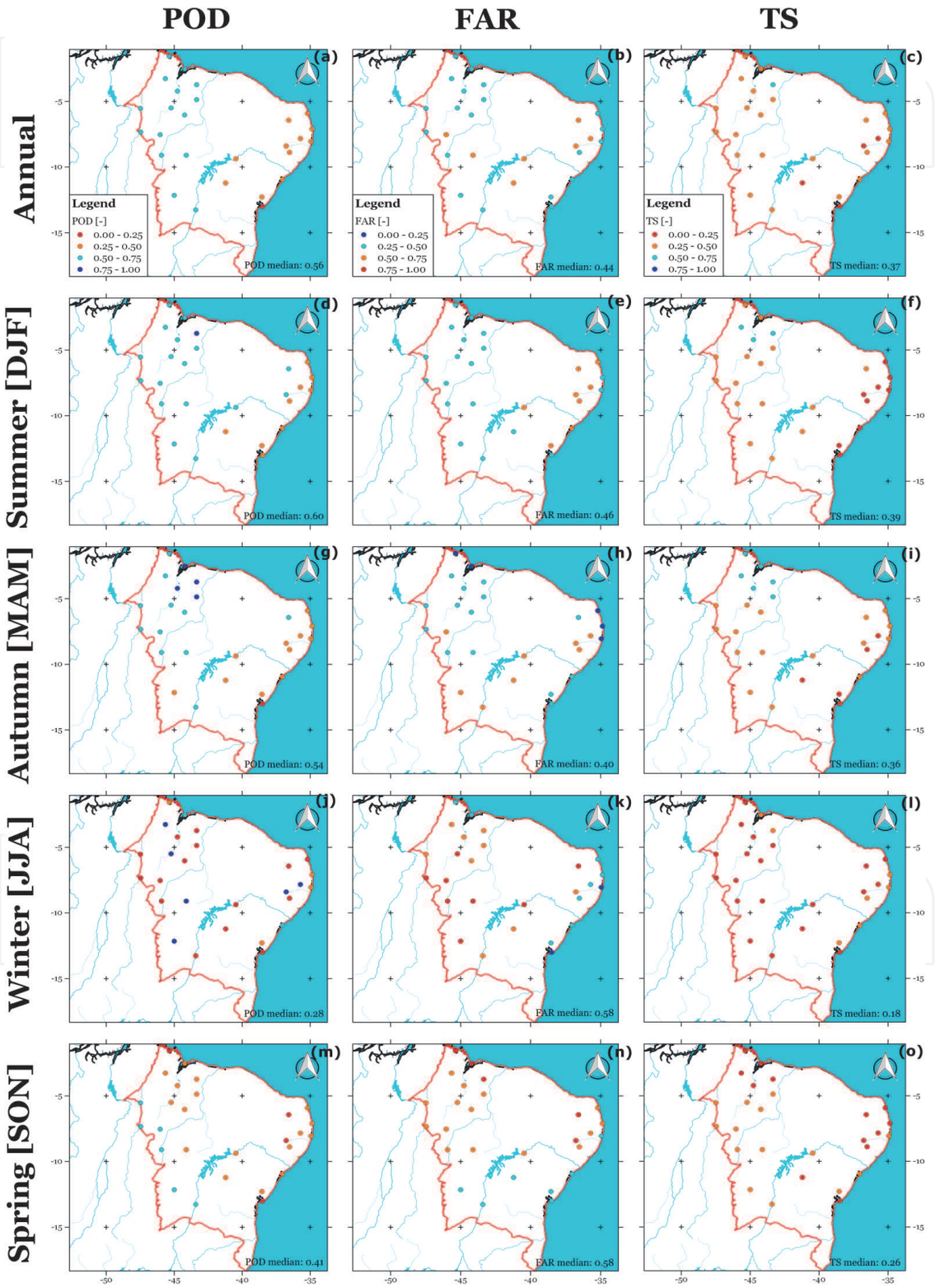

Figure 4.
Spatial distribution of POD, FAR, and TS derived from the CHIRPS rainfall estimates against ground observations for (a–c) annual; (d–f) summer; (g–i) autumn; (j–l) winter; and (m–o) spring. The median value of each score is reported.

(Sep-Oct-Nov) because the NEB is located in the southern hemisphere. The R, ubRMSE, and PBIAS median values listed in each subpanel were obtained by averaging these values from all stations via median to minimize the effects of extreme values. The CHIRPS product showed relatively good agreement with observations in terms of R, ubRMSE, and PBIAS at annual time scale (R median: 0.49; ubRMSE median: 9.73 mm/day; PBIAS, −4.10%), particularly in the northwest NEB ($R > 0.50$, ubRMSE and PBIAS near zero). Interestingly, the R median value begins to decrease from above 0.46 in summer to 0.32 in winter, but it rebounds and increases to values above 0.39 in spring. The ubRMSE values showed a similar pattern, with the higher ubRMSE values in summer and autumn (ubRMSE > 10 mm/day) and lower values in winter and spring (ubRMSE < 6 mm/day). The comparison revealed also that CHIRPS tends to underestimate the amount of rainfall in the course of a year (PBIAS annual median: −4.10%), especially during the transition from summer to winter (PBIAS median from −0.20% to −15.00%).

For the annual time scale, the POD, FAR, and TS mean values were 0.56, 0.44, and 0.37, respectively (**Figure 4**), indicating an acceptable rain detection ability in terms of POD, even though with a medium probability of false alarms in the central NEB. Similar to R and ubRMSE (**Figure 2**), the higher POD and TS values occurred

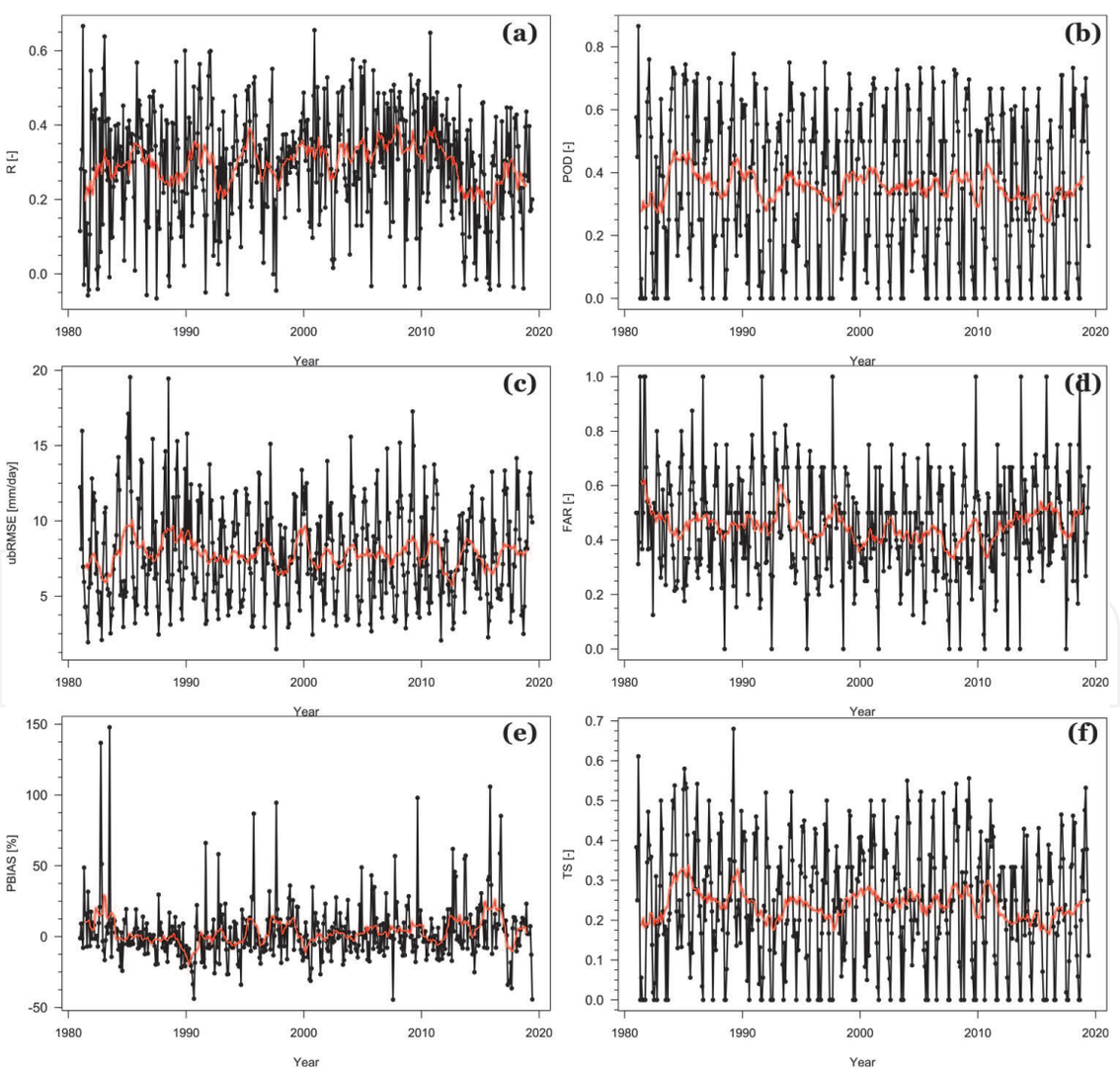

Figure 5.
Monthly time series for (a) R (dimensionless); (b) POD (dimensionless); (c) ubRMSE (mm/day); (d) FAR (dimensionless); (e) PBIAS (%); and (f) TS (dimensionless) derived from the CHIRPS rainfall estimates against ground observations (black line) for all NEB during the period 1981–2019. The red line depicts a 12-month moving average.

in summer and autumn (POD median > 0.50; TS median > 0.30), while lower values were observed in winter and spring. As expected, the FAR exhibited an inverse response to POD throughout the year (i.e., FAR median > 0.55 in winter and spring with lower values in summer and autumn).

3.2 Monthly variation of scores

Figure 5 shows the median of the scores for all stations, months, and years. The median values of R, ubRMSE, and PBIAS ranged between −0.06 and 0.66, 1.48 mm/day and 19.54 mm/day, and −44.50% and 147.80%, respectively. The lowest R values were observed in August (R median: 0.16) and the highest R values in March (R median: 0.41). According to the PBIAS time series, CHIRPS tends to underestimate (overestimate) the amount of rainfall between May and August (September and April), which is consistent with the findings from **Figure 3**. A moderate linear relationship between the monthly averaged values of PBIAS and ubRMSE was also found ($R = -0.35$, p-value < 0.05), suggesting that PBIAS tends to increase when ubRMSE decreases. Furthermore, R, ubRMSE, and PBIAS did not exhibit a long-term trend (not shown for brevity), even though they showed high values for the coefficient of variation (i.e., 51.86%, 41.82%, and 675.49%, respectively).

The temporal variation of POD, FAR, and TR is shown in **Figure 5**. They varied from 0.00 to 0.86, from 0.00 to 1.00, and from 0.00 to 0.68, respectively. The highest POD and TR values were observed in February and March and the lowest in July and August. This means that CHIRPS shows better performance during the rainy season in terms of detection of rain events, which is in line with those inferences obtained from **Figure 4**. Moreover, the lowest FAR values were observed in July and August, indicating a minimum rate of false alarms during the driest

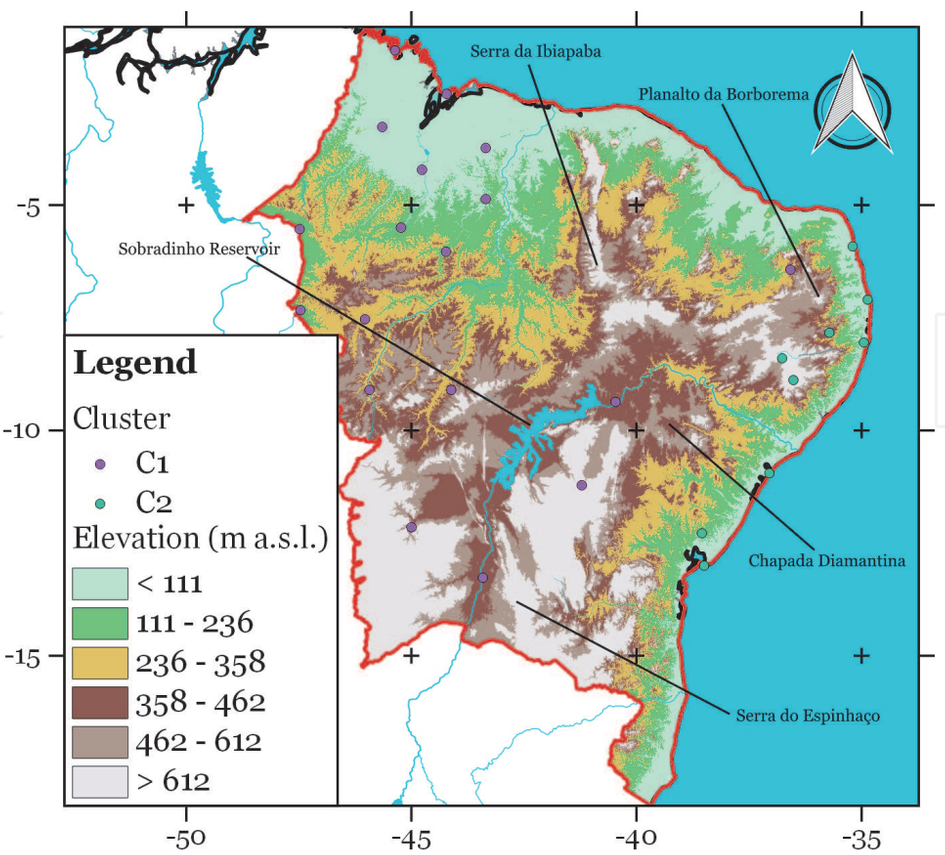

Figure 6.
Clustered stations according to their continuous and categorical scores at annual time scale. A 250-m digital elevation model derived from SRTM images is shown.

months. Similar to the continuous scores, these scores did not exhibit a long-term trend but a high temporal variation (i.e., 64.69%, 42.13%, and 63.97% for POD, FAR, and TR, respectively).

3.3 Clustering-based spatial performance

The previous statistical approaches provide a limited interpretation of the performance of CHIRPS, because they do not offer information about the degree of similarity among the selected stations in terms of their performance scores. Therefore, to identify the similar stations according to their scores, a medoid-based cluster analysis was applied. In order to adequately capture the spatiotemporal variability of the performance scores, an annual time scale was considered (i.e., **Figures 3a–c** and **4a–c**). The spatial distribution of the clustered stations is shown in **Figure 6** (N1, 18 stations; N2, 9 stations), while **Figure 7** displays the performance scores grouped by cluster.

Visual inspection of **Figure 7** reveals that the C1 stations showed the best performance in terms of R, ubRMSE, PBIAS, POD, and TS. The FAR values were similar in both clusters, indicating that CHIRPS tends to forecast false alarms in the

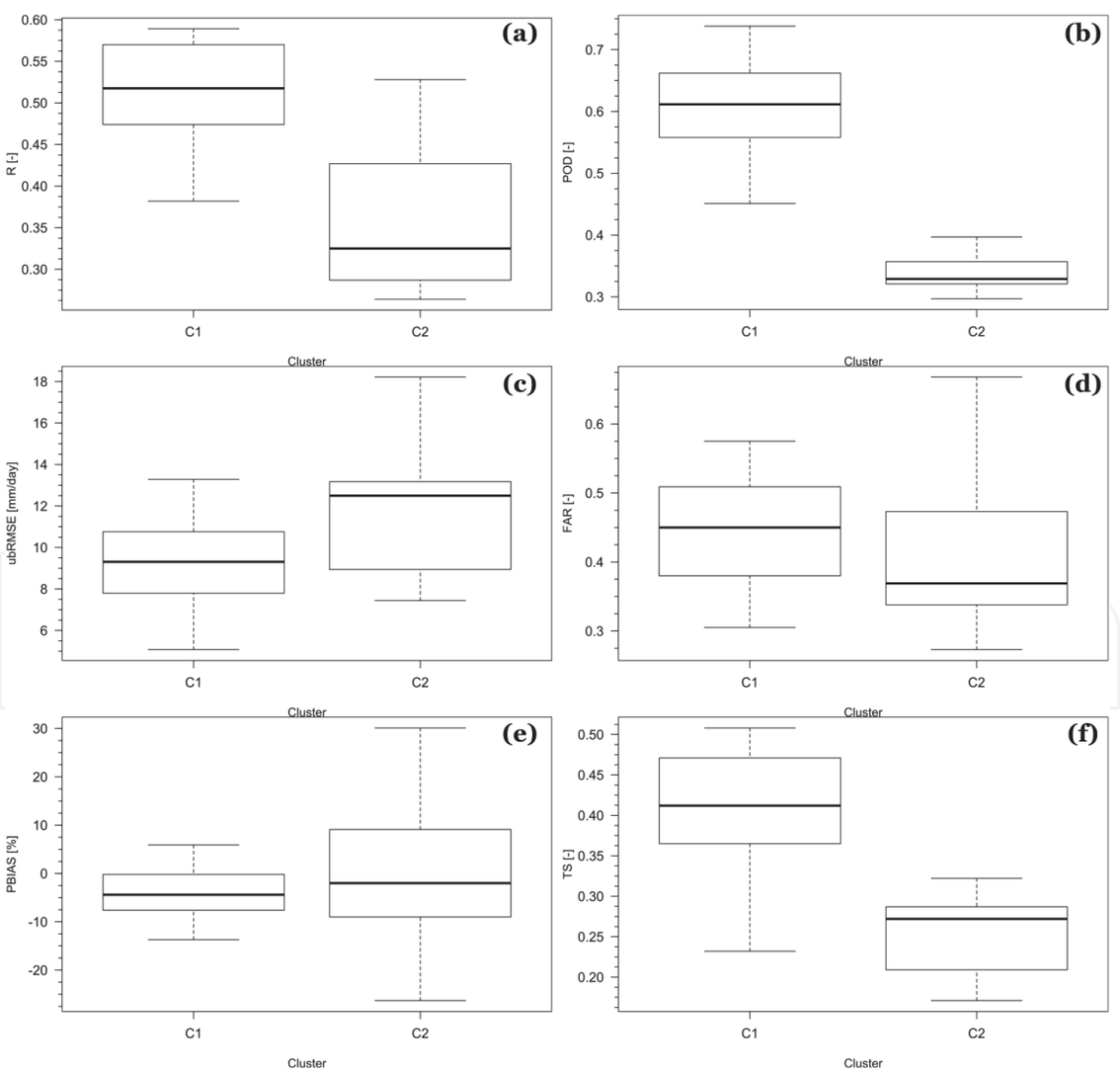

Figure 7.
Boxplots for (a) R (dimensionless); (b) POD (dimensionless); (c) ubRMSE (mm/day); (d) FAR (dimensionless); (e) PBIAS (%); and (f) TS (dimensionless) at annual time scale grouped by cluster, where the thick line depicts the median, while the other horizontal lines of the box depict the maximum, upper quartile, lower quartile, and minimum. For clarity the outliers were omitted.

entire NEB (i.e., CHIRPS estimates to occur a rainfall event, but did not occur), which is also evident in **Figure 4**. It is interesting to note that the C2 stations were mostly concentrated near the coast.

A more detailed comparison, considering the auxiliary data sets (see Section 2.3), showed that there were no significant differences between both clusters in terms of average annual precipitation and terrain elevation (test based on Wilcoxon's t-statistic at the 5% level was used). This means that these local factors did not affect the performance scores. However, regardless of the land cover, most of the C1 stations are located in open flatlands (i.e., terrain slope < 7%) with tropical savanna climate (i.e., Aw), which seem to be favorable surface conditions for better performance of CHIRPS.

4. Discussion

Several performance scores were used to evaluate the CHIRPS rainfall product against gauge observations in Northeast Brazil during the period from January 1981 to June 2019. This region is characterized by large interannual rainfall variations and severe droughts [6, 15]. In line with previous studies [22–24], the CHIRPS data set captured relatively well the spatiotemporal pattern of rainfall across NEB, showing acceptable accuracies (see **Figures 3** and **4**), thanks to the blending process to merge the CHIRP data set derived from IR brightness temperature and TRMM, with ground-based observations [16].

CHIRPS exhibited poorer performance at daily time scale in terms of R (R median: 0.49) than that obtained with monthly time scale (R median: 0.94, reported by Paredes et al. [22]), indicating that increasing temporal aggregation leads to better agreement between CHIRPS and ground-based observations in NEB. This was expected because errors at daily scale time showed closely symmetric characteristics (see **Figure 5**); therefore, they tend to cancel each other during the temporal aggregation [32]. By contrast, this procedure did not provide a significant improvement on the performance in terms of PBIAS (PBIAS median: −4.10% and −3.58% [22] for daily and monthly time scales, respectively), likely due to its high variability at daily time scale (about 700%).

These first results are consistent with the previous findings in other regions with similar climatic features such as South Sudan [33], where CHIRPS became more accurate in terms of R and RMSE as the duration of the integration time increased from months to years. It is important to note, however, that this characteristic is not unique to CHIRPS. Most of the satellite-based rainfall products tend to improve their general performance as the aggregation period increases owing to the effect of cancelation of errors [34, 35].

Overall, CHIRPS showed the best (worst) performance with the (lowest) highest of R and POD and the (highest) lowest bias and FAR during the (driest) wettest months of the year (see **Figures 3** and **4**). This result is consistent with the findings of Paredes-Trejo et al. [24] and Nogueira et al. [23], who found that CHIRPS tends to overestimate low and underestimate high rainfall values in NEB. Likewise, it should be mentioned that the PBIAS and R values were highly sensitive to drought conditions, such as those observed from 2012 to 2015, where CHIRPS showed lower R values (about 0.20) and higher overestimation of the rainfall amount (see **Figure 5a** and **e**). The degradation of the performance under extreme droughts may be attributed to the evaporation processes of raindrops in the dry atmosphere before reaching the surface [20]. In this context, CHIRPS forecasts a rainfall event, but does not occur. According to the equations listed in **Table 2**, this phenomenon leads to higher PBIAS values and near-zero values for R, POD, and TS.

The sub-cloud evaporation plays an important role in the overestimation of rainfall occurrence over different semiarid and arid regions in the world [19, 32, 36]. Therefore, it can help to explain the poor performance of CHIRPS over the driest region of NEB (i.e., the Sertão region), especially in autumn and winter (see **Figures 3** and **4**) and during drought years induced by climate anomalies from the tropical Pacific Ocean (i.e., El Niño-Southern Oscillation) [37]. When this occurs, the air in the lower atmosphere is drier and hotter than usual conditions over the Sertão region [4]. Then, an intensification of the sub-cloud evaporation processes might be expected.

On a seasonal time scale, the reliability of the CHIRPS product was evident in reproducing the seasonal rainfall pattern with results comparable with the ones previously published by Melo et al. [30] for the TRMM 3B42V7 rainfall product, which is its parent rainfall product [16] (see Section 2.2). Similar to TRMM, it was found that CHIRPS exhibits poorer performance over those stations near the coast than the ones located in inland regions of NEB (see **Figures 6** and **7**), particularly in winter (see **Figures 3** and **4**). The reason behind this can be attributed to the prevalence of warm-top stratiform cloud systems along the coastal region [38, 39]. Under these conditions, CHIRPS may not detect rainfall because the cloud tops tend to have a value warmer than the IRP CCD threshold value (i.e., 235 K) [19], leading to a large underestimation in the daily precipitation and poor detection of rainfall events.

As can be seen from **Figure 6**, the landscape at most of the stations is characterized by high topographic complexity, where warm-rain processes induced by orographic lifting are dominant [40, 41]. Similar to the warm-top stratiform cloud systems in the coastal areas mentioned above, CHIRPS has limitations in reproducing the orographic rainfall due to the adoption of a fixed IRP CCD threshold value (i.e., 235 K), leading to classify warm orographic clouds as nonprecipitating [19]. Even though orographic clouds are relatively warm, they can produce substantial amounts of rain [15].

Interestingly, although the number of stations used in the CHIRPS blending process as anchor stations showed a gradual temporal decrease in NEB during the period January 1981 until June 2019 (see https://bit.ly/2ZZFAvA), there was no statistically significant trend in their performance scores (see **Figure 5**). For this study, at least 12, 19, and 21 rain gauges not included as anchor stations for the calculation of CHIRPS rainfall estimations during 1981–1998, 1999–2013, and 2014–2019, respectively, were used. One implication of this situation is that it can be considered a relatively independent validation.

5. Conclusions

The synergetic use of ground-based rainfall observations and satellite-based rainfall estimates is of paramount importance in semiarid regions such as Northeast Brazil. CHIRPS is a state-of-the-art satellite rainfall data set characterized by its blending procedure using thermal infrared satellite observations, TRMM 3B42-based rainfall estimates, monthly precipitation climatology, and atmospheric model rainfall fields from NOAA CFS, with ground-based rainfall measurements [16]. This study set out with the aim of evaluating the performance of CHIRPS against ground-based observations in NEB. The analysis was performed on a pixel-to-station basis at daily time scale and during the period 1981–2019. The major novelty of this study with respect to previous studies [22, 23, 42] is the use of the newest in situ data from the INMET meteorological stations. The main conclusions reached are the following:

1. The CHIRPS rainfall data set exhibits better performance in inland regions with open flatlands than near the coast (see **Figures 6** and **7**).

2. The accuracy of CHIRPS is better in the wettest months (i.e., summer) than in the driest months (i.e., winter) (see **Figures 3** and **4**). In general, CHIRPS underestimates (overestimates) high (low) rainfall amounts.

3. CHIRPS appears to be sensitive to the precipitation from the warm-top stratiform cloud systems (e.g., near to the coast), the warm-rain processes induced by orographic lifting (e.g., the mountain areas of NEB), and the sub-cloud evaporation processes (e.g., the Sertão region). The first and second are mainly attributed to a fixed IRP CCD threshold (i.e., 235 K) used by CHIRPS (see Section 2.2), which may be too cold for regions where the warm-rain processes are dominant [34], while the third is a usual phenomenon in semiarid regions [19].

Based on the abovementioned conclusions, CHIRPS can serve as an alternative source of data for operational applications that require rainfall data, especially over the inland regions of NEB (see the C1 stations in **Figure 6**), during the wettest months of the year (see **Figures 3** and **4**), and at monthly or annual time scales taking advantage of the cancelation of errors of CHIRPS rainfall estimates as the duration of the integration increases [34]. However, future investigations are needed to adequately choose the operational applications of CHIRPS for each sub-region of the NEB.

Acknowledgements

This work was funded by the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq) (Grant no. 88887.091737/2014-01: Edital Pró-Alertas no 24/2014 under project Análise e Previsão dos Fenômenos Hidrometeorológicos Intensos do Leste do Nordeste Brasileiro). We acknowledge to the National Institute of Meteorology (INMET) and the University of California Santa Barbara's Climate Hazards Group (CHG) for providing data that made this study possible.

Conflict of interest

The authors declare no conflict of interest.

Author details

Franklin Paredes-Trejo[1], Humberto Alves Barbosa[2*],
Tumuluru Venkata Lakshmi Kumar[3], Manoj Kumar Thakur[4]
and Catarina de Oliveira Buriti[5]

1 University of the Western Plains Ezequiel Zamora, San Carlos, Venezuela

2 Laboratory for Analyzing and Processing Satellite Images, Federal University of Alagoas, Brazil

3 Atmospheric Science Research Laboratory, Department of Physics, SRM Institute of Science and Technology, Tamilnadu, India

4 Tribhuvan University, Kathmandu, Nepal

5 National Semi-Arid Institute (INSA), Ministry of Science, Technology, Innovations and Communications (MCTIC), Brazil

*Address all correspondence to: barbosa33@gmail.com

References

[1] Beck HE, Vergopolan N, Pan M, Levizzani V, Van Dijk AIJM, Weedon GP, et al. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences. 2017;**21**(12): 6201-6217. DOI: 10.5194/hess-21-6201-2017

[2] Zambrano F, Wardlow B, Tadesse T, Lillo-Saavedra M, Lagos O. Evaluating satellite-derived long-term historical precipitation datasets for drought monitoring in Chile. Atmospheric Research. 2017;**186**:26-42. DOI: 10.1016/j.atmosres.2016.11.006

[3] Funk CC, Peterson PJ, Landsfeld MF, Pedreros DH, Verdin JP, Rowland JD, et al. A quasi-global precipitation time series for drought monitoring. US Geological Survey Data Series. 2014; **832**(4):1-12. DOI: 10.3133/ds832

[4] Marengo JA, Bernasconi M. Regional differences in aridity/drought conditions over Northeast Brazil: Present state and future projections. Climatic Change. 2015;**129**(1–2): 103-115. DOI: 10.1007/s10584-014-1310-1

[5] Brito SSB, Cunha APMA, Cunningham CC, Alvalá RC, Marengo JA, Carvalho MA. Frequency, duration and severity of drought in the Semiarid Northeast Brazil region. International Journal of Climatology. 2018;**38**(2):517-529. DOI: 10.1002/joc.5225

[6] Barbosa HA, Lakshmi Kumar TV. Influence of rainfall variability on the vegetation dynamics over Northeastern Brazil. Journal of Arid Environments. 2016;**124**:377-387. DOI: 10.1016/j.jaridenv.2015.08.015

[7] Michaelides S, Levizzani V, Anagnostou E, Bauer P, Kasparis T, Lane JE. Precipitation: Measurement, remote sensing, climatology and modeling. Atmospheric Research. 2009; **94**(4):512-533. DOI: 10.1016/j.atmosres.2009.08.017

[8] Bai L, Shi C, Li L, Yang Y, Wu J. Accuracy of CHIRPS satellite-rainfall products over mainland China. Remote Sensing. 2018;**10**(3):362. DOI: 10.3390/rs10030362

[9] Kidd C, Becker A, Huffman GJ, Muller CL, Joe P, Skofronick-Jackson G, et al. So, how much of the Earth's surface is covered by rain gauges? Bulletin of the American Meteorological Society. 2017;**98**(1):69-78. DOI: 10.1175/BAMS-D-14-00283.1

[10] Brocca L, Filippucci P, Hahn S, Ciabatta L, Massari C, Camici S, et al. SM2RAIN–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations. Earth System Science Data. 2019;**11**(4):1583-1601. DOI: 10.5194/essd-11-1583-2019

[11] Raghavan S. Radar Meteorology. Dordrecht: Springer; 2013. p. 549. DOI: 10.1007/978-94-017-0201-0

[12] Maggioni V, Sapiano MRP, Adler RF. Estimating uncertainties in high-resolution satellite precipitation products: Systematic or random error? Journal of Hydrometeorology. 2016; **17**(4):1119-1129. DOI: 10.1175/JHM-D-15-0094.1

[13] Loew A, Bell W, Brocca L, Bulgin CE, Burdanowitz J, Calbet X, et al. Validation practices for satellite-based Earth observation data across communities. Reviews of Geophysics. 2017;**55**(3):779-817. DOI: 10.1002/2017RG000562

[14] Kumar TVL, Barbosa HA, Thakur MK, Paredes-Trejo F. Validation

of satellite (TMPA and IMERG) rainfall products with the IMD gridded data sets over monsoon core region of India. In: Rustamov RB, editor. Satellite Information Classification and Interpretation. Rijeka: IntechOpen; 2019. p. 13. DOI: 10.5772/intechopen.77202

[15] Correia Filho WLF, De Oliveira-Júnior JF, De Barros SD, De Bodas Terassi PM, Teodoro PE, De Gois G, et al. Rainfall variability in the Brazilian northeast biomes and their interactions with meteorological systems and ENSO via CHELSA product. Big Earth Data. 2019;**3**(4):315-337. DOI: 10.1080/20964471.2019.1692298

[16] Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Scientific Data. 2015;**2**: 150066. DOI: 10.1038/sdata.2015.66

[17] Paredes Trejo FJ, Barbosa HA, Peñaloza-Murillo MA, Alejandra Moreno M, Farías A. Intercomparison of improved satellite rainfall estimation with CHIRPS gridded product and rain gauge data over Venezuela. Atmosfera. 2016;**29**(4):323-342. DOI: 10.20937/ATM.2016.29.04.04

[18] Prakash S. Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India. Journal of Hydrology. 2019;**571**: 50-59. DOI: 10.1016/j.jhydrol.2019.01.036

[19] Dinku T, Funk C, Peterson P, Maidment R, Tadesse T, Gadain H, et al. Validation of the CHIRPS satellite rainfall estimates over Eastern Africa. Quarterly Journal of the Royal Meteorological Society. 2018;**144**(S1): 292-312. DOI: 10.1002/qj.3244

[20] Rivera JA, Marianetti G, Hinrichs S. Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. Atmospheric Research. 2018; **213**:437-449. DOI: 10.1016/j.atmosres.2018.06.023

[21] Zambrano-Bigiarini M, Nauditt A, Birkel C, Verbist K, Ribbe L. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. Hydrology and Earth System Sciences. 2017;**21**:1295-1320. DOI: 10.5194/hess-21-1295-2017

[22] Paredes F, Barbosa HA, Lakshmi-Kumar T. Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. Journal of Arid Environments. 2016;**139**:26-40. DOI: 10.1016/j.jaridenv.2016.12.009

[23] Nogueira SMC, Moreira MA, Volpato MML. Evaluating precipitation estimates from Eta, TRMM and CHRIPS data in the south-southeast region of Minas Gerais state-Brazil. Remote Sensing. 2018;**10**(2):313. DOI: 10.3390/rs10020313

[24] Paredes-Trejo F, Barbosa H, Rossato L. Assessment of SM2RAIN-derived and state-of-the-art satellite rainfall products over Northeastern Brazil. Remote Sensing. 2018;**10**(7): 1093. DOI: 10.3390/rs10071093

[25] Instituto Brasileiro de Geografia e Estatística. 2010 Census (Censo 2010). [Online] IBGE. Available from: https://bit.ly/2Nufhsu [Accessed: 07 March 2019]

[26] Junquas C, Li L, Vera CS, Le Treut H, Takahashi K. Influence of South America orography on summertime precipitation in Southeastern South America. Climate Dynamics. 2016;**46**(11–12):3941-3963. DOI: 10.1007/s00382-015-2814-8

[27] Molion LCB, Bernardo S. de O. Uma revisão da dinâmica das chuvas no nordeste brasileiro. Revista Brasileira de Meteorologia. 2002;**17**(1):1-10

[28] Houghton RA, Bontemps S, Peng S, Lamarche C, Li W, MacBean N, et al. Gross and net land cover changes based on plant functional types derived from the annual ESA CCI land cover maps. Earth System Science Data Discussions. 2017;**10**(1):1-23. DOI: 10.5194/essd-2017-74

[29] Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data. 2018;**5**: 180214. DOI: 10.1038/sdata.2018.214

[30] Melo D d CD, Xavier AC, Bianchi T, Oliveira PTS, Scanlon BR, Lucas MC, et al. Performance evaluation of rainfall estimates by TRMM multi-satellite precipitation analysis 3B42V6 and V7 over Brazil. Journal of Geophysical Research. 2015;**120**(18):9426-9436. DOI: 10.1002/2015JD023797

[31] Arora P, Varshney S, et al. Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science. 2016;**78**:507-512. DOI: 10.1016/j.procs.2016.02.095

[32] Ayehu GT, Tadesse T, Gessesse B, Dinku T. Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. Atmospheric Measurement Techniques Discussions. 2017;**11**(November):1-24. DOI: 10.5194/amt-2017-294

[33] Basheer M, Elagib NA. Performance of satellite-based and GPCC 7.0 rainfall products in an extremely data-scarce country in the Nile Basin. Atmospheric Research. 2019;**215**:128-140. DOI: 10.1016/j.atmosres.2018.08.028

[34] Belay AS, Fenta AA, Yenehun A, Nigate F, Tilahun SA, Moges MM, et al. Evaluation and application of multi-source satellite rainfall product CHIRPS to assess spatio-temporal rainfall variability on data-sparse western margins of Ethiopian highlands. Remote Sensing. 2019;**11**(22):1-22. DOI: 10.3390/rs11222688

[35] Dembélé M, Zwart SJ. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. International Journal of Remote Sensing. 2016;**37**(17):3995-4014. DOI: 10.1080/01431161.2016.1207258

[36] Saeidizand R, Sabetghadam S, Tarnavsky E, Pierleoni A. Evaluation of CHIRPS rainfall estimates over Iran. Quarterly Journal of the Royal Meteorological Society. 2018;**144**: 282-291. DOI: 10.1002/qj.3342

[37] Marengo JA, Torres RR, Alves LM. Drought in Northeast Brazil—Past, present, and future. Theoretical and Applied Climatology. 2017;**129**(3–4): 1189-1200. DOI: 10.1007/s00704-016-1840-8

[38] Rozante J, Vila D, Barboza Chiquetto J, Fernandes A, Souza AD. Evaluation of TRMM/GPM blended daily products over Brazil. Remote Sensing. 2018;**10**(6):882. DOI: 10.3390/rs10060882

[39] Fedorova N, Levit V, Fedorov D. Fog and stratus formation on the coast of Brazil. Atmospheric Research. 2008; **87**(3–4):268-278. DOI: 10.1016/j.atmosres.2007.11.008

[40] Anders AM, Nesbitt SW. Altitudinal precipitation gradients in the tropics from Tropical Rainfall Measuring Mission (TRMM) precipitation radar. Journal of Hydrometeorology. 2015;**16**(1): 441-448. DOI: 10.1175/JHM-D-14-0178.1

[41] Giovannettone JP, Barros AP. Probing regional orographic controls of precipitation and cloudiness in the central Andes using satellite data. Journal of Hydrometeorology. 2009; **10**(1):167-182. DOI: 10.1175/2008JHM973.1

[42] Paredes-Trejo F, Barbosa H, dos Santos CAC, Paredes-Trejo F, Barbosa H, dos Santos CAC. Evaluation of the performance of SM2RAIN-derived rainfall products over Brazil. Remote Sensing. 2019;**11**(9):1113. DOI: 10.3390/RS11091113