

<https://doi.org/10.1590/2318-0331.262120210071>

Validation of TRMM data in the geographical regions of Brazil

Validação de dados TRMM nas regiões geográficas do Brasil

Rafael Brandão Ferreira de Moraes¹  & Fábio Veríssimo Gonçalves¹ 

¹Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, Brasil

E-mails: rafael.moraes@ufms.br (RBFM), fabio.goncalves@ufms.br (FVG)

Received: May 18, 2021 - Revised: October 08, 2021 - Accepted: October 14, 2021

ABSTRACT

The low density of precipitation gauges, the areas of difficult access and the high number of missing values hinder a rapid and effective hydrological monitoring. Thus, the present study aims to statistically validate the precipitation estimates by the data Tropical Rainfall Measuring Mission (TRMM) in relation to the data observed in the Conventional Meteorological Stations (CMSs) in the geographic regions of Brazil. The statistical indicators used were: Correlation Coefficient (r), Mean Absolute Error (MAE), Percentage of Bias (Pbias), T-Test and Mann-Witney Test. It is concluded that the precipitation data estimated via TRMM are effective and reliable alternatives for hydrological studies in areas that do not have in-situ gauges and/or need to fill missing values in the five regions of the country, especially in the driest months and in larger time scales.

Keywords: Estimate; Precipitation; TRMM; Conventional meteorological station; Geographic region.

RESUMO

A baixa densidade de estações meteorológicas superficiais, áreas de difíceis acessos e elevada quantidade de valores ausentes prejudicam um monitoramento hidrológico rápido e eficaz. Assim, o presente estudo tem como objetivo validar estatisticamente as estimativas de precipitação pelo produto de satélite *Tropical Rainfall Measuring Mission* (TRMM) em relação aos dados observados nas Estações Meteorológicas Convencionais (EMCs) nas regiões geográficas do Brasil. Os indicadores estatísticos utilizados foram: Coeficiente de Correlação (r), Erro Médio Absoluto (EMA), Percentual de Biais (Pbias), Teste-T e Teste de Mann-Witney. Conclui-se que os dados de precipitação estimados via TRMM são alternativas eficazes e confiáveis para estudos hidrológicos em áreas que não possuem medidores in-situ e/ou necessitem de preenchimento de valores ausentes nas cinco regiões do país, sobretudo nos meses mais secos e em escalas maiores de tempo.

Palavras-chave: Estimativa; Precipitação; TRMM; Estação meteorológica convencional; Região geográfica.

INTRODUCTION

Effective hydrological monitoring is essential for the planning and operation of various sectors of society, such as agriculture, livestock, navigation, water supply, flood control, among other analyzes involving hydrological basin balance and river flow regime (Soares et al., 2016). According to World Meteorological Organization (2008), a minimum density of precipitation gauges is important to avoid problems in the development and management of water resources in physiographic units. World Meteorological Organization (2008) recommends at least one Conventional Meteorological Station for each: 900 km² of coastal, 250 km² of mountains, 575 km² of interior plains, 25 km² of small islands and 10 000 km² of polar or arid region.

However, these recommendations are not always possible, since several regions are difficult to access and do not have sufficient resources for such installations, monitoring and periodic maintenance. In addition, these instruments measure the local rain, not capturing its spatial distribution, especially in land with complex topography, being susceptible to precipitation detection failures (Pereira et al., 2013), which are not always filled in a way effective and immediate. In this scenario, the precipitation estimate using satellite appears as a viable option to complement the spatial and temporal constraints from traditional stations. Hence, studies were carried out in order to identify the best satellite to estimate the rainfall regime in some areas of Brazil, from the comparison of data estimated via satellite with data observed in-situ.

Reis et al. (2017) evaluated the performance of precipitation estimates of the Tropical Rainfall Measuring Mission (TRMM) and Hydroestimator (HYDROE) satellites in the Sapucaí River Basin. Nogueira et al. (2018) analyzed the performance of the products of the TRMM and Climate Hazards Group InfraRed Precipitation (CHIRPS) satellites in the state of Minas Gerais. Ringard et al. (2015) examined the estimates of TRMM, Climate Prediction Center Morphing Technique (CMORPH) and Precipitation Estimation from Remotely-Sensed Information using Artificial Neural Network (PERSIANN) satellites in the North Region of Brazil. In general, all these authors concluded that the TRMM

satellite provides the best precipitation estimate in their respective study areas, showing good performance even in more extreme events such as El Niño (Erazo et al., 2018), encouraging validation of this dataset in other regions of the country, in order to use it in future hydrological studies.

Some authors are already using only the precipitation dataset from the TRMM satellite for hydrological studies around the world, such as Santos et al. (2019) to analyze the spatio-temporal variability of rain in the State of Paraíba in Brazil, Darzi (2018) to analyze precipitation in the Pantanal Biome in Brazil, Corporal-Lodangco & Leslie (2017) to identify climatic zones in the Philippines and Islam & Uyeda (2007) in Bangladesh. Nevertheless, it is recommended that before using satellite precipitation estimates to fill missing values or even substitute in-situ gauges in studies, it is necessary to perform the statistical validation of these data for the respective region to be evaluated identifying possible inconsistencies and not compromising the result of the research (Silva et al., 2019).

Considering the challenges encountered to characterize the rainfall regime of a region from in-situ gauges and the need to validate the estimated data via satellite, this study aims to statistically validate the TRMM precipitation estimates in relation to the data observed in the Conventional Meteorological Stations in the geographic regions of Brazil, in order to use them in hydrological monitoring in an effective and reliable way in all areas of the country, even the most remote ones.

MATERIAL AND METHODS

The study areas of this work comprise the South, Southeast, Midwest, North and Northeast Regions of Brazil, characterized by economic, cultural, and climatic differences. Table 1 shows the states, area, population, topographic surface, climate, vegetation and main rivers, agricultural production and economy of each region, according to the Brazilian Institute of Geography and Statistics (IBGE) (Instituto Brasileiro de Geografia e Estatística, 2010) and Brazilian Agricultural Research Corporation (EMBRAPA) (Empresa Brasileira de Pesquisa Agropecuária, 2021). The dataset used are

Table 1. Main characteristics of the Brazilian geographic regions.

Features	South	Southeast	Midwest	North	Northeast
States	RS, SC e PR	SP, MG, RJ e ES	MS, MT e GO	AM, PA, AC, RR, RO, AP e TO	MA, PI, CE, RN, PB, PE, AL, SE e BA
Area (km ²)	576 743	924 565	1 606 239	3 851 281	1 551 991
Population	2 9975 984	88 371 433	16 297 074	18 430 980	57 071 654
Relief	Plateaus and plains	Mountain ranges and plains	Plateaus and plains	Plains and depressions	Plains, depressions and plateaus
Climate	Subtropical	Tropical and tropical altitude	Semi-humid tropical	Equatorial	Equatorial humid, semi-arid tropical and humid coast
Vegetation	Araucaria forest and grasses	Atlantic forest	Wetland and savannah	Amazon rainforest	Caatinga
Main rivers	Paraná, Jacuí e Itajaí	Tietê, Paraíba do Sul e Paraná	Paraguai e Xingu	Amazonas e Tocantins	São Francisco e Parnaíba
Main agricultural production	Rice	Sugar cane	Soy	Extractivism	Sugar cane and fruitful
Main Economy	Agriculture, industry and tourism	Industry and agriculture	Agriculture and livestock	Fishing, extractivism and livestock	Tourism

the satellite precipitation estimates, product of Tropical Rainfall Measuring Mission (TRMM) 3B42 V7 - denominated virtual stations - and the precipitation data observed in the Conventional Meteorological Stations (CMSs) – in-situ gauges - from January 2010 to December 2019. The study was carried out on the monthly and annual scales based on the accumulated precipitation.

Figure 1 represents the study areas with the locations of the analyzed stations. The South Region comprises the states of Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR). Southeast by São Paulo (SP), Minas Gerais (MG), Rio de Janeiro (RJ) and Espírito Santo (ES). Midwest by Mato Grosso do Sul (MS), Mato Grosso (MT) and Goiás (GO). North by Amazonas (AM), Pará (PA), Acre (AC), Roraima (RR), Rondônia (RO), Amapá (AP) and Tocantins (TO). Finally, Northeast Region by Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE) e Bahia (BA).

TRMM is a joint mission between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) and aims to measure the intensity and area of rain cover around the tropical and semitropical area, where two thirds of the world's rains occur (Tan et al., 2015). TRMM data were collected from the Agrometeorological Monitoring System, administered by the IT sector of the Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária, 2020) directly in the form of accumulated precipitation in millimeters (mm) at the virtual stations. Whereas the observed data from CMSs were acquired from the Meteorological Database for Teaching and Research of the National Institute of Meteorology (INMET) (Instituto Nacional de Meteorologia, 2020).

First, CMSs were screened, selecting only those that provided complete monthly and annual data from 2010 to 2019, in other words, without missing values. Therefore, 17 CMSs were identified in the South Region, 28 in the Southeast, 17 in the Midwest, 23 in the North and 26 in the Northeast. The Table S1, available in supplementary section, shows the municipalities in which INMET CMSs are located with their respective state and geographic region. From their location, the virtual stations (TRMM) were chosen by using the nearest-neighbor interpolation technique. Next, all data were separated and organized on monthly and annual spreadsheets. Subsequently, variability analysis and statistical tests were carried out in order to validate the estimated TRMM data in relation to the data observed in the CMSs.

For the analysis of data variability, the following statistical indicators were calculated: Correlation Coefficient (r), Mean Absolute Error (MAE) and Bias Percentage (Pbias). The r is the measure of the strength and direction of a linear relationship between two variables (Equation 1) (Larson & Farber, 2010). The MAE quantifies the error associated with rainfall estimates, corresponding to the average of the absolute differences between the observed and estimated values (Equation 2). While Pbias measures the average tendency of TRMM data to over or underestimate the observed data (Equation 3).

$$r = \frac{(E - Em) \cdot (O - Om)}{\sqrt{\sum_{i=0}^n (Ei - Em)^2 \cdot \sum_{i=0}^n (Oi - Om)^2}} \quad (1)$$

$$EMA = \frac{\sum_{i=0}^n |Ei - Oi|}{n} \quad (2)$$

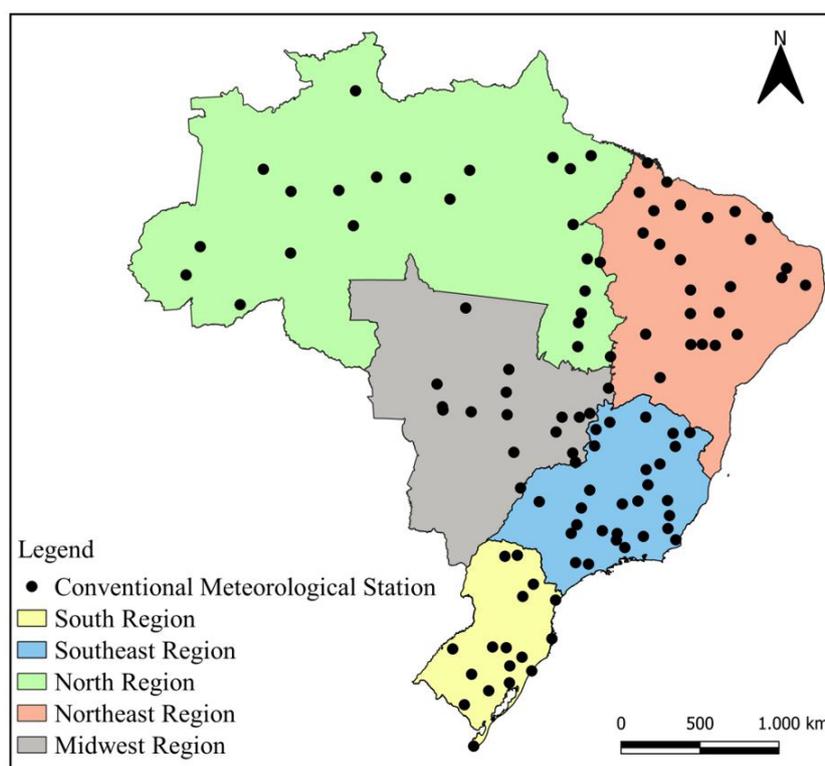


Figure 1. Geographic regions of Brazil and location of in-situ gauges.

$$P_{bias} = \sum_{i=0}^n \frac{(E_i - O_i)}{\sum_{i=0}^n O_i} \cdot 100 \quad (3)$$

Where: E_i = value estimated by the satellite in time interval i ; O_i = value observed on the surface in time interval i ; n = number of data analyzed; E_m = average value estimated by the satellite and O_m = average value observed on the surface.

The correlation coefficients were classified as positive correlation: perfect (1), very strong (0.90 to 0.99), strong (0.7 to 0.89), moderate (0.50 to 0.69), weak (0.01 to 0.49), no correlation (0) and negative (-0.01 to -1), a convention like the one proposed by Callegari-Jacques (2003). The closer to 1, 0 and -1 it means that the data are directly proportional, without correlation and inversely proportional, respectively (Larson & Farber, 2010).

Statistical tests were started from the verification of the normality of each group of data using the Shapiro-Wilk Test. If one group or both (TRMM and/or INMET) were non-normal, then the Mann-Witney Test was performed ($\alpha = 5\%$). However, if the two groups were normal, their respective variances were calculated using the Levene Test. In the condition that the two variances are statistically equal, if the T-Test were calculated assuming equal variances, otherwise it would be T-Test assuming different variances.

Finally, if P from the T-Test or Mann-Witney Test was greater than 0.05, the two groups were considered (January TRMM and January INMET, February TRMM and February INMET, etc. on the monthly scale or 2010 TRMM and 2010 INMET, 2011 TRMM and 2011 INMET, etc. on the annual scale) with statistically equal means, at the 95% significance test ($\alpha = 0.05$). Otherwise, $P < 0.05$, groups with statistically different means

(Larson & Farber, 2010). Figure 2 shows the flowchart of the methodology of statistical tests in a summarized way.

It is important to emphasize that more than one statistical indicator was evaluated along with statistical tests, due to the period responsible for low rainfall, especially winter in most of Brazil. Because if an CMS measures a local rain of 2 mm and the TRMM does not estimate it over its respective catchment area, or vice versa, consequently there will be a difference of 200% between the data or even a negative correlation coefficient, for example. Nonetheless, this does not mean a poor performance of the TRMM satellite, since the difference in mm is small and the other statistical analysis can find satisfactory results (t , MAE, Mann-Witney Test and T-Test).

After validating the estimated precipitation data via satellite, the Thiessen Polygon methodology was used in the Quantum Geographic Information System (QGIS) version 3.16 software to calculate the average annual precipitation of a Brazilian State from data observed in conventional meteorological stations and data estimated on the TRMM satellites, in order that show the best distribution of the last group of data, based on the generation of a greater number of polygons and probably results closer to reality.

RESULTS AND DISCUSSION

Figure 3 shows the graphs referring to the occurrence rates of the coefficients of perfect positive correlation, very strong, strong, moderate, weak, and negative between the estimated TRMM data and observed INMET data on the monthly scale by geographic region of Brazil. It can be seen that the months between May and September, the driest period in most of Brazil, are responsible for the occurrence of the highest rates of strongest

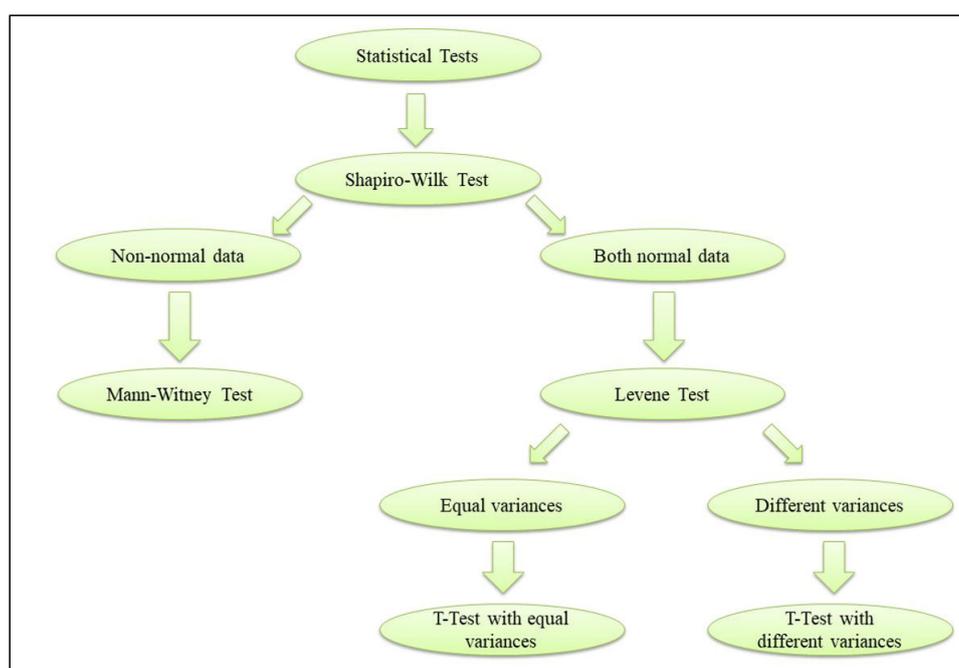


Figure 2. Flowchart of statistical tests.

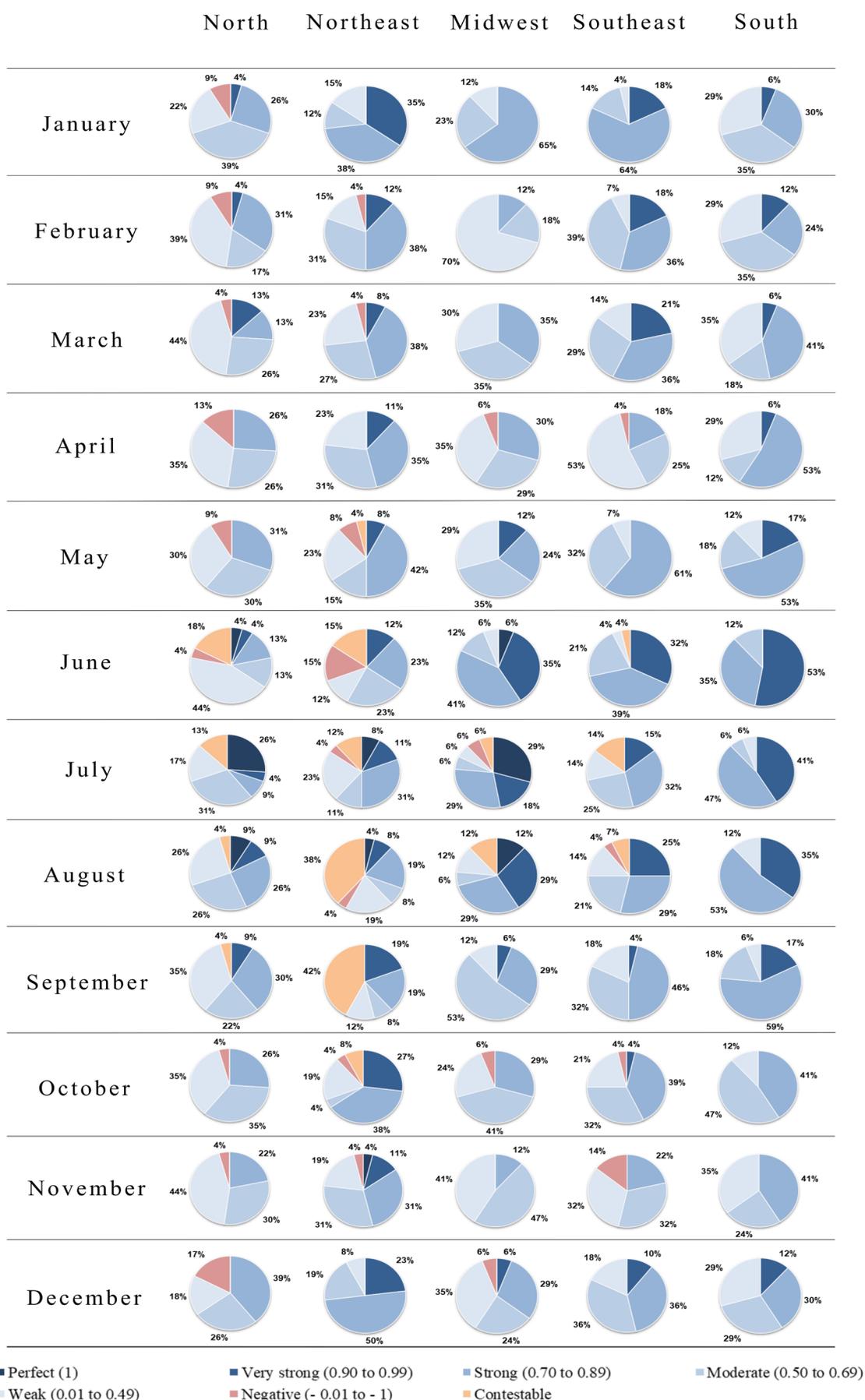


Figure 3. Graphs related to the performance of the monthly correlation coefficient analysis by geographic region.

positive correlations ($r \geq 0.70$) among the data from the TRMM and INMET in the South and Southeast Regions. While in the Midwest these correlations are concentrated only in the months of June, July and August. The North Region centralizes the most significant positive correlations only in the months of June and August, whereas in the Northeast this occurs from December to March.

In general, between the months of October to April, the wettest period in most of the Brazilian territory, the amount of moderate and weak positive correlation coefficients increases considerably in the South and Midwest Regions. While in the Southeast, the correlations are more distributed between strong, moderate and weak in the same period, except for April, the month with the highest weak correlation rate (53%) between the TRMM and INMET data.

In the North, the data correlation resulted in correlation coefficients classified as moderate and weak ($r < 0.70$) in the approximately 60% of the statistical analysis for the first four months of the year. On the other hand, the Northeast Region presents this inferior correlation performance mainly in the months of April and June. Still, in relation to the monthly variability of data from satellites and in-situ gauges, Figure 4 and 5 shows the results of Pbias and MAEs, respectively, of part of the analysis carried out in the South, Southeast and Midwest Regions.

From Figure 4 it is possible to conclude that in the months of strongest correlations in the South, Southeast and Midwest Regions, from May to September mainly, TRMM data are overestimated when compared to those observed in CMS. In the South they are overestimated between +1 and +25% predominantly, from May to July. In the Southeast they are overestimated between +1 and +25% mainly in the month of May and reach very significant percentages in August and September (Arinos - MG with +298% and Januária - MG with +280%, respectively, for example), as well as in some analysis located in other Brazilian regions, such as the Midwest (Aragarças - GO with +161%, Matupá - MT with +343%, Poxoréo - MT with +130%, etc.). By contrast June and July there is no definite trend in the Southeast. The results of Pbias by municipality can be seen in Figure S1, as well as MAE results in Table S2.

However, many of these sub and/or overestimated results, as well as some correlation coefficients (Figure 3) are classified as contestable, since the Southeast, Midwest, North (part of the state of Tocantins) and Northeast are characterized by having part of their territories located in areas with hot and dry climate. According to Empresa Brasileira de Pesquisa Agropecuária (2021), among the predominant climates of the Southeast, Midwest and Northeast Regions are tropical (dry winter), semi-humid tropical (dry winter) and semi-arid (hot and dry), respectively. Thus, low or nonexistent rainfall from May to September is common in these regions, it can cause large differences between the data and even a negative correlation coefficient, because small local precipitation can be captured by in-situ gauges and not by the TRMM satellite or vice versa.

In this context, although significant overestimations were found in some analysis of the Midwest, such as +355% in Matupá - MT in July and +63% in Goiânia - GO in September, their MAEs were 2.8 and 3.1 mm, respectively, according to Table S2.

Furthermore, even though relatively low overestimations were identified in the month of August in the South Region, high MAEs were found, such as 75 and 60 mm in Porto Alegre and Londrina in the same period, respectively. These facts reinforce the importance of analyzing several statistical indicators simultaneously, since an isolated result classified as terrible may not represent reality, such as the overestimates of Matupá, Goiânia and Aragarças, among others.

Moreover, it is important to highlight that the analysis of statistical variability in this study were carried out based on initial data of monthly accumulated precipitation at the level of each season, while other studies such as Pereira et al. (2013), used average precipitation values for geographic regions. This can be a justification for results that refer to a lower performance of the TRMM, because when comparing the spatial averages of estimated and observed precipitation, super and underestimation errors tend to compensate, increasing the chance of greater agreement between the two variables (Soares et al., 2016).

Figure 6 represents the graph with the percentages of the statistical tests for each region that resulted in a significant difference between the averages from the estimated accumulated rainfall and observed on the monthly scale, from the Shapiro-Wilk, Levene, T-Student and Mann-Witney Tests. It is possible to conclude a good performance of the precipitation estimates in the months of January, February, May to August, November and December for the South Region, since no statistically significant differences were identified between the means of the two groups of data ($P > 0.05$, $\alpha = 5\%$) in none of the parametric and non-parametric tests performed. The exceptions were March, April, September and October, which obtained a significant difference between the data averages in 5.9% of the analysis of the first three months mentioned and 11.8% in October.

In the Southeast Region, it is noteworthy that although most months between October and February have the highest rates of weak and moderate correlations between the data, no statistically significant difference was noticed between their respective averages from the T and Mann-Witney Tests ($P > 0.05$, $\alpha = 5\%$), as shown in Figure 6. The exception is April, which has a weak positive correlation between the data in more than 50% of the 28 statistical analysis carried out this month. Indeed, April was identified a significant difference between the group averages of approximately 7% of the parametric and non-parametric tests performed in the Southeast Region, that is, the greatest proportion of difference found, next to July, when compared to the other months.

Despite the poor performance in April, it is possible to use precipitation estimates derived from the TRMM data in locations that do not have rain monitoring networks in the Southeast of Brazil, as concluded by Camparotto et al. (2013) in the state of São Paulo, Aires et al. (2016) in the Paraopeba River Basin and Silva et al. (2019) in the states of São Paulo and Paraná.

In the Midwest Region, the results show that from April to August, the driest period, no significant differences were identified between the means of the estimated TRMM data in relation to those observed by INMET. While in the rainy season, from September to March, at least one comparison showed a significant difference between the data averages, with emphasis

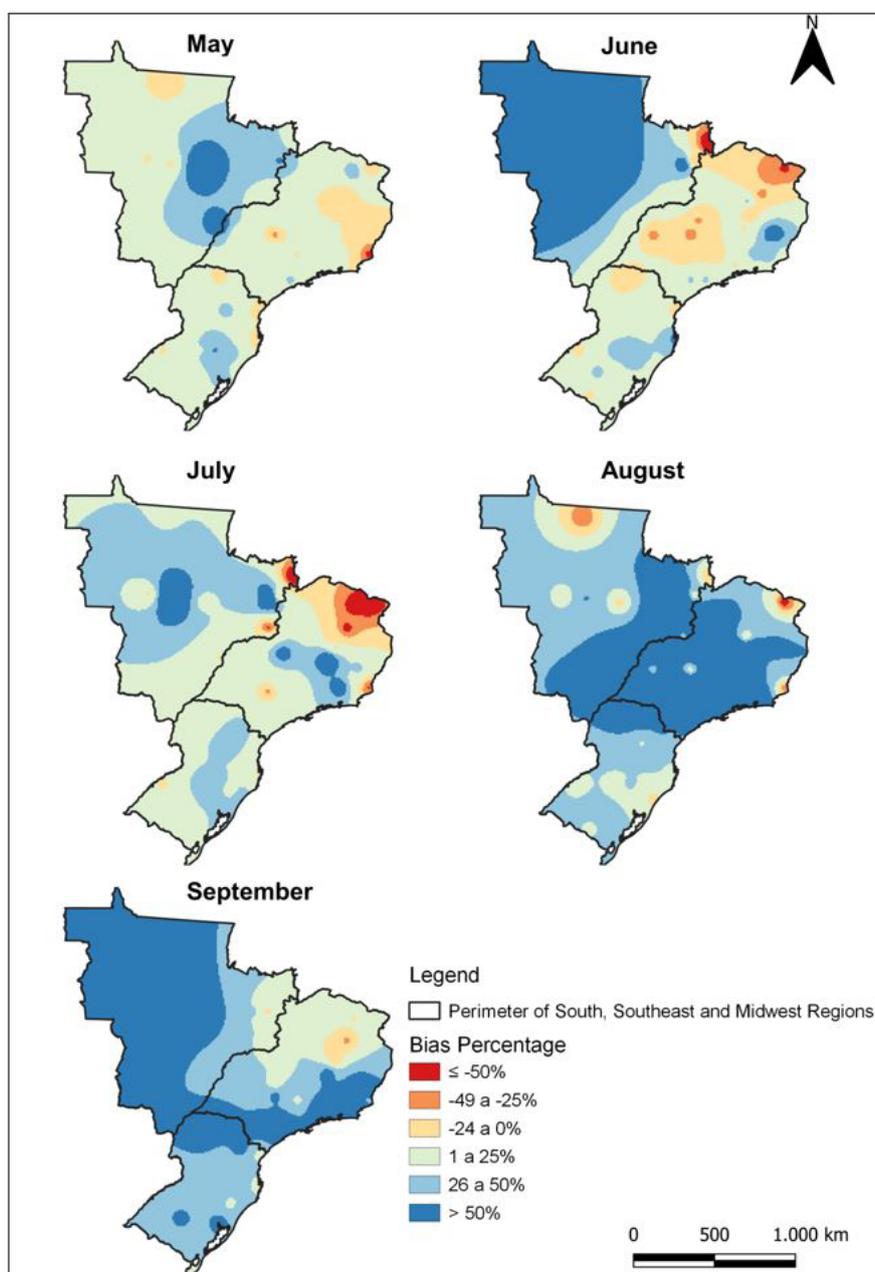


Figure 4. Result of the Percentage of Bias (Pbias) of the analysis carried out in the months of May to September in the Southeast, Midwest, and South Regions.

on the month of November. At the same time, it is important to highlight that the months of February, April and December presented the worst correlations between the data in this region. This may be associated with the amount of rainfall for each month and the more accuracy of estimating non-occurrence of rain than exact amount of rain in greater volume (Soares et al., 2016).

In general, the Midwest Region achieved an interesting performance of the estimates generated by the TRMM, especially in drier months. Thus, it is possible to use this satellite in places that do not have CMSs in this region and/or need to fill missing values. In fact, Pessi et al. (2019) reached this same conclusion for the State of Mato Grosso, when they identified a strong correlation between the precipitation estimates through the TRMM

3B43 satellite and the observations coming from the network of conventional pluviometric stations in this state.

In the North Region, only two months did not show significant differences between the averages of the estimated and observed data - July and November. However, this does not detract from the good performance of the TRMM satellite, since in the other months the rate of municipalities with a significant difference between the data averages was low, being 4.3% in May and December and 8.7% in January, February, April, June, September and October. The exceptions were March and August with 13 and 17.4%, respectively.

In comparison to the Midwest Region, for example, which did not present any statistical divergence between the data averages

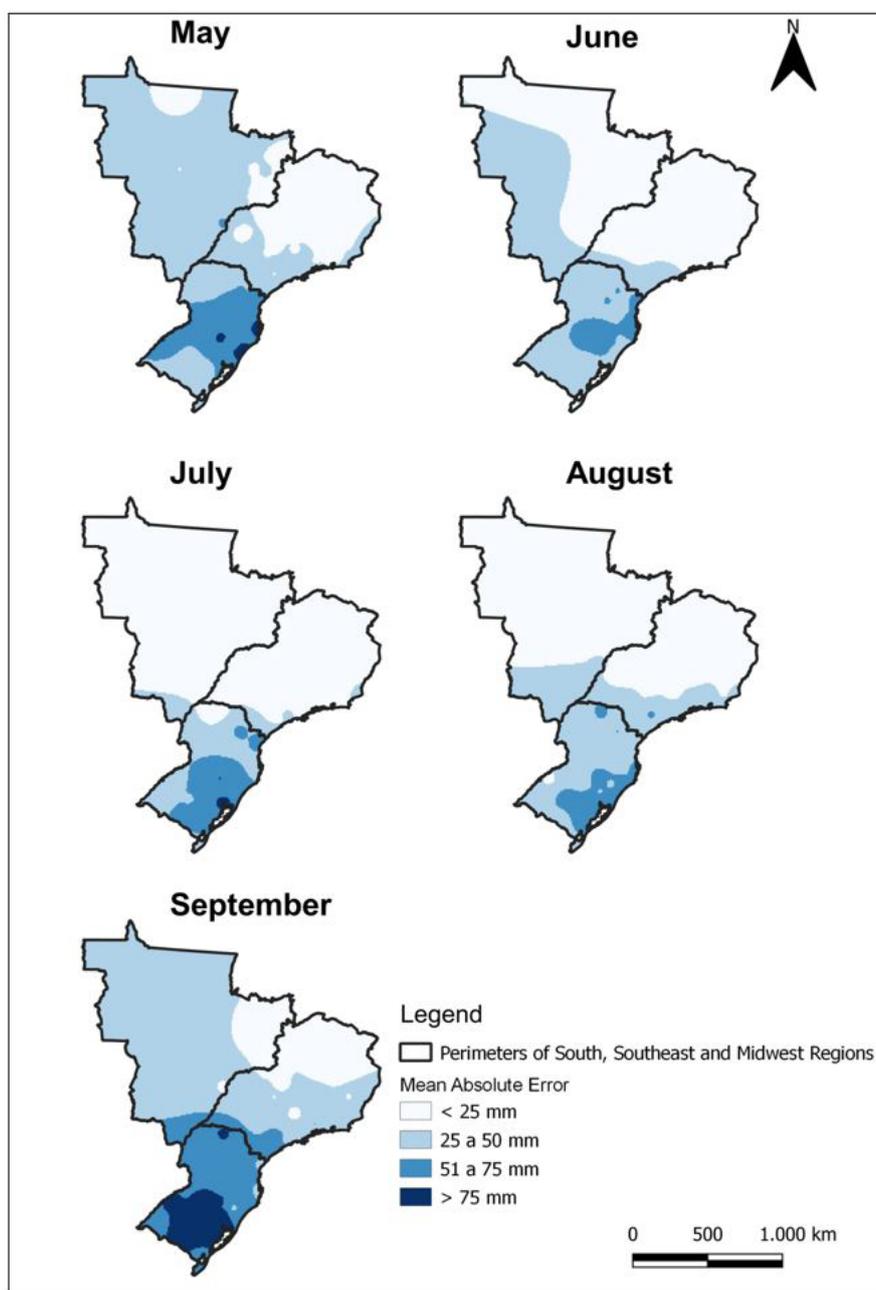


Figure 5. Mean Absolute Errors (MAEs) from May to September for partial analysis of the South, Southeast and Midwest Regions.

from April to August, a possible justification is that the North Region has a predominantly equatorial climate, that is, characterized by high rates of precipitation and well distributed (Empresa Brasileira de Pesquisa Agropecuária, 2021). Besides, due to the large territorial extension of the North, more local rains may occur and not be accounted for by CMSs, as these stations are fixed and cannot measure precipitation in a distributed manner, such as the TRMM satellite.

Despite this, the statistical results show a good performance among the estimated and observed data for the North Region. Santos et al. (2018) concluded that precipitation estimates from the TRMM V7 satellite can be used to represent precipitation rates in the Iriri River Basin, located in North Brazil, for hydrological modeling purposes. Erazo et al. (2018) concluded that this same

satellite provides reliable rainfall estimates even in more extreme events, such as El Niño, validating its use in regions with sparse rainfall stations and high rainfall variability.

Furthermore, in a study carried out in the North Region, Almeida et al. (2015) concluded that the TRMM product rainfall estimates are a good alternative source of data for the Amazon region according to statistical parameters, representing well the seasonal variability of rainfall. Thus, according to the same authors, the TRMM satellite can assist in rainfall studies of regions with low density of surface information, as well as in filling missing values and homogenizing precipitation data in CMS.

Regarding the monthly analysis of precipitation, there is the Northeast Region, which is composed of nine states with

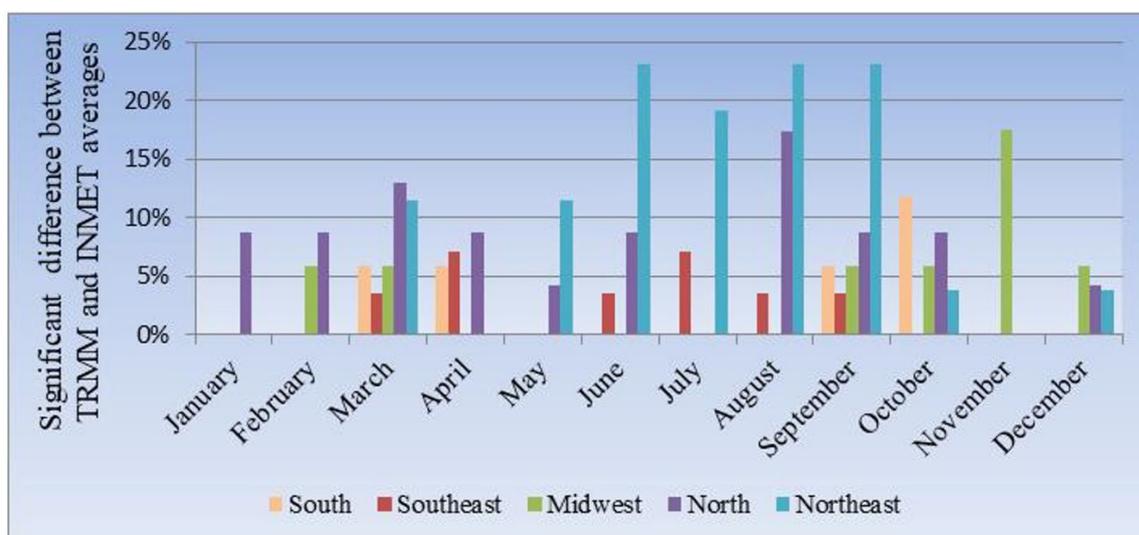


Figure 6. Percentage of monthly analyzes performed that showed a significant difference between the averages of INMET and TRMM data by Brazilian geographic region.

different physical, social and economic characteristics. According to Empresa Brasileira de Pesquisa Agropecuária (2021), this region has several characteristic climates, such as: humid equatorial, tropical, semi-arid and humid coastal. In view of the extensive territorial area and climatic diversity found in the region, the results found in Figure 6 show that the Northeast Region presents a lower performance when compared to the others, since in the monthly scale significant differences were observed between the means of the two groups of data in 23.1% of the analysis in the months of June, August, and September, 19.2% in July and 11.5% in the months of March and May.

This inferior performance may be related to the great climatic variety of the region, which presents poorly distributed rain, isolated and sometimes with an irrelevant daily amount, factors that together make it difficult to capture the sensors and consequently in the final monthly sum. In addition, the extensive coastal strip of the Northeast Region presents maritimicity and a precipitation regime quite different from the interior, which can compromise the quality of the satellite estimates, as observed by Bernardi (2016), who found worse performance of the TRMM estimates in the coastal strip of your study area.

In this context, the TRMM satellite estimates the precipitation in a distributed area and not local like CMS. Thus, the northeastern interior, which presents little rainfall, which often occurs in more isolated parts, can contribute to reduce the satellite's efficiency. However, despite the low performance when compared to other regions, estimates of rainfall from the TRMM generally reproduce the temporal pattern of the pluviometric regime in the Northeast Region, both in terms of seasonality and in terms of spatial distribution of rainfall, such as concluded by Soares et al. (2016) for the state of Paraíba.

Figures 7 and 8 highlight the behavior of precipitation estimates via TRMM satellite compared to data observed by CMS in the the annual scale, with respect to the correlation coefficient (r) and the significant statistical differences between the means of the groups studied, respectively.

In general, the results in Figure 7 express that the estimated data are more correlated to those observed by CMS in the annual scale than in the monthly scale in the five Brazilian geographic regions. However, the Southeast, Midwest, North and Northeast Regions stand out, as they portray $r \geq 0.70$ in more than 80% of the statistical analysis carried out in their respective stations, especially the Midwest with 44 and 45% of very strong and strong positive correlation, respectively. As well as the Northeast Region with a 48% very strong correlation between the estimated and observed data in its perimeter.

On the other hand, the South Region shows a lower performance compared to the others in the annual scale, since it obtained only 11% of very strong correlation between the data and the highest rates of weak and moderate correlations (10 and 29%, respectively). Nonetheless, this result does not invalidate the TRMMs accuracy in this region, since, according to Figure 8, no significant differences were found between the averages of the two groups of data in practically no year of the studied period, except for 2014.

In summary, all regions showed excellent performance on the annual scale, both in the correlation of the data and in the absence of a significant difference between the precipitation averages of the TRMM and CMS data, mainly in the Midwest and Southeast. This corroborates with the observations of Iqbal & Athar (2018) and Reis et al. (2017) who concluded that the larger the time scale of analysis, the better the precipitation estimate of the TRMM satellite, since a longer period allows temporal errors in the rain estimates to be compensated so that the total accumulated in the period is closer to that observed (Soares et al., 2016).

Finally, the average annual precipitation calculation for the state of Goiás (GO) and the Distrito Federal (DF) was carried out from 2010 to 2016, using the Thiessen Polygon methodology to compare the results from CMSs and TRMM (Table 2). The Figure 9 shows the Thiessen Polygons traced in the state of GO and DF according to the meteorological stations used, that is, in-situ (INMET) or virtual (TRMM).

The results in Table 2 show that the average annual rainfall calculated, only from the CMSs and TRMM located in the same

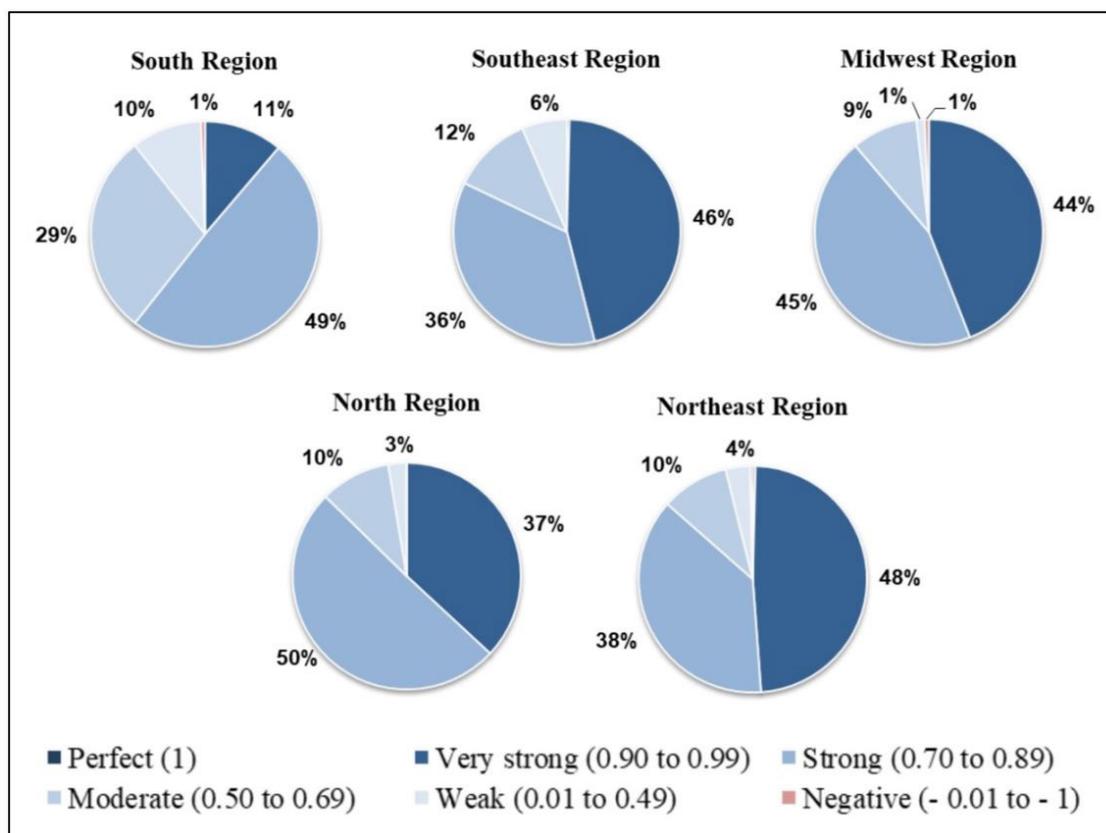


Figure 7. Graphs relating to the performance of the annual correlation coefficient analysis by geographic region.

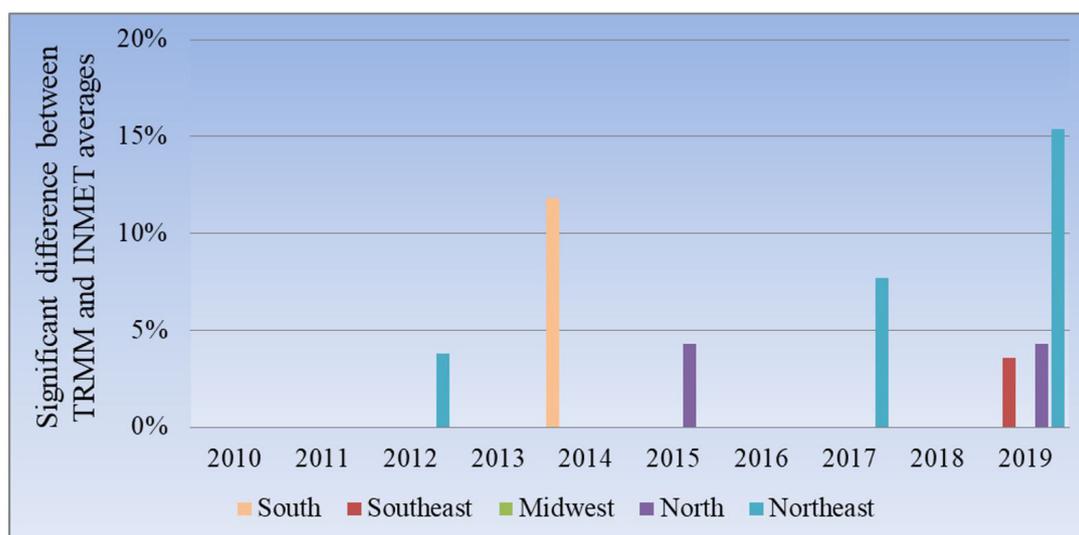


Figure 8. Percentage of annual comparisons made that showed a significant difference between the averages of the INMET and TRMM data by Brazilian geographic region.

geographic coordinates or as close as possible (Figures 9A and 9B), show relatively close results and a very strong positive correlation between the observed data and estimated ($r = 0.90$). According to Instituto Brasileiro de Geografia e Estatística (2020), the areas of the State of GO and the DF together are 346 006 km², while the areas of the Thiessen Polygons totaled 345 889 km² (Figures 8A and 8B)

and 345 888 km² (Figure 9C), that is, a difference of only 0.03% in relation to the official measure, representing practically perfect agreement between them.

Moreover, the average annual precipitation was calculated based on TRMM data in the same locations as the in-situ gauges of INMET, in addition to other TRMM inserted in areas that

Table 2. Average annual rainfall in the state of Goiás and the Distrito Federal calculated from CMSs and TRMM.

Stations used	Year	Precipitation by Thiessen Polygons (mm)
INMET		1,477.6
TRMM*	2010	1,413.8
TRMM		1,374.9
INMET		1,601.4
TRMM*	2011	1,596.6
TRMM		1,603.2
INMET		1,498.6
TRMM*	2012	1,534.4
TRMM		1,516.1
INMET		1,676.5
TRMM*	2013	1,732.1
TRMM		1,748.6
INMET		1,519.5
TRMM*	2014	1,468.9
TRMM		1,486.6
INMET		1,366.0
TRMM*	2015	1,449.1
TRMM		1,498.7
INMET		1,337.5
TRMM*	2016	1,203.8
TRMM		1,244.2

*Only TRMM data located in the same geographic coordinates as INMET CMSs or as close as possible.

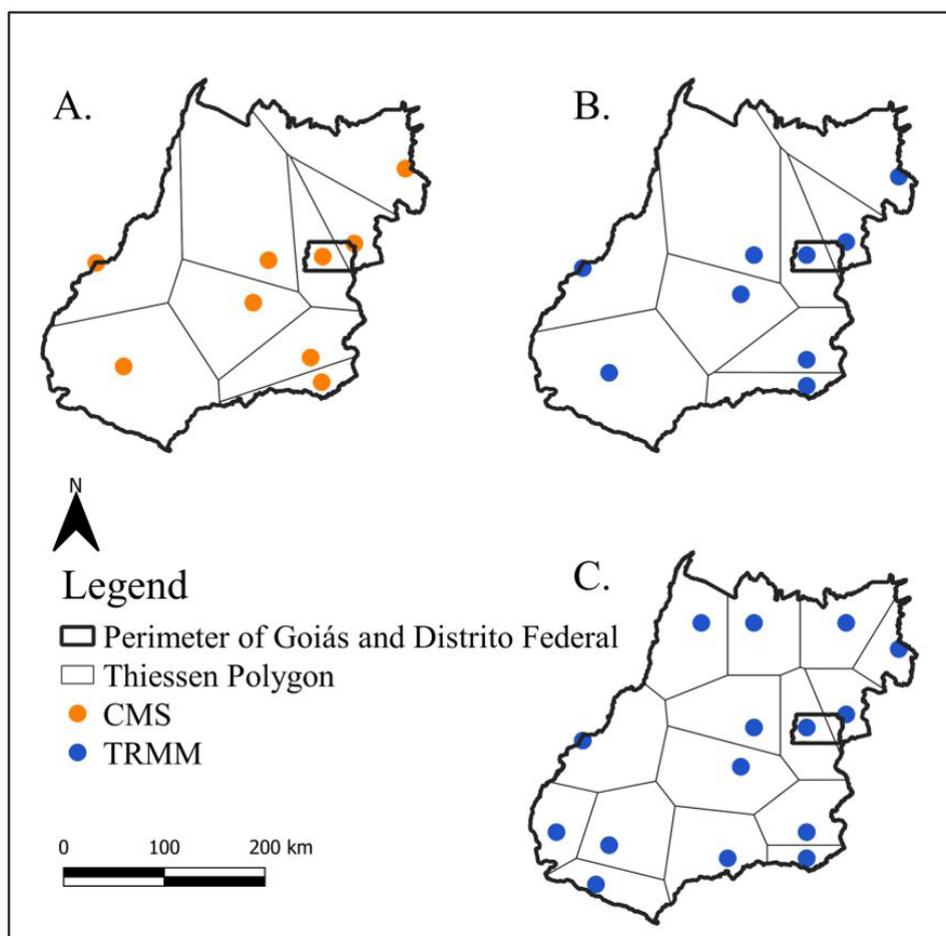


Figure 9. Maps with Thiessen Polygons from data from INMET (A) and TRMM (B and C) in State of Goiás and Distrito Federal. CMS = Conventional Meteorological Station and TRMM = Tropical Rainfall Measuring Mission.

were empty (Figure 8C), to homogenize the distribution of TRMM virtual stations in the state. These results obtained a greater difference when compared to the first (only CMSs), due to the fact of recalculating the Thiessen Polygons based on more TRMM data and, therefore, probably closer to reality, since the greater the number of Thiessen Polygons, the more distributed is the rainfall throughout the state. In fact, this is another advantage of the data estimated via satellite, that is, in addition to being reliable and effective, they have a greater spatial distribution in the region to be evaluated.

CONCLUSIONS

The present study aimed to statistically validate satellite precipitation estimates in relation to data observed in conventional meteorological stations in the five geographic regions of Brazil. The correlation coefficient, mean absolute error, percentage of bias, mann-witney test and t-test were calculated between the two groups of data in the monthly and annual scales.

Regarding numerous problems and low density of surface weather stations with complete rainfall data in different regions of Brazil, it is concluded that the data estimated via TRMM satellite are effective, reliable and better distributed. Furthermore, if the precipitation data estimated via satellite have already been validated and extracted for a particular region of interest, they can optimize hydrological monitoring studies in areas without conventional meteorological stations or with a high number of missing data.

Regarding the monthly scale, the statistical indicators analyzed between estimated and observed precipitation data obtained the best performances over the months considered driest in the five geographic regions. More specifically, May to September in the South, Southeast and Midwest, June and August in the North and December to March in the Northeast. The monthly statistical tests had excellent performances in every month of the five geographic regions, concentrating a significant difference between the averages of the monthly estimated and observed data in less than 15% of the analyzes carried out in each region.

Moreover, it is concluded that the performance of precipitation estimates is more satisfactory on a larger scale of time, that is, annual when compared to the monthly one, and that much more estimated precipitation data per square kilometer is available than data observed in in-situ gauges, significantly increasing the precision and accuracy of rain analysis in a region. Finally, for future work on the validation of precipitation data via satellite in other regions of the world, it is recommended to use more than one statistical indicator of variability, in addition to parametric and/or non-parametric statistical tests, to increase the accuracy and not compromise the conclusion of the respective study, positively or negatively, based on only one result. Thus, the existence of CMSs is essential for the validation of precipitation data obtained via satellite, requiring the continuity and expansion of the current network of in-situ gauges.

ACKNOWLEDGEMENTS

To the Support Foundation for the Development of Education, Science and Technology of the State of Mato Grosso

do Sul (FUNDECT) for the granting of the scholarship during the Doctorate. To the Federal University of Mato Grosso do Sul (UFMS) for the opportunity of the Doctorate. To the National Institute of Meteorology (INMET) and the Brazilian Agricultural Research Corporation (EMBRAPA) for making data available from conventional and TRMM meteorological stations, respectively.

REFERENCES

- Aires, U. R. V., Melo Neto, J. O., & Mello, C. (2016). Estimativas de precipitação derivadas do satélite TRMM para a bacia hidrográfica do Rio Paraopeba, MG. *Revista Scientia Agraria*, 17(2), 57-66. <http://dx.doi.org/10.5380/rsa.v17i2.46384>.
- Almeida, C. T., Delgado, R. C., Oliveira Junior, J. F., Gois, G., & Cavalcanti, A. S. (2015). Assessment of Rainfall Estimates from the TRMM-3B43 Product in the State of Amazonas. *Floresta e Ambiente*, 22(3), 279-286. <http://dx.doi.org/10.1590/2179-8087.112114>.
- Bernardi, E. C. S. (2016). *Qualidade das estimativas de precipitação do satélite TRMM no Estado do Rio Grande do Sul* (Dissertação de mestrado). Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria.
- Callegari-Jacques, S. M. (2003). *Bioestatística: princípios e aplicações*. Porto Alegre: Artmed.
- Camparotto, L. B., Blain, G. C., Giarolla, A., Adami, M., & Camargo, M. B. P. (2013). Validation of temperature and rainfall data obtained by remote sensing for the State of Sao Paulo. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 17(6), 665-671. <http://dx.doi.org/10.1590/S1415-43662013000600013>.
- Corporal-Lodangco, I. L., & Leslie, L. M. (2017). Defining Philippine climate zones using surface and high-resolution satellite data. *Procedia Computer Science*, 114, 324-332. <http://dx.doi.org/10.1016/j.procs.2017.09.068>.
- Darzi, L. B. G. (2018). *Análise da precipitação no Pantanal utilizando dados de sensoriamento remoto* (Dissertação de mestrado). Programa de Pós-graduação em Recursos Naturais, Universidade Federal de Mato Grosso do Sul, Campo Grande.
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. (2020). *Sistema de monitoramento agrometeorológico*. Retrieved in 2020, November 10, from <https://www.agritempo.gov.br/agritempo/index.jsp>
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. (2021). *Contando ciência*. Retrieved in 2021, May 20, from <https://www.embrapa.br/>
- Erazo, B., Bourrel, L., Frappart, F., Chimborazo, O., Labat, D., Dominguez-Granda, L., Matamoros, D., & Mejia, R. (2018). Validation of satellite estimates (Tropical Rainfall Measuring Mission, TRMM) for rainfall variability over the Pacific slope and Coast of Ecuador. *Water*, 10(2), 213. <http://dx.doi.org/10.3390/w10020213>.

- Instituto Brasileiro de Geografia e Estatística – IBGE. (2010). *Cidades e Estados*. Retrieved in 2021, March 22, from <https://www.ibge.gov.br/cidades-e-estados/df.html>
- Instituto Brasileiro de Geografia e Estatística – IBGE. (2020). *Cidades e Estados*. Retrieved in 2021, March 22, from <https://www.ibge.gov.br/cidades-e-estados/go.html>
- Instituto Nacional de Meteorologia – INMET. (2020). *Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP)*. Retrieved in 2020, November 8, from <https://bdmep.inmet.gov.br/>
- Iqbal, M. F., & Athar, H. (2018). Validation of satellite based precipitation over diverse topography of Pakistan. *Atmospheric Research*, 201, 247-260. <http://dx.doi.org/10.1016/j.atmosres.2017.10.026>.
- Islam, M. N., & Uyeda, H. (2007). Use of TRMM in determining the climatic characteristics of rainfall over Bangladesh. *Remote Sensing of Environment*, 108(3), 264-276. <http://dx.doi.org/10.1016/j.rse.2006.11.011>.
- Larson, R., & Farber, B. (2010). *Estatística aplicada*. São Paulo: Pearson Prentice Hall.
- Nogueira, S. M. C., Moreira, M. A., & Volpato, M. M. L. (2018). Evaluating precipitation estimates from ETA, TRMM and CHRIPS data in the South-Southeast Region of Minas Gerais State – Brazil. *Remote Sensing*, 10(3), 313. <http://dx.doi.org/10.3390/rs10020313>.
- Pereira, G., Silva, M. E. S., Moraes, E. C., & Cardozo, F. S. (2013). Avaliação dos dados de precipitação estimados pelo satélite TRMM para o Brasil. *Revista Brasileira de Recursos Hídricos*, 18(3), 139-148. <http://dx.doi.org/10.21168/rbrh.v18n3.p139-148>.
- Pessi, D. D., dos Santos, C. S. A., Nonato, J. J., Dourado, L. G. A., Silva, O. P., Bassini, R. T., & José, J. V. (2019). Validation of the monitors of the TRMM soil satellite in the State of Mato Grosso, Brazil. *Revista de Ciências Agrárias*, 42(1), 79-88. <http://dx.doi.org/10.19084/RCA18217>.
- Reis, J. B. C., Rennó, C. D., & Lopes, E. S. S. (2017). Validation of satellite rainfall products over a mountainous watershed in a humid subtropical climate region of Brazil. *Remote Sensing*, 9(12), 1240. <http://dx.doi.org/10.3390/rs9121240>.
- Ringard, J., Becker, M., Seyler, F., & Linguet, L. (2015). Temporal and spatial assessment of four satellite rainfall estimates over French Guiana and North Brazil. *Remote Sensing*, 7(12), 16441-16459. <http://dx.doi.org/10.3390/rs71215831>.
- Santos, C. A. G., Brasil Neto, R. M., da Silva, R. M., & Costa, S. G. F. (2019). Cluster analysis applied to spatiotemporal variability of monthly precipitation over Paraíba state using Tropical Rainfall Measuring Mission (TRMM) data. *Remote Sensing*, 11(6), 637. <http://dx.doi.org/10.3390/rs11060637>.
- Santos, V., Laurent, F., Abe, C., & Messner, F. (2018). Hydrologic response to land use change in a large basin in eastern Amazon. *Water*, 10(4), 429. <http://dx.doi.org/10.3390/w10040429>.
- Silva, C. B., Silva, M. E. S., Ambrizzi, T., Tommaselli, J. T. G., Patucci, N. N., Mataveli, G. A. V., Lima, B. S., & Correa, W. C. (2019). Precipitação na América do Sul – Dados obtidos em estações meteorológicas automáticas e sistemas orbitais. *Revista Brasileira de Climatologia*, 25, 54-79. <http://dx.doi.org/10.5380/abclima.v25i0.58813>.
- Soares, A. S. D., Paz, A. R., & Picilli, D. G. A. (2016). Assessment of rainfall estimates of TRMM satellite on Paraíba state. *Revista Brasileira de Recursos Hídricos*, 21(2), 288-299. <http://dx.doi.org/10.21168/rbrh.v21n2.p288-299>.
- Tan, M., Ibrahim, A., Duan, Z., Cracknell, A., & Chaplot, V. (2015). Evaluation of six high-resolution satellite and ground-based precipitation products over Malaysia. *Remote Sensing*, 7(2), 1504-1528. <http://dx.doi.org/10.3390/rs70201504>.
- World Meteorological Organization – WMO. (2008). *Guide to hydrological practices: hydrology from measurement to hydrological information* (6th ed., Vol. 1, No. 168, Cap. 2, pp. 24-27). Geneva: WMO.

Authors contributions

Rafael Brandão Ferreira de Moraes: Writing, theoretical development, statistical analysis and making figures and tables.

Fábio Veríssimo Gonçalves: Writing, analysis of results and reviews.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Fernando Mainardi Fan

SUPPLEMENTARY MATERIAL

Supplementary material accompanies this paper.

Table S1. Location of Conventional Weather Stations

Table S2. Mean Absolute Errors (MAEs) from May to September for partial analysis of the South, Southeast and Midwest

Figure S1. Result of the Percentage of Bias (Pbias) of the analysis carried out in the months of May to September in the Southeast, Midwest, and South Regions

This material is available as part of the online article from <https://www.scielo.br/j/rbrh>