Improving GSMaP V06 Precipitation Products Over the Upper Tocantins River Basin in the Brazilian Cerrado, Based on Local Rain-Gauge Network

Rodrigo Pereira (✉ rodrigomouracbs@gmail.com)

Universidade de Brasília <https://orcid.org/0000-0002-0964-8079>

Vinícius Bof Bufon

Brazilian Agricultural Research Corporation

Felipe Cardoso de Oliveira Maia

Brazilian Agricultural Research Corporation

Research Article

Keywords: satellite observation, rainfall assimilation, Cerrado biome

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1062102/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Theoretical and Applied Climatology on March 8th, 2022. See the published version at <https://doi.org/10.1007/s00704-022-03985-x>.

Abstract

This study aimed to evaluate the performance of GSMaP (Global Satellite Mapping of Precipitation) in estimating rainfall in central Brazil, using the Upper Tocantins River sub-basin as a specific area of analysis. GSMaP data were compared with data from a rain gauge network between 2000 and 2019. Evaluations were made at daily and monthly temporal scales. In general, GSMaP products show an overestimate bias for drizzle ($0.1\sim1\text{ mm day}^{-1}$) and underestimate for rainfalls above 10 mm day^{-1} . The use of monthly scale data significantly reduces the bias observed in the daily scale, but with an underestimation trend of -28.3% and -39.7% for the dry and rainy periods, respectively. Categorical indices showed that the GSMaP system had better hit rates for rain detection in the rainy season (October-April) than in the dry season (May-September). For the studied region, the use of GSMaP data on daily and monthly scales should be preceded by a bias analysis as a function of rain gauge network data. The use of bias coefficient corrected observed rainfall data underestimation on daily and monthly scales, improved correlation between GSMaP and observed rainfall data and reduced errors associated with rainfall network data within the basin influence area.

1 Introduction

Rainfall plays a crucial role in the global energy balance and hydrological cycle, interacting with the hydrosphere, atmosphere, lithosphere, and biosphere (Yuan et al. 2017). Reliable regional- and global-scale rainfall data are crucial for hydrological modeling, water resources management, agriculture, and natural disaster prevention (Germann et al. 2007). However, obtaining reliable rainfall information at regional scales is still a challenge due to its high spatiotemporal distribution heterogeneity (Chen et al. 2019). Although accurate rainfall data can be acquired from rain gauges, their sparse location in developing countries hinders reliable rainfall mapping (Hrachowitz & Weiler, 2011). Remote sensing data can provide information at varied spatiotemporal resolutions on a regional scale, which can be used for systematic mapping of rainfall distribution on the surface of the Earth (Sharifi et al. 2019).

Since the launch of the first atmospheric satellites in the 1960s, many sensors have been developed for monitoring rainfall from space (Sun et al. 2018). They all focus on two main spectral categories: visible and infrared (VIS/IR) onboard geostationary and orbital satellites and low-orbit passive microwave (PMW) sensors (Levizzani et al. 2002). The Global Precipitation Measurement mission (GPM) is a result of a collaboration between the American (NASA) and Japanese (JAXA) space agencies to unify and promote advances in rainfall measurements from an operational constellation of microwave sensors, which have systematically provided global rainfall data on an hourly scale and at different correction levels (Hou et al. 2014).

Google Earth Engine (GEE) is a cloud-based computing platform designed to store and process large sets of atmospheric and earth surface data for analysis and decision making (Kumar & Mutanga, 2018). The platform includes data from several satellites; vectorial, social, demographic, and meteorological data; digital elevation models; and climate data layers (Mutanga & Kumar, 2019). This tool hosts satellite

images and stores them in a public data archive that includes historical Earth images dating back more than forty years (<https://earthengine.google.com/faq/>). Both GSMaP (Global Satellite Mapping of Precipitation) products on the GEE platform were produced with the V06 algorithm of the GPM mission. The GSMaP-Reanalysis product was produced from the resampling of data from several sensors and provides rainfall data for the period between 2000 and 2014. The GSMaP-Operational product includes data from the GPM Core Observatory satellite obtained from March 2014 onwards.

In this sense, this study aimed to evaluate the performance of GSMaP V06 products in the region of the Upper Tocantins River sub-basin, located in central Brazil. The evaluation was carried out by comparing the rainfall estimates from the GSMaP products with rainfall data measured in the rain gauge network within the basin influence area on daily and monthly scales from 2000 to 2019.

2. Study Area

The study was carried out in the area delimited by the Upper Tocantins River basin, located in the state of Goiás, Brazil (between latitudes 13.5° and 16.1° S, and longitudes 47.5° and 50.0° W). The basin has an area of 56344 km² and its main watercourse is the Tocantins River. According to Köppen's climate classification, the region has an *Aw*-type climate, which stands for tropical. The dry season lasts from May to September, while the rainy season is along the remaining months, which account for more than 90% of the average annual rainfall. The average annual pluviometric rainfall is 1800 mm and the average temperature in the winter ranges between 10 and 27°C, while in the summer between 18 and 35°C, sometimes reaching 38°C (Cardoso et al. 2014). The Cerrado is the characteristic biome of the region and native vegetation is characterized by forest, savanna, and grassland formations (Ribeiro & Walter, 1998).

3. Materials And Methods

3.1 Ground data

Daily rainfall data for the period between 2000 and 2019 were obtained from pluviometric stations managed by the National Water Agency (NWA) within the influence area of the basin. The pluviometric stations that had rainfall data in each year of the period evaluated were considered, which resulted in the variation of the annual distribution of the number of user stations, as shown in Figure 2.

3.2 Satellite precipitation products

Products coming from the GSMaP (Global Satellite Mapping of Precipitation) system are made available by the Japan Aerospace Exploration Agency (JAXA - <https://sharaku.eorc.jaxa.jp/GSMaP/index.htm>) for hourly global rainfall data distribution, at a spatial resolution of 0.1° x 0.1° ($\approx$ 10 km). The system uses data from several sensors under the Global Precipitation Measurement mission (GPM), which currently has a constellation of low-orbit satellites that operate passive microwave data, and geostationary and

low-orbit satellites that operate in the infrared range (Hou et al. 2014). In the present study, GSMap daily rainfall data from the GSMap Reanalysis V06 and GSMap Operational V06 products were used, both obtained on the GEE (Google Earth Engine) platform. Rainfall data were evaluated based on the separation between dry and rainy seasons in the Upper Tocantins River basin. Thus, a compartmentalized analysis was performed for the period between 2000 and 2019, separating the data between May and September (dry season) from those between October and April (rainy season).

3.3 Method

3.3.1 Scale handling of rain gauges and pixels for evaluation of rainfall events

To compare GSMap and rainfall data, the point-by-pixel method was used. The GSMap rainfall data was gathered from the rain gauge locations within the basin area to obtain rain gauge-satellite merged data on daily and monthly scales. Rainfall intensity was analyzed by dividing daily data into five categories (Wang et al. 2019): drizzle ($0.1 \sim 1 \text{ mm day}^{-1}$), light ($1 \sim 10 \text{ mm day}^{-1}$), moderate ($10 \sim 25 \text{ mm day}^{-1}$), heavy ($25 \sim 50 \text{ mm day}^{-1}$), and torrential ($>50 \text{ mm day}^{-1}$).

3.3.2 Statistical metrics for evaluating satellite precipitation products

Four statistical indices were used to evaluate continuous estimates from the GSMap products. First, Pearson's correlation coefficient (CC, Eq. 1) reflects the consistency between GSMap data and gauge observations. Second, root means square error (RMSE, Eq. 2) measures the deviation between GSMap data and gauge observations. Then, mean absolute error (MAE, Eq. 3) describes the average difference between GSMap data and gauge observations. The use of MAE and RMSE together is recommended since MAE applies the same weight to all errors, while RMSE penalizes the variance for assigning greater weights to errors with a higher absolute value (Chai & Draxler, 2014). Lastly, percentage bias (PBias, Eq. 4) was used to check for systematic bias, either under or overestimating GSMap data concerning reference data (rain gauges).

$$CC = \frac{\left[\sum_{i=1}^n (S_i - \bar{S})(G_i - \bar{G}) \right]^2}{\sum_{i=1}^n (S_i - \bar{S})^2 \sum_{i=1}^n (G_i - \bar{G})^2}$$

1

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - G_i)^2}$$

2

$$MAE = \frac{1}{n} \sum_{i=1}^n |S_i - G_i|$$

3

$$PBias = \frac{\sum_{i=1}^n (S_i - G_i)}{\sum_{i=1}^n G_i} * 100$$

4

where n is the total number of satellite product and gauge observation data, i is the i^{th} data value of the satellite product and gauge observation, G_i is the gauge observations and G is the average of the gauge observations, S_i and S are satellite estimates and averages, respectively.

A set of categorical metrics was used to assess the ability of GSMap products to describe rain or no rain events (Hossain & Huffman 2008). Probability of detection (POD, Eq. 5) evaluates the satellite hit rates concerning the total rain events observed, penalizing by detection failure. False alarm ratio (FAR, Eq. 6) evaluates the proportion of falsely detected rain events to the total rain events correctly detected by the satellite Frequency of hit (FOH, Eq. 7) measures the frequency at which the satellite detects rain when it occurs, penalizing by the false alarm. Finally, the Heidke skill score (HSS, Eq. 8) measures the accuracy of rainfall estimates, considering detections due to the random effect (Dinku et al. 2008; Mashingia et al. 2014).

$$POD = \frac{a}{a + c}$$

5

$$FAR = \frac{b}{a + b}$$

6

$$FOH = \frac{a}{a + b}$$

7

$$HSS = \frac{2(a \cdot d - b \cdot c)}{(a + b) \cdot (b + d) + (a + c) \cdot (c + d)}$$

8

where a represents the hit (i.e., the event was detected and observed), b represents the false alarm (i.e., the event was detected but not observed), c represents the missing event (i.e., the event was not detected but observed), and d represents the negative hit (i.e., the event was not detected nor observed).

3.3.3 Bias correction procedure for GSMaP rainfall estimate

After comparing the GSMaP and rainfall data, a bias correction method was tested for GSMaP data, using a bias coefficient acquisition (Bias Factor, Eq. 9) and a bias coefficient application (GSMaPcorr, Eq. 10) (Saber & Yilmaz, 2018). The acquisition and application of the bias correction coefficient for GSMaP were applied using observed (rain gauge) and estimated (GSMaP) monthly rainfall during the dry and rainy seasons of the region. To do so, we considered historical series between March/2000 and February/2014 (GSMaP Reanalysis), and between March/2014 and December/2019 (GSMaP Operational).

$$\text{BiasFactor}(T_m) = \frac{\text{GSMaP}(T_m)}{\text{RainGauge}(T_m)}$$

9

$$\text{GSMaP}_{\text{corr}}(P_{(x,y)}, T_i) = \frac{\text{GSMaP}(P_{(x,y)}, T_i)}{\text{BiasFactor}(T_m)}$$

10

where $\text{GSMaP}(T_m)$ and $\text{Rain Gauge}(T_m)$ are the GSMaP-based and rain gauge-based rainfall estimates on a monthly scale with one bias factor calculated for each month within a year, $\text{GSMaP}(P_{(x,y)}, T_i)$ and $\text{GSMaP}_{\text{corr}}(P_{(x,y)}, T_i)$ are the GSMaP data for day T_i at point $P_{(x,y)}$ before and after bias correction procedure, respectively.

3.3.4 Performance of GSMaP corrected products over the Upper Tocantins River Basin

After applying the bias correction to GSMaP daily historical rainfall data, the method was assessed for performance, using the following statistical indices: Kling-Gupta efficiency - KGE (Gupta et al. 2009), which is used to diagnose the relative importance of correlation components, bias, and variability between estimated and reference data; Nash-Sutcliffe efficiency – NSE (Nash & Sutcliffe, 1970), which determines the relative magnitude of residual variance compared to observed data variance; as well as Correlation Coefficient (CC), RMSE, PBias, and MAE.

$$\text{KGE} = 1 - \sqrt{\left[s_r \cdot (r - 1)\right]^2 + \left[s_\alpha \cdot (\alpha - 1)\right]^2 + \left[s_\beta \cdot (\beta - 1)\right]^2}$$

11

$$NSE = 1 - \frac{\sum_{i=1}^n |O_i - P_i|^j}{\sum_{i=1}^n |O_i - B_i|^j}$$

12

where r is the linear correlation coefficient between observed (O_i) and estimated (P_i) data, α is the variability ratio or ratio between estimated and observed value standard deviations (σ_p/σ_o), β is the ratio between estimated and observed means (μ_p/μ_o or $\bar{P}/\bar{O}$), s_r , s_α , and s_β are scaling factors. The weights of scaling factors are conventionally equal, i.e., $s_r = s_\alpha = s_\beta = 1$, B_i is the benchmark series at time-step i . In its original form $j = 2$ and $B_i = \bar{O}$.

4. Results And Discussions

4.1 Precipitated totals evaluation

Figure 3a shows the average number of rainy days for values $\geq 0.5 \text{ mm day}^{-1}$. Figure 3b demonstrates the annual averages of observed (rain gauge network) and estimated (GSMaP-Reanalysis and GSMaP-Operational products) rainfall, for the periods from 2000 to 2014 and from 2014 to 2019, respectively. Considering the number of rainy days, the observed rainfall data were frequently smaller than the GSMaP estimated data throughout the period. However, except for 2007, annual averages between GSMaP-estimated and observed rainfall values showed similar trends. Moreover, the observed rainfall data showed greater variability than did the GSMaP-estimated data, which was expected since rainfall detection from rain gauge are punctual. Conversely, GSMaP estimation integrates the rainfall within a pixel area ($\sim 100 \text{ km}^2$). By comparing annual averages between observed and satellite-estimated rainfall data (Figure 3B), GSMaP data showed accumulated values lower than the observed ones for the entire period evaluated. Still, observed data had greater variability than that estimated by GSMaP.

The number of rainy days estimated by the GPM can be influenced by its algorithms and measurements. Such tools are developed to obtain greater accuracy in instantaneous rainfall estimates, particularly for light rainfalls ($>0.5 \text{ mm h}^{-1}$) (Hou et al., 2014). When evaluating GSMaP over a Cerrado area in Brazil, Salles et al. (2019) pointed out that spatial differences between area (GSMaP grid) and point (rain gauge) scale influence conclusions due to rain gauge locations and distribution. Therefore, rainfall events identified by satellites may not be detected by rain gauges, as it might rain at other locations within the GSMaP grid area.

Regarding annual rainfall values, large-scale rainfalls are commonly underestimated by GSMaP. Deng et al. (2018) analyzed the performance of GSMaP over the Hanjiang River basin in China and found that it can overestimate and underestimate small-scale and large-scale rainfalls, respectively. After integrating the data into annual accumulated, annual averages of observed rainfall in a river basin can be underestimated.

Figure 4 shows the monthly rainfall averages between 2000 and 2019 estimated by the GSMap products and observed in the local rain gauge network. These results show GSMap capacity to reproduce inter-seasonal variations between dry and rainy periods in the region. The months between May and September are marked as the dry season, while from October to April is the rainy season. Just as the annual cumulative, monthly rainfall is underestimated by GSMap concerning the observed in the region, mainly between November and April.

4.2 Quantitative statistics

A comparison of statistical indices (CC, MAE, RMSE, and PBias) between GSMap and rain gauge data demonstrated differences in satellite data performance between dry and rainy seasons (Figure 5). During the dry season, CC values (Figure 5a) had greater variability. Although some CC values can be considered high, it was not verified here since 50% was below 0.4 for the dry seasons from 2000 to 2019. In the rainy season, the variability of correlations between GSMap-estimated and observed data was smaller; yet, 75% of the CC for this period were below 0.5. This difference was already expected since, in dry periods, rains are substantially less frequent and more irregular, in addition to being in lesser quantity or almost zero, compared to rainy periods; consequently, the variability between rainfall observation (rain gauge) and detection (satellite) during this period is greater. During the rainy months in the Cerrado environment, seasonal rainfall is marked by high variability in the dry-rainy and rainy-dry transition months (Nimer, 1989). Furthermore, the punctual observation from the rain gauge network had greater variability compared to GSMap-estimated data. Such variability increases errors when comparing the ground and satellite data (Darand & Siavashi, 2021). The Upper Tocantins River basin lies in Brazil midwest region, and its rainfall regime is governed by atmospheric mechanisms, fostering a regional and homogeneous climatic pattern. However, the local terrain, with altitudes between 325 and 1606 m (Figure 1), can create a certain heterogeneity in rainfall distribution over the basin area. In almost the entire Midwest region, more than 70% of the total rainfall accumulated during the year occurs from November to March. Conversely, the months from June to September are excessively dry, with an average of 4 to 5 days of rainfall per month (Nimer, 1989).

Another factor affecting ground-observed and GSMap-estimated rainfall correlations is the low coverage density of rain gauges, despite their unlimited database between 2000 and 2019. In this regard, Darand & Siavashi, (2021) applied GSMap over Iran to compare regions with different rain gauge densities. These authors found correlations between 0.1 and 0.7 for regions with low rain gauge densities, and between 0.4 and 0.9 (more often above 0.6) for regions with higher densities. Moreover, Hrachowitz & Weiler (2011) found that more sparse rain gauge networks tend to have different deviations, depending on the local rainfall regime. The larger and the more uniform the rainfall events occur within a basin, the smaller the data deviations observed in rain gauges. On the other hand, small summer storms or localized rain events are registered differently by rainfall gauge networks, wherein some events may not be observed, or only detected individually.

Given the variability in the rainfall formation process, studies using different algorithms and sensors have shown that topographic variations and convective rains have a significant influence on satellite rainfall estimation. This can therefore result in unexpected errors between embedded pixel values and gauge point values (Ma et al. 2016; Wang et al. 2019). Our rainfall analysis between 2000 and 2019 may also have been influenced by local atmospheric irregularities, which can cause rainfall to have distinct totals each year and far-from-normal values (Nimer, 1989).

The MAE and RMSE values (Figures 5b, 5c) demonstrate a great difference in data performance between dry and rainy seasons. In the dry season, rainfall events and volumes both for observed (rain gauge) and estimated (GSMaP) values were extremely lower than those in the rainy season. This can cause misinterpretation regarding the GSMaP system performance in the dry season. Darand & Siavashi et al. (2021) found RMSE values from 5 to 15 mm day⁻¹ between March and December of 2018 in Iran. These authors highlighted the use of GSMaP-Gauge product calibrated with data from rain gauges rather than GSMaP-NRT (Near Real-Time) and GSMaP-MVK (research product) products to obtain more accurate estimates.

PBias (Figure 5d) had greater variability and amplitude for comparisons in the dry season. In the rainy season, PBias values were mostly negative, which denotes rainfall underestimation by GSMaP. Ma et al. (2016) evaluated the TRMM 3B42V7 and GPM IMERG rainfall data for the Tibet region and observed BIAS rates between -20 and 80% and between -30 and 70%, respectively, during the rainy seasons between 2014 and 2017. In the Upper Tocantins River basin, between 2000 and 2019, average PBias values for GSMaP data ranged between -80% and 103%, and between -68% and 80% for the dry and rainy periods, respectively.

4.3 Qualitative analysis

Figure 6 shows a set of categorical metrics used in qualitatively evaluating the performance of GSMaP products. In the dry season, POD values (Figure 6a) varied highly, with 50% being below 0.8, and reaching a minimum of 0.3. In the rainy season, POD values are concentrated above 0.7, showing greater efficiency or probability of success in detecting rains between October and April. Such a higher detection probability in the rainy season reduced false alarm rates (FAR) (Figure 6B). Otherwise, the lower performance in terms of POD in the dry period (May to September) promoted higher FAR values. Notably, FAR has a negative effect since the lower the value, the lower the false alarm rate. Therefore, during the rainy season, the sensor could properly detect the rainfall events observed in the rain gauge network (Wilks, 2006). Lastly, FOH values (Figure 6C) showed significant differences in averages between both periods; therefore, the frequency of accurate estimates was higher in the rainy season, corroborating FAR results.

HSS index (Figure 6d) compares a system quality or ability of prediction to other forecasts occurring randomly, in other words, statistically independent from reference observations (rain gauges). A perfect prediction would have an HSS equal to 1, while a random forecast receives 0 (Tuan et al. 2019). As in

other indices, HSS variability in the dry period was greater, but its average was slightly higher than that in the rainy season. In short, only one observation in the rainy season was deemed random based on HSS.

The high variability of categorical indices in both seasons can be attributed to the seasonal effect of rainfall in the region under study. Salles et al. (2019) evaluated IMERG-v5 and GSMaP-v7 rainfall estimates for the Distrito Federal in Brazil, which is near the Upper Tocantins (Figure 1), and observed that both products had the best performances during the rainy season and the worst in the dry one. Moreover, median performances were obtained in intermediate periods between the dry and rainy seasons.

4.4 Performance analysis of monthly and daily GSMaP estimate

Figure 7 displays the accuracy of GSMaP Reanalysis and GSMaP Operational products on daily and monthly scales. After comparing with rain gauge data, rainfall tended to be underestimated on both temporal scales and analyzed periods. PBias values on the daily scale were -53.3% and -52% for the rainy and dry seasons, respectively, and on a monthly scale were -39.7% and -28.3%, respectively. Data fit was smaller on a daily scale than on a monthly scale, in which CC for daily data was 0.42 in the rainy and 0.35 in the dry season. On a monthly scale, CC values were similar between both products (0.86 for the rainy season and 0.85 for the dry season). For other regions of the globe, the best fit of satellite data with rain gauge data on a monthly scale has often been found in comparative studies with other sensors and GPM mission data. Wang et al. (2019) assessed GPM IMERG and TRMM 3B42V7 products and observed CC of 0.85 and 0.92 on a monthly scale, and 0.50 and 0.41 on a daily scale, respectively.

In terms of bias, GSMaP products have shown in previous studies a negative trend, or underestimation, in Bolivia and Distrito Federal, Brazil, -22.4% and -10%, respectively (Satgé et al. 2017; Salles et al. 2019a). To circumvent this problem, a bias correction procedure has been recommended before using a GSMaP product. Boluwade et al. (2017) showed one example of a bias correction method for satellite rainfall estimates.

Satellite rainfall underestimates or overestimates are troublesome in hydrological applications. For instance, problems in streamflow or surface runoff estimates can affect flood predictions or aquifer recharge estimates (Tian et al. 2010). GSMaP products, as well as other satellite rainfall data, require prior comparison with data from rain gauges distributed throughout the land to identify site-specific errors in satellite estimates (Salles et al. 2019b). Sungmin et al. (2017) evaluated the GPM products IMERG Early, Late, and Final over Austria and faced the need for reference data to assess bias, not only for rainfall volume but also its temporal scale acquisition. Such an approach is recommended to achieve the required accuracy in satellite-estimated rainfall data. Another issue with GSMaP products is the difference in sampling rates between infrared (IR) and passive microwave (PMW) input sources used by the system. Chen et al. (2019) found that a significantly different pattern in sampling frequency for the GPM-GSMaP input sources. In short, IR data is more used than PMW sensor data.

Figure 8 shows the probability distributions (PDFv) in terms of daily rainfall volume of rain gauge and GSMaP data. PDFv is a widely used metric for satellite rainfall products and measures accumulated rainfall rates within a rainfall volume ranking (Kirstetter et al. 2012; Tang et al. 2016; Xu et al. 2017). Both products tested here tended to overestimate rainfall for the categories drizzle ($0.1\sim1\text{ mm day}^{-1}$) and light ($1\sim10\text{ mm day}^{-1}$). For Wang et al. (2019), satellite overestimation of low rainfall volumes may be related to evaporation, since small rain droplets do not fall into the atmospheric profile. For rainfall volumes above 10 mm, an inverse behavior was observed, with rainfall events being underestimated by both GSMaP products. Similar behavior was observed by Wang et al. (2019), who used GPM IMERG and TRMM 3B42V7 data and noted an overestimating trend for drizzle data ($0\sim1\text{ mm day}^{-1}$). Concerning heavy ($25\sim50\text{ mm day}^{-1}$) and torrential ($> 50\text{ mm day}^{-1}$) volumes, both GSMaP products showed a low probability of detecting events compared to the rain gauges in the Upper Tocantins River basin. This fact might be related to the distribution of isolated rains within the total pixel area.

4.5 Bias correction for GSMaP rainfall data in the Upper Tocantins River basin.

Previous analysis demonstrated that the GSMaP products underestimate rainfall data in the Upper Tocantins River basin from 2000 and 2019. Therefore, a bias correction procedure was required. We thus applied a multiplicative factor to correct GSMaP data bias and improve comparison statistical indices between GSMaP and rain gauge data (Table 5). Bias correction improved KGE, NSE, CC, RMSE, and MAE of GSMaP data in the dry and rainy seasons. Based on PBias values, one can say that bias correction practically eliminated data underestimation before application (Tables 4, 5). For daily scale (Table 4), correlation (CC) between GSMaP and rain gauge data increased in the dry period, besides substantial KGE and NSE increases. Both RMSE and MAE indexes could also be reduced for the dry period. On the other side, PBias correction and KGE and NSE improvements stood out in the rainy season, while RMSE and MAE were not substantially changed on a daily scale. In Turkey, Saber & Yilmaz (2018) applied a multiplicative correction factor on GSMaP data and obtained R^2 from 0.81 to 0.98, RMSE from 6.97 to 1.21, NSE from 0.57 to 0.99, and PBias from -56.44 to 0.20. These authors assessed these statistical indices, first on an hourly scale after a daily scale, for different rainfall volume thresholds (from 1 to 10 mm day^{-1}). Condom et al. (2011) also used a multiplicative correction model on TRMM data in Peru and obtained greater RMSE reductions compared to additive bias correction models. The results of bias correction on a monthly scale (Table 5) improved almost all indices. Therefore, the method was efficient for GSMaP data in the Upper Tocantins River basin, both during dry and rainy seasons from 2000 to 2019.

Table 4
Statistical analysis comparing original and corrected daily GSMaP data for the dry season (May – September 2000-2019).

Statistical Measures	GSMaP before Correction		GSMaP after Correction	
	Dry	Wet	Dry	Wet
KGE	0.15	0.11	0.74	0.34
NSE	0.09	0.03	0.55	0.14
CC	0.36	0.34	0.76	0.36
RMSE (mm day ⁻¹)	17.70	13.49	4.80	14.54
MAE (mm day ⁻¹)	9.90	7.32	1.01	8.22
PBias (%)	-36.8	-38.6	0.9	0.1

Table 5
Statistical analysis comparing original and corrected monthly GSMaP data for the rainy season (October – April 2000-2019).

Statistical Measures	GSMaP before Correction		GSMaP after Correction	
	Dry	Wet	Dry	Wet
KGE	0.43	0.33	0.82	0.82
NSE	0.42	0.20	0.72	0.72
CC	0.76	0.77	0.85	0.85
RMSE (mm month ⁻¹)	87.77	117.65	61.13	69.85
MAE (mm month ⁻¹)	58.94	83.42	42.64	48.99
PBias (%)	-28.1	-39.5%	0	0.1

Figure 9 shows the scatterplot of monthly rainfall data obtained by GSMaP products, before and after bias correction, for the dry and rainy periods between 2000 and 2019. If compared to its direct use (without bias correction), bias-corrected GSMaP data had a general improvement in terms of correlation with rain gauge data.

Summary And Conclusion

The comparative analysis of GSMaP and rain gauge network rainfall data demonstrates that the system could detect seasonal variation in rainfall in the Upper Tocantins River basin. However, it overestimates the number of rainy days and underestimates the volume of rain when compared to rain gauge data.

Overall, the GSMap system's ability to detect rain events, which was demonstrated by categorical indices (POD, FAR, FOH, and HSS), is superior in the rainy season, between October and April.

Bias correction using Bias coefficient reduced underestimation bias of GSMap historical data in the Upper Tocantins River basin, improved correlations between rainfall data, and reduced errors associated with rainfall observations from 2000 to 2019.

Declarations

Conflict of interest

The authors have no competing interests to declare that are relevant to the content of this article.

Funding Statement

No funding was received for conducting this study.

Author's Contribution

Conceptualization: Rodrigo Pereira, Vinícius Bufon, Felipe Maia; Data curation: Rodrigo Pereira, Felipe Maia; Formal analysis: Rodrigo Pereira, Vinícius Bufon; Writing – original draft: Rodrigo Pereira; Writing – review & editing: Rodrigo Pereira, Vinícius Bufon.

Code availability

Not applicable.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. Corresponding author: Rodrigo Pereira, email: rodrigomouracbs@gmail.com

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Boluwade A, Stadnyk T, Fortin V, Roy G (2017) Assimilation of rainfall Estimates from the Integrated Multi-satellite Retrievals for GPM (IMERG, early Run) in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology: Regional Studies* 14:10-22
2. Cardoso MRD, Marcuso FFN, Barros JR (2014) Classificação climática de Koppen-Geiger para o Estado de Goiás e o Distrito Federal. *Acta Geográfica* 8(16):44-55
3. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7:1247-1250
4. Chen H, Yong B, Gourley JJ, Liu J, Ren L, Wang W, Hong Y, Zhang J (2019) Impact of the crucial geographic and climatic factors on the input source errors of GPM-based global satellite rainfall estimates. *Journal of Hydrology* 575:1-16
5. Darand M, Siavashi Z (2021) An evaluation of Global Satellite Mapping of Precipitation (GSMaP) datasets over Iran. *Meteorology and Atmospheric Physics* 133:911-923
6. Deng P, Zhang M, Guo H, Xu C, Bing J, Jia J (2018) Error analysis and correction of the daily GSMaP products over Hanjiang river basin of China. *Atmospheric Research* 214:121-134
7. Dinku T, Chidzambwa S, Ceccato P, Connor SJ, Ropelewski CF (2008) Validation of high-resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing* 29(14):4097–4110
8. Germann U, Galli G, Boscacci M, Bolliger M (2006) Radar rainfall measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society* 132:1669-1692
9. Hrachowitz M, Weiler M (2011) Uncertainty of Precipitation Estimates Caused by Sparse Gauging Networks in a Small, Mountainous Watershed. *Journal of Hydrologic Engineering* 16:460-471
10. Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, Oki R, Nakamura K, Iguchi T (2014) The global rainfall measurement mission. *Bulletin of the American Meteorological Society* 95(5):701-722
11. Hossain F, Huffman GJ (2008) Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales. *Journal of Hydrometeorology* 9:563-575
12. Kirstetter PE, Hong Y, Gourley JJ, Chen S, Flamig Z, Zhang J, Schwaller M, Petersen W, Amitai E (2012) Toward a framework for systematic error modeling of spaceborne rainfall radar with NOAA/NSSL ground radar-based national mosaic QPE. *Journal of Hydrometeorology* 13(4):1285–1300
13. Kumar L, Mutanga O (2018) Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sensing* 10(10):1-15
14. Levizzani V, Amorati R, Meneguzzo F (2002) A review of satellite-based rainfall estimation methods. ed. ISAC-CNR, Bologna
15. Ma Y, Tang G, Long D, Yong B, Zhong L, Wan W, Hong Y (2016) Similarity and Error Intercomparison of the GPM and Its Predecessor-TRMM Multisatellite Precipitation Analysis Using the Best Available Hourly Gauge Network over the Tibetan Plateau. *Remote Sensing* 8(569):1-17

16. Mashingia F, Mtalo F, Bruen M (2014) Validation of remotely sensed rainfall over major climatic regions in northeast Tanzania. *Physics and Chemistry of the Earth* 67-69:55-63
17. Mutanga O, Kumar L (2019) Google Earth Applications. *Remote Sensing* 11(591):1-4
18. Nimer E (1989) *Climatologia do Brasil*. IBGE – Departamento de Recursos Naturais e Estudos Ambientais, Rio de Janeiro
19. Ribeiro JF, Walter BMT (1998) Fitofisionomias do bioma Cerrado. In: Sano S. M., Almeida, S. P. Eds. *Cerrado: Ambiente e Flora*. Planaltina
20. Saber M, Yilmaz KK (2018) Evaluation and bias correction of satellite-based rainfall estimates for modelling flash floods over the Mediterranean region: Application to Karpuz River Basin, Turkey. *Water* 10(657):1-24
21. Salles L, Satgé F, Roig H, Almeida T, Olivetti D, Ferreira W (2019) Seasonal Effect on Spatial and Temporal Consistency of the New GPM-Based IMERG-v5 and GSMaP-v7 Satellite Precipitation Estimates in Brazil's Central Plateau Region. *Water* 11(4):1-18
22. Satgé F, Xavier A, Zolá R, Hussain R, Timouk F, Garnier J, Bonnet MP (2017) Comparative Assessments of the Latest GPM Mission's Spatially Enhanced Satellite Rainfall Products over the Main Bolivian Watersheds. *Remote Sensing* 9(4):1-16
23. Sharifi E, Saghaian B, Steinacker R (2019) Downscaling Satellite Precipitation Estimates with Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *Journal of Geophysical Research: Atmospheres* 124:789-805
24. Sungmin O, Foelsche U, Kirchengast G, Fuchsberger J, Tan J, Petersen WA (2017) Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using Wegener Net gauge data in southeastern Austria. *Hydrology and Earth System Sciences* 21:6559-6572
25. Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A Review of Global Precipitation Data Sets: DataSources, Estimation, and Intercomparisons. *Review of Geophysics* 56:79-107
26. Tang G, Ma Y, Long D, Zhong L, Hong Y (2016) Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over mainland China at multiple spatiotemporal scales. *Journal of Hydrology* 533:152–167
27. Tian Y, Lidard-Peters CD, Adler RF, Kubota T, Ushio T (2010) Evaluation of GSMaP Precipitation Estimates over the Contiguous United States. *Journal of Hydrometeorology* 11:566-574
28. Tuan-Trinh L, Matsumoto J, Ngo-Duc T, Nodzu MI, Inoue T (2019) Evaluation of satellite rainfall products over Central Vietnam. *Progress in Earth and Planetary Sciences* 6(54):1-16
29. Wang S, Liu J, Wang J, Qiao X, Zhang J (2019) Evaluation of GPM IMERG V05B and TRMM 3B42V7 Precipitation Products over High Mountainous Tributaries in Lhasa with Dense Rain Gauges. *Remote Sensing* 11(2080):1-20
30. Wilks DS (2006) *Statistical methods in the atmospheric sciences*. Academic Press, Burlington
31. Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., Hou, A. 2017. Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on high- density rain gauge network.

32. Yuan F, Zhang L, Win KWW, Ren L, Zhao C, Zhu Y, Jiang S, Liu Y (2017) Assessment of GPM and TRMM Multi-Satellite Precipitation Products in Streamflow Simulations in a Data-Sparse Mountainous Watershed in Myanmar. Remote Sensing 9(302):1-23

Tables

Tables 1-3 are not provided with this version.

Figures

Figure 1

Study area location, distribution of rain gauges, and GSMaP grid.

Figure 2

Number of rain gauges with complete data during the period of 2000-2019 at the Upper Tocantins basin.

Figure 3

Number of rainfall days (A) and rainfall amount per year (B) observed by the rain gauge network and estimated by GSMaP Reanalysis Operational. Error bars represent the standard deviation of the GSMaP and observed data.

Figure 4

Average monthly rainfall recorded by the rain gauge network and estimated by GSMaP for the Upper Tocantins River basin.

Figure 5

Boxplot of correlation coefficient (CC), mean absolute error (MAE), root mean square error (RMSE), and bias percentage (PBias) for dry and wet seasons between satellite-based and rain gauge recorded daily

rainfall over the Upper Tocantins River basin, from 2000 to 2019.

Figure 6

Boxplot of rainfall prediction ability of GSMaP products (2000-2019) in dry and rainy seasons compared to rain gauge observations in the Upper Tocantins River basin.

Figure 7

Scatterplots of rainfall between GSMaP data and rain gauge data at daily (A, B) and monthly (C, D) timescales during the whole study period. The diagonal reference line is indicated by a dashed line, and the fitting line (determined via the least-squares method) is indicated by a red solid line.

Figure 8

Probability distribution function (PDFv) of daily rainfall volume with different rainfall rates for rain gauges, and GSMaP and GSMaP Operational Reanalysis products.

Figure 9

Scatterplot of GSMaP monthly data before and after bias correction procedure for rainy (A, B) and dry (C, D) seasons during the period from 2000 to 2019.