Science of the Total Environment 799 (2021) 149492

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Geospatial cluster analysis of the state, duration and severity of drought over Paraíba State, northeastern Brazil

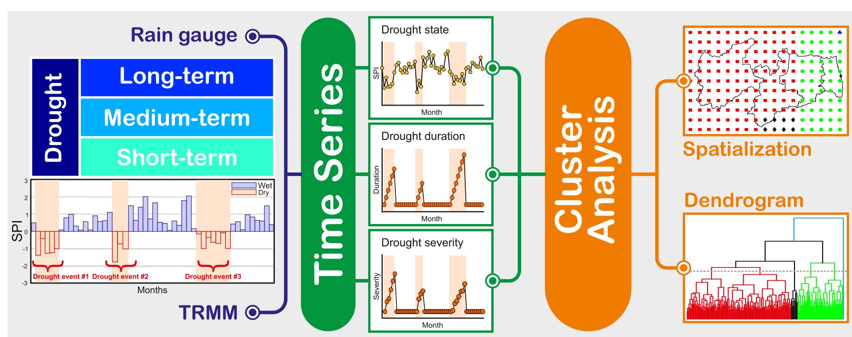
Reginaldo Moura Brasil Neto [a], Celso Augusto Guimarães Santos [a,*], Richarde Marques da Silva [b], Carlos Antonio Costa dos Santos [c], Zhong Liu [d,e], Nevil Wyndham Quinn [f]

[a] *Department of Civil and Environmental Engineering, Federal University of Paraíba, Brazil*
[b] *Department of Geosciences, Federal University of Paraíba, Brazil*
[c] *Academic Unit of Atmospheric Sciences, Federal University of Campina Grande, Brazil*
[d] *Center for Spatial Information Science and Systems (CSISS), College of Science, George Mason University, USA*
[e] *NASA Goddard Earth Sciences Data and Information Services Center GES DISC, Greenbelt, MD, USA*
[f] *Department of Geography and Environmental Management, University of the West of England, Bristol, UK*

HIGHLIGHTS

- Homogenous zones were identified based on drought state, duration, and severity.
- Cluster analysis was based on both gauge-measured and TRMM-estimated rainfall data.
- Integrated results of drought state, duration, and severity identified typical areas.
- This division was evident when assessing short-term droughts.
- Proximity to the ocean, climatic systems, and relief influences the drought regime.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 March 2021
Received in revised form 16 July 2021
Accepted 2 August 2021
Available online 5 August 2021

Editor: Christian Herrera

Keywords:
SPI
Drought
Remote sensing
TMPA
Clusters
Semiarid region

ABSTRACT

Droughts threaten water resources, agriculture, socio-economic activities and the population at the global and regional level, so identifying areas with homogeneous drought behaviors is an important consideration in improving the management of water resources. The objective of this study is to identify homogenous zones over Paraíba State in relation to the state, duration and severity of droughts that have occurred over the last 20 years (1998–2017) using hierarchical cluster analysis based on both gauge-measured and Tropical Rainfall Measuring Mission (TRMM) estimated rainfall data (TMPA 3B42). The drought series were calculated using the Standardized Precipitation Index (SPI) based on eight time scales and were grouped according to drought state, duration and severity time series. The integrated results of state, duration and severity of droughts indicate that there is a basis for dividing Paraíba State into two major regions (a) Zone I, formed by Mata Paraibana and Agreste Paraibano, and (b) Zone II, composed by Borborema and Sertão Paraibano. This division is evident when assessing short-term droughts, but in the case of long-term droughts, Paraíba State has a high similarity in terms of drought state, duration, and severity. Factors such as proximity to the ocean, active climatic systems, and the local relief configuration were identified as influencing the drought regime. Finally, it is concluded that TMPA rainfall estimates represent a valuable source of data to regionalize and identify drought patterns over this part of Brazil and that other studies of this type should be carried out to monitor these phenomena based on other satellite-based rainfall data, including the Global Precipitation Mission (GPM).

* Corresponding author.
E-mail address: celso@ct.ufpb.br (C.A.G. Santos).

https://doi.org/10.1016/j.scitotenv.2021.149492

1. Introduction

The future impacts of climate change for semiarid regions can cause the intensification and prolongation of droughts and generate serious problems, such as water scarcity and collapse in the water supply (Li et al., 2020). Droughts can also cause socio-environmental impacts of various magnitudes, such as desertification, reduction of agricultural potential and the rural exodus (Vieira et al., 2020). Drought events in many semiarid regions are frequent and are expected to increase in frequency and severity in the coming decades (IPCC, 2014). Precipitation measurements are therefore essential for water resources monitoring and in evaluating regional and global climate change (Mossad and Alazba, 2018).

Many developing or less developed countries have problems with collecting and storing high-quality and long-term meteorological data due to poorly developed and maintained hydrometric infrastructure and limited financial resources, which is especially difficult in arid and semiarid regions (Tan, 2019). Therefore, satellite precipitation products are used as an alternative data source to study the climate system in large or in non-instrumented watersheds (Tan and Duan, 2017). Among them, the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) dataset (Huffman et al., 2007; Liu et al., 2012) is a suite of precipitation products that includes daily long-term records of precipitation with acceptable accuracy for various regions of the globe (Qin et al., 2014; Prakash et al., 2015), and where reliability has been widely evaluated in relation to the general aspects of precipitation measurement (AL-Falahi et al., 2020) and hydrological modeling (Ur Rahman et al., 2020; de Medeiros et al., 2019; Silva et al., 2018).

TRMM precipitation data have also been used extensively in meteorological drought analysis. TMPA was one of the most important satellite products used to monitor rainfall in its era and used a fine spatial scale (i.e., 0.25°), whereas the resolution of other datasets, such as CRU (Climate Research Unit) or GPCC (Global Precipitation Climatology Centre), for example, is four times larger (i.e., 0.50°), which makes TMPA a useful dataset for regional analysis. Naumann et al. (2012) tested the feasibility of using TRMM 3B43 estimates for monitoring drought conditions and their uncertainties over four river basins in Africa, concluding that TRMM-based SPI estimation was reliable. Zeng et al. (2012) demonstrated the value of TRMM in mapping drought in the susceptible Lancang River Basin, China and validated the monitoring accuracy of TMPA 3B43 for two severe droughts, confirming the potential of TMPA for drought monitoring in data-poor regions. TRMM data were also found to be well correlated with ground observations in the Loess Plateau of Northwest China (Zhao et al., 2018). For Jiangsu Province, China, Tao et al. (2017) suggested that TRMM 3B43 data performed well for short-term drought monitoring, but the accuracy decreased for longer time scales.

Chen et al. (2020) reported good correlations at a shorter time scale (1, 3, 6 months) and conclude that the TRMM 3B43 precipitation product is reliable in drought monitoring over the Yangtze River Basin. Similar comparative studies have been undertaken in other localities, for example, Mexico (De Jesús et al., 2016), Malaysia (Tan et al., 2017), Brazil (Ferreira da Silva et al., 2020; Brasil Neto et al., 2020), Iran (Amini et al., 2019), Iraq (Suliman et al., 2020), Nepal (Sharma et al., 2020) and Morocco (Hadria et al., 2019), all of which broadly confirm the use of TRMM data in regional drought analysis. Although these studies show the potential value of TRMM data in drought analysis, further studies are still needed to analyze the quality of the TRMM rainfall products in the identification of state, duration, and severity of droughts for different regions (dry and wet), as is the case of Paraíba State.

Paraíba State is a predominantly semiarid region located in the Northeast of Brazil, one of the most vulnerable areas in the world due to climate change and where droughts are frequent (Dantas et al., 2020). Furthermore, Paraíba State has diverse climatic and geomorphological characteristics that control the spatiotemporal distribution of rainfall (Santos et al., 2019a), which makes monitoring precipitation and droughts at more detailed space-time scales a complex task. For this reason, it is a good locality for evaluating the utility of TRMM in the monitoring and analysis of drought across different climatic zones. Furthermore, the region has recently experienced one of the most severe drought events of recent times, with significant socio-economic consequences for the population of the state (Marengo et al., 2017; Santos et al., 2019b). For this reason, improved drought monitoring is therefore critical for improved resilience in water resource management.

To monitor the state, duration and severity of droughts, several metrics have been developed (RajKhatiwada and Pandey, 2019). These indices integrate various variables such as precipitation, temperature, flow, evapotranspiration, and humidity and can be interpreted on a severity scale (normal, wet, medium, or dry) to provide a comprehensive view of this phenomenon for decision-making. However, each drought index has different characteristics and is suitable for specific environments (Zhang et al., 2017), a factor that has stimulated several comparisons of alternative indices in various climatic regions of the planet. Studies of this nature are scarce in Paraíba State because the hydrometeorological time series have many gaps, challenging the analysis using multiple indices (Santos et al., 2019b; Brasil Neto et al., 2020; Brasil Neto et al., 2021). In ungauged, remote, and complex regions, as is the study area, obtaining a reliable rainfall data time series is easier than obtaining time series for other meteorological variables. The advantage of using the SPI in areas such as Paraíba State is that the SPI is only based on precipitation data and can be used for monitoring drought and wet conditions. In this context, SPI (McKee et al., 1993) is an important tool to assess the geospatial distribution of meteorological drought over Paraíba State.

Identifying areas with similar drought characteristics is an important but challenging task, as it typically requires high levels of local knowledge, understanding and experience for each region. Hierarchical cluster analysis methods offer a means of extracting greater understanding from different time series and have become notable as one of the most suitable instruments for defining pluviometrically homogeneous regions and their climate trends at regional and global scales (Unal et al., 2003; Keller Filho et al., 2005; Lyra et al., 2014; Teodoro et al., 2016; Oliveira-Júnior et al., 2017; Brito et al., 2017; Santos et al., 2019a). Drought zoning based on the state, duration and severity of these phenomena is a theme of interest in some studies (Rad and Khalili, 2015; Li et al., 2015; Wang et al., 2015; McGree et al., 2016; Shiau and Lin, 2016; Yang et al., 2017), but there is a lack of more detailed studies in the arid and semiarid regions.

This study aims to address this specific knowledge gap and further contribute to the literature on monitoring, classifying and mapping meteorological drought using hierarchical cluster analysis, a methodology that should be applied for other regions. The findings of our paper allow not only the comparison between rain gauge-measured and TRMM-estimated data, but also make it possible to identify regions based on different drought characteristics and different time scales. As far as we are aware, this is the first study to evaluate the performance of the TRMM product for drought regionalization over multiple time scales and characteristics over this area. The results can inform decision-making by different water resources sectors, such as agriculture and public water supply, which is particularly relevant in ungauged, remote, and complex regions, such as this. The specific objective of this study is to identify homogenous zones over Paraíba State as to the state, duration and severity of droughts that occurred over the last 20 years (1998–2017) using hierarchical cluster analysis based on both gauge-measured and TMPA-estimated rainfall data. In so doing, we provide an important understanding for the management of scarce water resources in a region of Brazil characterized by frequent and severe droughts.

2. Material and methods

2.1. Study area

The study area is Paraíba State, with a total area of 56,469.78 km^2 and a population of about four million inhabitants living in 223 municipalities (IBGE, 2016). Paraíba State is located between latitudes 5.875°S and

8.375°S and longitudes 38.875°O and 34.625°O (Fig. 1). Paraíba State has a rectangular shape and is subdivided into four administrative mesoregions, namely Mata Paraibana, Agreste Paraibano, Borborema and Sertão Paraibano (Fig. 1). Its rectangular shape influences different factors that interfere with the circulation of winds and the climate of the region; standing out among these factors is the proximity to the Atlantic Ocean, the existence of plateaus, the mountain range, and depressions. Details about Paraíba State can be found in Santos et al. (2019a), Santos et al. (2019b) and Brasil Neto et al. (2020).

2.2. Rainfall datasets

2.2.1. In-situ measurement data

Gauge-measured rainfall data for the period 1998 to 2017 were provided by the Agência Executiva de Gestão de Águas (AESA). Although there are 251 rainfall stations across the region, a prior analysis of rainfall network quality and time-series consistency over Paraíba State identified limitations (Brasil Neto et al., 2020). For the present analysis, all stations with missing data were therefore excluded, resulting in 78 complete series of daily data which were then accumulated at a monthly level for SPI calculation. More details regarding the qualitative and quantitative analysis of the available data can be found in Brasil Neto et al. (2020).

2.2.2. Estimated rainfall dataset

To carry out drought monitoring using complete and equally distributed satellite estimated rainfall data over Paraíba State, the Tropical Rainfall Measuring Mission (TRMM) datasets were used. TRMM was a 17-year joint mission between the American Space Agency (NASA) and the Japanese Space Agency (JAXA) (Huffman et al., 2007; Liu et al., 2012). Launched at the end of 1997, the TRMM satellite was developed to monitor rainfall in tropical regions but suffered technical problems around 2014 and started to fall slowly while continuing to collect data (Xia et al., 2018). The mission has published critical datasets, including the TRMM Multi-satellite Precipitation Analysis (TMPA) (Huffman et al., 2010), which is a product that combines the precipitation data estimated by the TRMM satellite and remote sensing measurements from multiple satellites with the available observations of rain gauge for bias correction.

TMPA products cover extensive space domains, between latitudes 50°N and 50°S and longitudes 180°W and 180°E, with a refined spatial resolution of 0.25° × 0.25°, allowing the monitoring of rainfall in various areas of the globe (Zhao et al., 2018). In Paraíba State, several studies have used TMPA estimates, and the results indicate that these estimates are a viable alternative to conventional rainfall measurement for resources management purposes (Soares et al., 2016; Santos et al., 2019a; Santos et al., 2019b; Brasil Neto et al., 2020). In this work, data from TMPA 3B42v7 were used, hereafter called TRMM, and the study

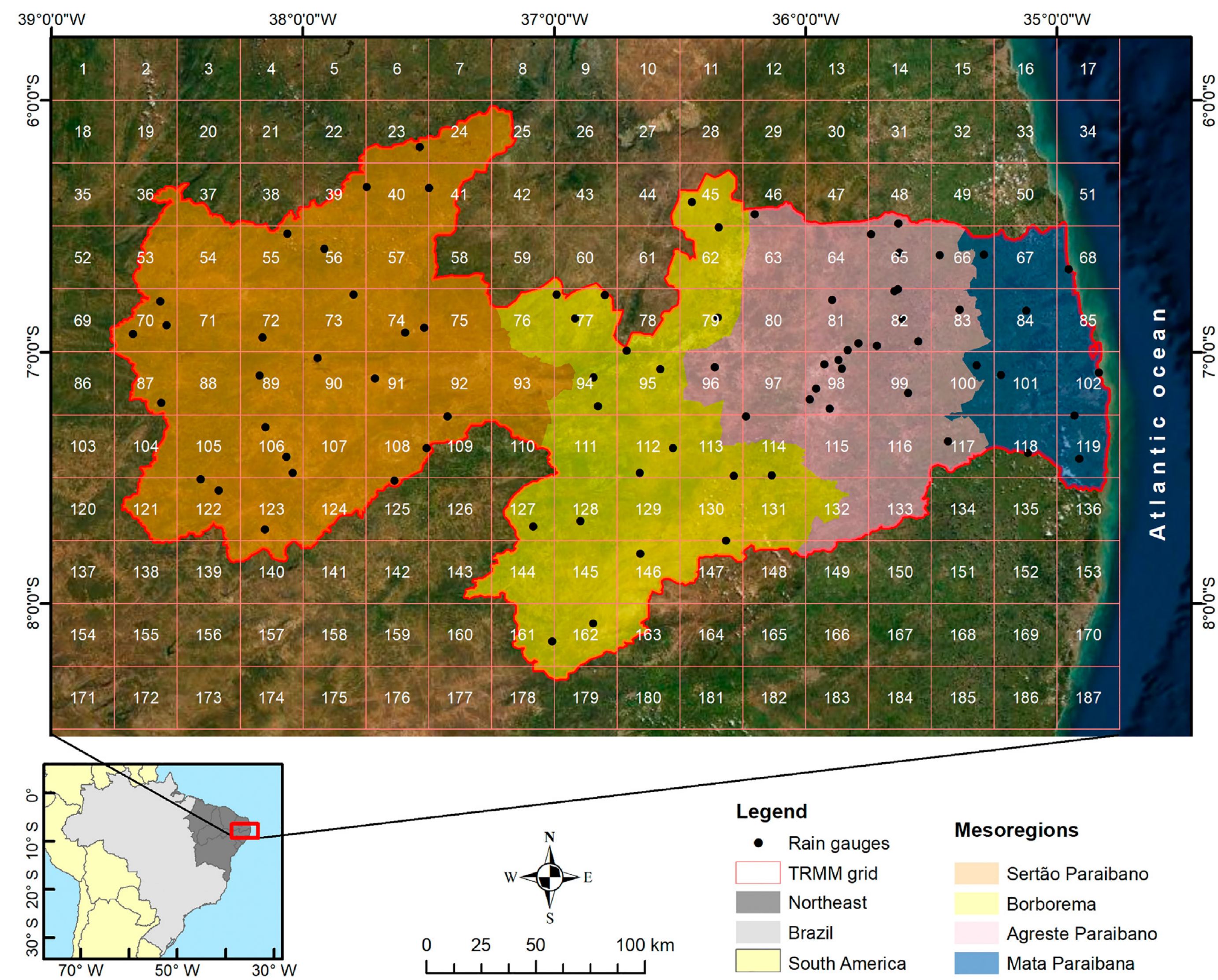


Fig. 1. Location of Paraíba State, the spatial distribution of the TMPA grid, the selected rain gauges, and its mesoregions.

area was divided into 187 grids (11 × 17). Fig. 1 shows the spatial distribution of the TRMM cell grids and the rain gauges used in this work. The daily rainfall time series were accumulated at a monthly level from January 1998 to December 2017, totaling approximately 45,000 monthly rainfall data points (187 TRMM-estimated series × 20 years × 12 months) estimated by satellite.

2.3. SPI: Run Theory and time series development

The drought analysis from January 1998 to December 2017 was based on eight SPI multitemporal scales, i.e., (a) short-term droughts: SPI-1, SPI-3 and SPI-6, (b) medium-term droughts: SPI-9 and SPI-12, and (c) long-term droughts: SPI-18, SPI-24 and SPI-48. All eight-time scales were calculated by adjusting the precipitation data to fit a gamma distribution of two parameters α and β. The SPI series were calculated for each of the 78 gauge-measured and 187 TRMM-estimated rainfall time series. Details regarding the calculation of the SPI index can be found in Santos et al. (2017). In this study, each drought event was characterized by the continuity of dry events, i.e., SPI ≤ 0, based on the premise of Run Theory (Yevjevich, 1967). The rationale for this is that anything less than 0 is drier than the median and therefore represents 'dry' rather than 'wet' conditions. Further division of 'dry' conditions into categories of drought is somewhat objective, and drought and dry events can be classified differently, depending on purpose or application (McKee et al., 1993; Liu et al., 2011; Azhdari et al., 2020; Bazrafshan et al., 2020). In this study, the classification proposed by Santos et al. (2017), Santos et al. (2019b), Brasil Neto et al. (2020) and Brasil Neto et al. (2021) was used, primarily for the reason of wanting to maintain consistency with this previous research.

The SPI is sensitive to the series size and using less than 30 years of data can have consequences, depending on the evaluated timescale. However, it is important to note that the time series of the rain gauges distributed across Paraíba State have many gaps and extending the period of analysis to increase series size would also mean reducing the number of rain gauge time series used as a reference, which introduces a new problem of under-representation. In addition, the behavior of the SPI time series tended to be the same when considering the SPI in different time scales ($R > 0.90$) (Brasil Neto et al., 2020), and the length of the time series did not make the study unfeasible. Furthermore, there were interesting and notable recorded drought events from 1998 to 2017, making this period a suitable choice for evaluating SPI, based on different data sources, as a meteorological drought monitoring tool.

Finally, Zamani and Bazrafshan (2020) highlighted that in computing the SPI, selecting an accurate distribution can be a basic and key step in estimation-desired index and drought monitoring. The goodness of fitness of gauge-measured and satellite-estimated rainfall data to the gamma distribution of two parameters α and β was carried out based on the Lilliefors test (Lilliefors, 1967) with $\alpha = 0.05$. In the present case, if the null hypothesis is rejected, the time series do not fit the gamma distribution, whereas if the null hypothesis is accepted, the precipitation data fit the gamma distribution. This study evaluated the adequacy of all available precipitation series, months, and time scales, totaling more than 25,000 analyses (265 time series × 8 time scales × 12 months). Fig. 2 shows the percentage of time series that do not fit the gamma distribution for each time scale.

For example, we detail the results of rain gauge-measured rainfall data for SPI-3 and evaluate the adequacy of these time series considering the accumulated quarterly rainfall in each rainfall month and gauge. White circles indicate that the data fit the gamma distribution, while red circles indicate that the temporal series does not fit this distribution. From this figure, the percentage of time series that do not fit the gamma distribution is small and that there is a variation between months and regions. In general, the results present in the graph indicate that in less than 10% of cases (i.e., 78 rain gauges × 12 months), the null hypothesis was rejected, which shows that, in general, the results are satisfactory, and the time series fit the gamma distribution.

With the exception of the results obtained for the SPI-48, the percentage of time series that did not fit the gamma distribution is less than 10%. It can be observed that as the time scale increases, the percentage of time series that do not fit the gamma distribution also increased. This must be closely related to the size of the time series being evaluated and was more evident when evaluating the series coming from the TRMM. Indeed, we recommend that longer data series be used whenever possible to improve the reliability of the results, but we believe that using this period (1998–2017) was not inappropriate, given that this dataset has been successfully used in other studies (Brasil Neto et al., 2020; Brasil Neto et al., 2021) and the objective of the study has not been undermined.

Fig. 3 illustrates the definition of a drought event and the behavior of the three time series evaluated in this study: (a) the drought state time series (SPI), (b) the drought duration time series (DDS), and (c) the drought severity time series (DSS). The DDS is the series that increases incrementally during a drought event, and the DSS is the result of the accumulated SPI values during the drought event. For these two series, when the events are no longer dry (i.e., SPI ≤ 0), the series are null. The drought state series, in turn, reflects the SPI values themselves over time. In the case of the example in Fig. 3, all three series are composed of 50 values.

2.4. Cluster analysis

Hierarchical cluster analysis techniques were used to divide Paraíba State into homogeneous regions based on drought state, duration, and severity. The analyses were performed for the eight time scales and considered the gauge-measured rainfall data and the satellite-estimated rainfall data, totaling 48 cluster analyses (2 databases × 3 types series × 8 time scales). The following section describes the basic steps of the hierarchical cluster analysis, such as choosing the metric of dissimilarity, the method of the linkage between clusters, and the optimal number of clusters (Keller Filho et al., 2005). To ensure the reliability of results, where relevant, the effect of different choices was evaluated against statistical criteria.

Pearson's linear coefficient was selected as the dissimilarity metric and calculated between the time series, considering that the time series will be grouped based on the similarity of their temporal variation. Thus, it was possible to evaluate how similar the state, duration and severity time series at multiple time scales are over time, which is important information to assess the influence of the weather phenomena active in the region, for example. Eq. (1) shows how the correlation distance between two different time series was calculated:

$$d = d(x_s, x_t) = 1 - \frac{\sum_{i=1}^{n} (x_s - \overline{x_s})(x_t - \overline{x_t})}{\sqrt{\sum_{i=1}^{n} (x_s - \overline{x_s})^2}\sqrt{\sum_{i=1}^{n} (x_t - \overline{x_t})^2}} \tag{1}$$

where d is the correlation distance between two time series x_s and x_t, $\overline{x_s}$ and $\overline{x_t}$ represent the averages of the historical series x_s and x_t that contain n data.

Then, we defined which linkage method was the most appropriate to perform the cluster analysis, and for that, the results were evaluated based on three different methods: single, complete, and average. The single method considers the distance between the clusters as the shortest distance between the elements (Eq. (2)); in the complete method, the largest distance between the components of different clusters is considered (Eq. (3)), and in the average method, the average of the distances between the series of cluster r with those of cluster s is considered (Eq. (4)). Eqs. (2)–(4) illustrates the difference between the calculation of distances between two different clusters:

$$D(r, s) = \min(d(x_{ri}, x_{sj})) \tag{2}$$

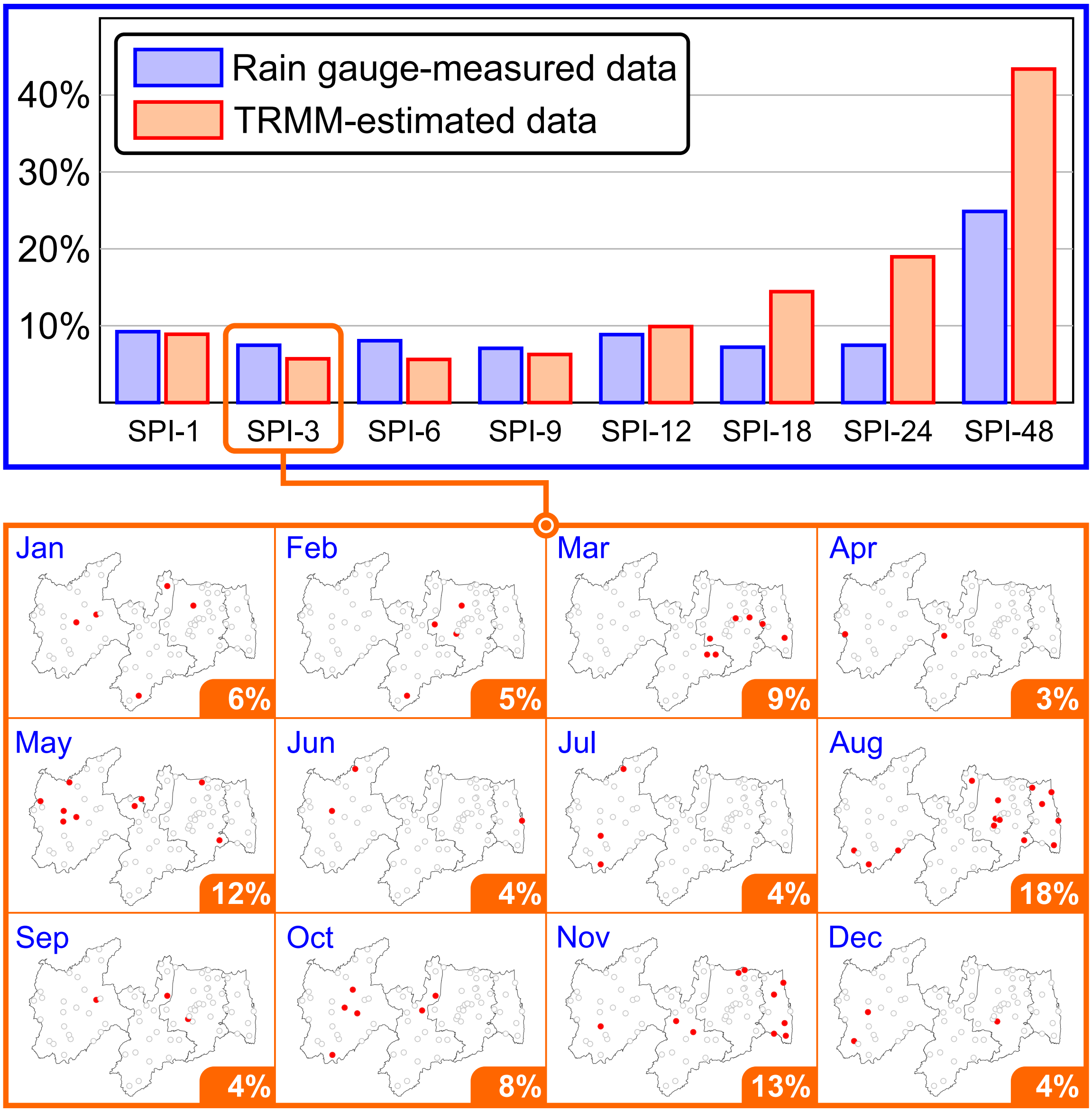


Fig. 2. Analysis of goodness of fitness to assess the fitting accuracy of gamma distribution over Paraíba State.

$$D(r,s) = \max(d(x_{ri}, x_{sj})) \tag{3}$$

$$D(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_{ri}, x_{sj}) \tag{4}$$

where $D(r,s)$ is the distance between clusters r and s, n_r is the number of components in cluster r, n_s is the number of elements in cluster s, d represents the metric of dissimilarity between the time series x_r and x_s, x_{ri} is component i of cluster r, and x_{sj} is element j of cluster s.

Additionally, the cophenetic correlation coefficient c was computed to assess the consistency and similarity of representativeness between the clusters. This coefficient measures the appropriateness of the choice of linkage method to perform the cluster analysis. The closer the value of coefficient c is to 1, the more appropriate the choice of the dissimilarity metric and the linkage method. In other words, since Pearson's linear coefficient was chosen as the dissimilarity metric, the cophenetic correlation coefficient c evaluated which linkage method (i.e., single, complete or average) was most appropriate for the clusters analysis. The cophenetic correlation coefficient was calculated according to Eq. (5):

$$c = \frac{\sum_{i=1}^{j} (x(i,j) - \bar{x})(t(i,j) - \bar{t})}{\sqrt{\sum_{i=1}^{j} (x(i,j) - \bar{x})^2 \sum_{i=1}^{j} (t(i,j) - \bar{t})^2}} \tag{5}$$

where $x(i,j)$ is the distance between the time series i and j based on the chosen dissimilarity metric and $t(i,j)$ is the dendrogram distance between the time series i and j based on the chosen method.

Finally, to define the optimal number of clusters to perform the regionalization, the silhouette method (Rousseeuw, 1987), the Calinski-Harabasz criterion (Calinski and Harabasz, 1974), and the variation curve of the distance between the clusters were used. Based on the variation curve of the distance between clusters by the number of clusters, it was assumed that the optimum quantity is equivalent to the number of clusters in which this variation curve remained constant. The idea of adopting this criterion is that when the derivative of this curve is practically null, there is no advantage in dividing the time series into several clusters because the variation between the clusters is not relevant. The silhouette method measures how similar the time series of a given

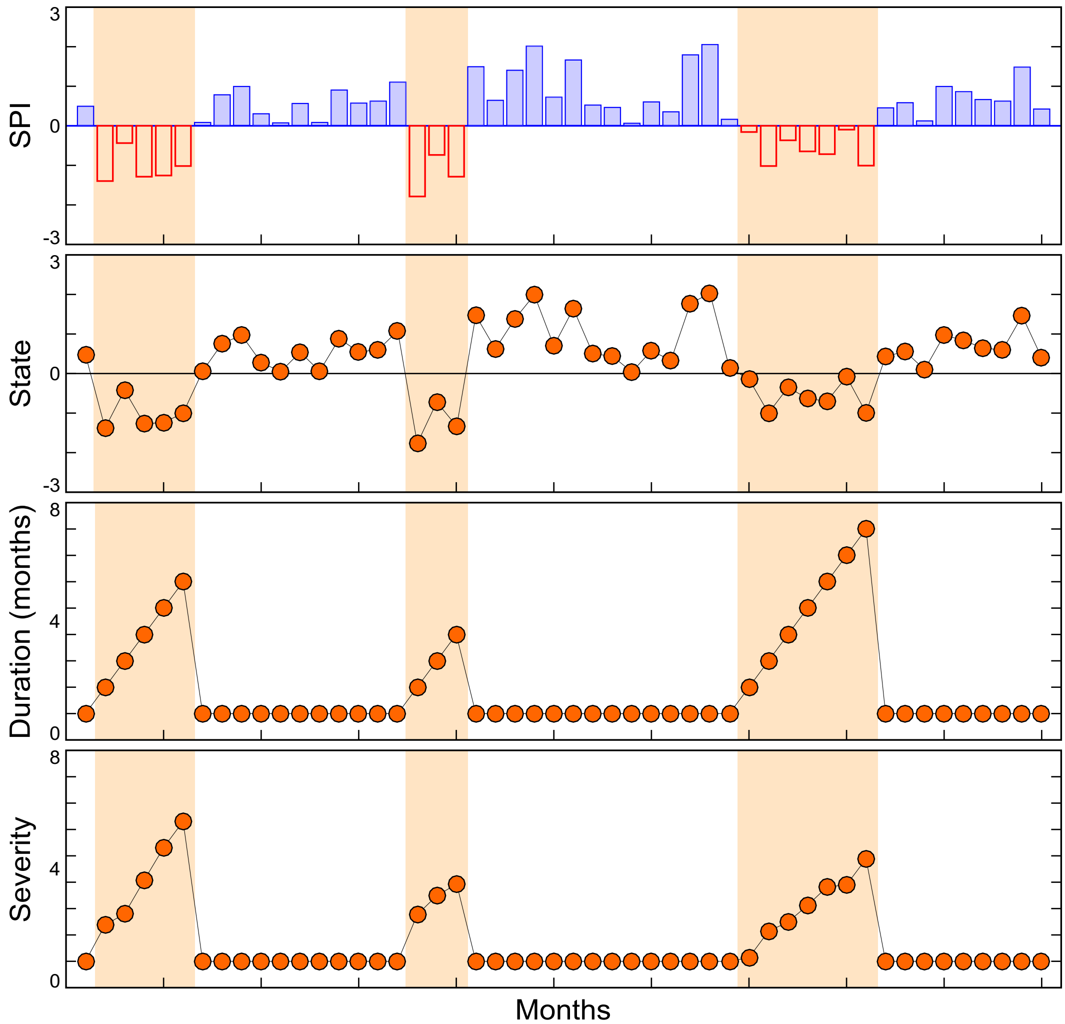


Fig. 3. Definition of a drought event and its main characteristics.

cluster are in relation to the original cluster when compared to the time series of other clusters. The final value of the silhouette is the average value among the all-time series, and the values range from −1 to 1, with the best result being 1 (Eq. (6)):

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{6}$$

where S_i is the silhouette value of time series i, a_i is the average distance from time series i to the time series of the origin cluster and b_i is the distance from series i to the time series that forms the other clusters.

The Calinski-Harabasz criterion, in turn, expresses the ratio between the variance between the different clusters and the variance within the different clusters. In general, based on this criterion, well-defined clusters present high variance between different clusters and small variance within clusters, and therefore, in the case of the *CH* ratio, the higher the value, the more adequate the cluster analysis was (Eq. (7)):

$$CH = \frac{SS_b}{SS_w} \times \frac{(N-k)}{(k-1)} \tag{7}$$

where SS_b is the variance between clusters, SS_w is the variance between clusters, N is the number of time series analyzed and k is the number of clusters.

3. Results and discussion

3.1. Definition of the linkage method

To avoid random choices, the best linkage method between the clusters was evaluated using the cophenetic coefficient *c* of state, duration and severity time series of droughts which was calculated for eight-time scales, three linkage methods and two databases (Fig. 4). Because of the variability of the results, it is worthwhile to provide a brief explanation of how to interpret them: the *c* calculated from the drought state time series based on the 187 TRMM cells grids was 0.80 when using the average distance as the linkage method and 0.76 when using single and complete distances, in the case of SPI-1 (see top left panel in Fig. 4).

From gauge-measured rainfall data, these values were 0.84, 0.83 and 0.60 when using the average, complete and single linkage methods, respectively. It is noteworthy that the results show high variability when considering all combinations and as for the variation between the types of drought time series, the results tend to vary according to the time scale. For short- (SPI-1, SPI-3 and SPI-6) and medium-term (SPI-9 and SPI-12) droughts, the values of *c* are higher for drought state time series, but for long-term droughts (SPI-18, SPI-24 and SPI-48), the duration and severity time series presented the best results, i.e., higher *c* values.

Regarding the variation of *c* values between the time scales, it is important that for the drought state time series, the best results were found in the case of medium-term droughts, but there is no great difference between these results and those found in the case of short- and

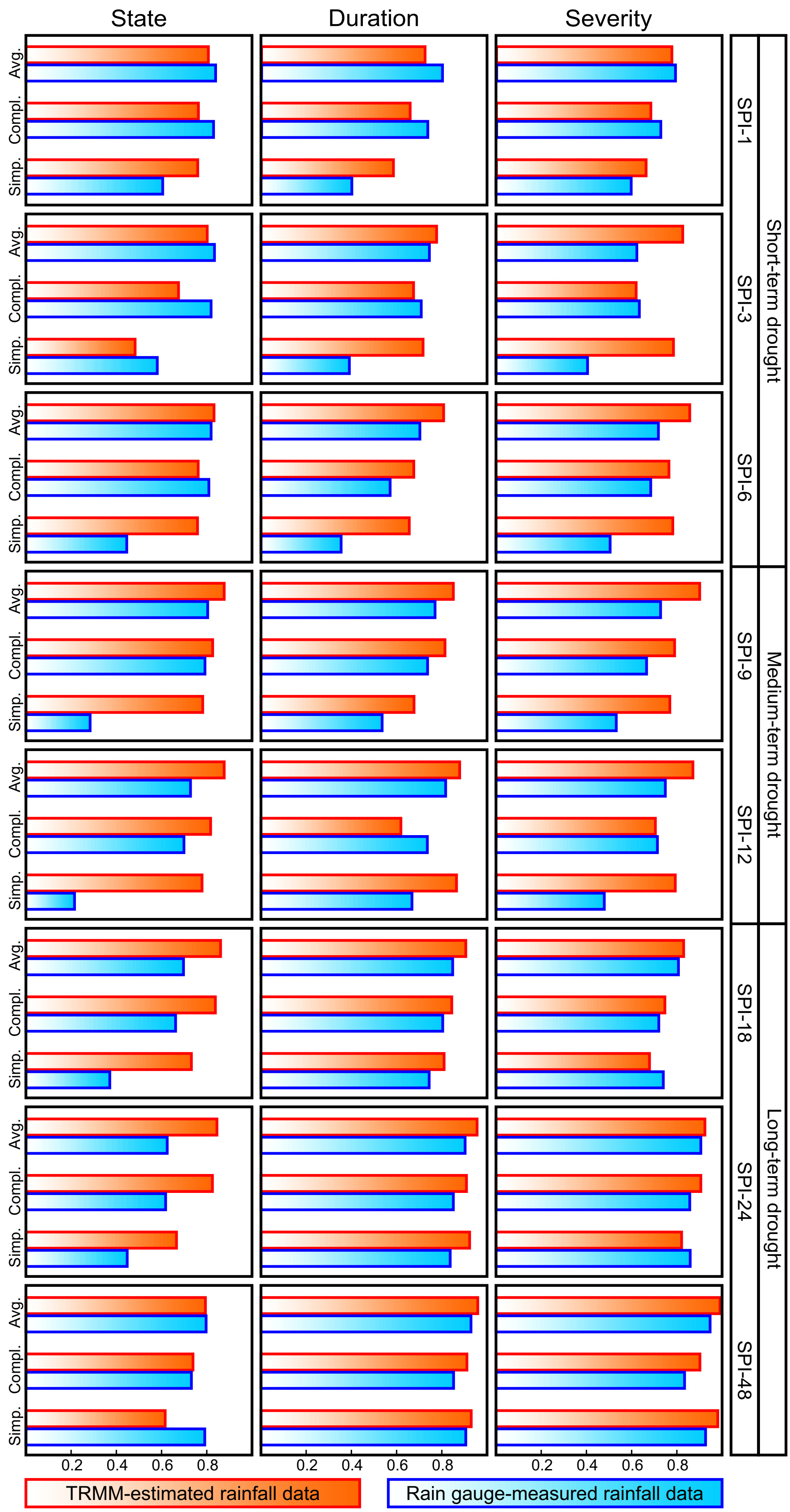
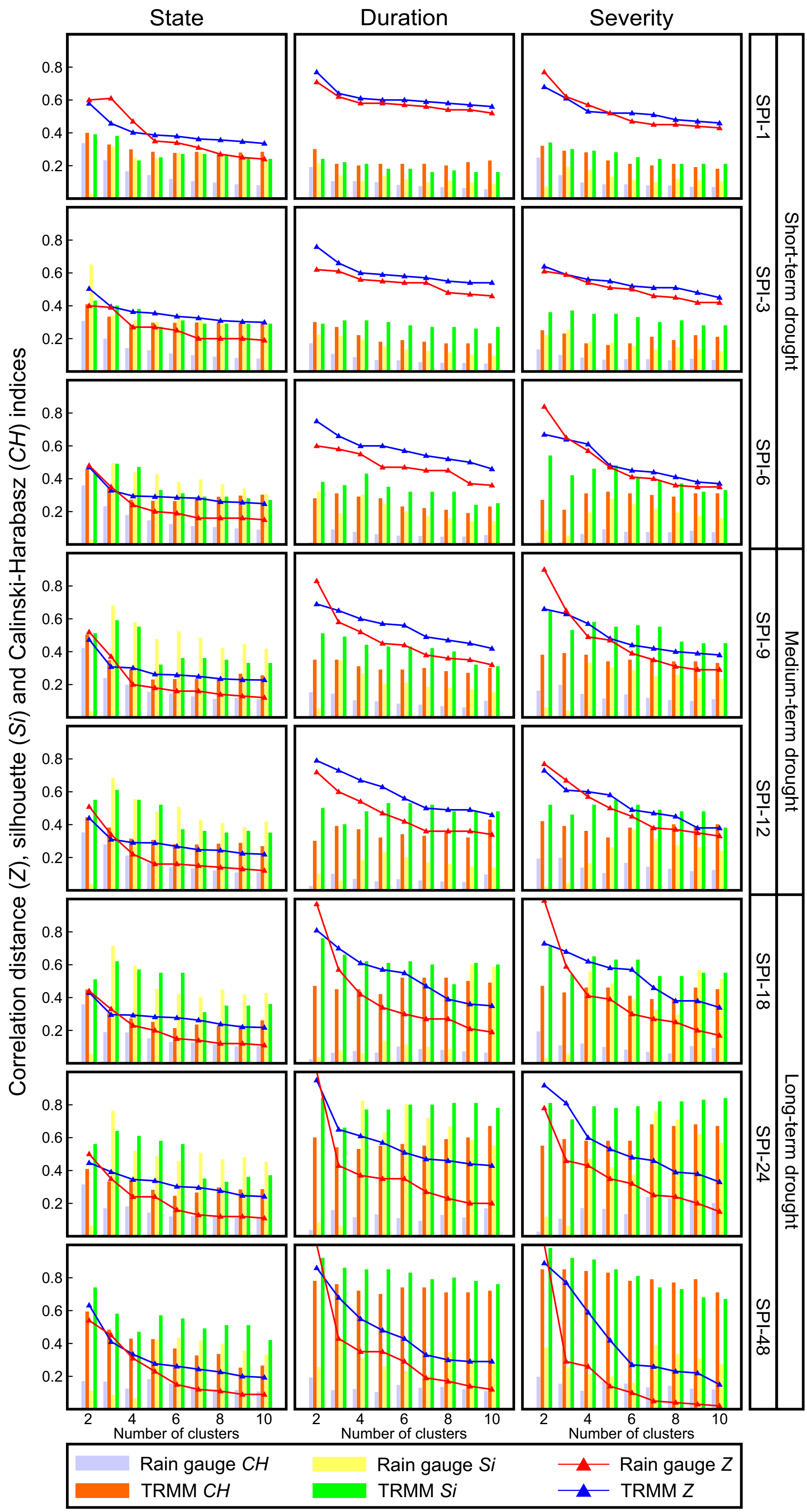


Fig. 4. Variation of the cophenetic correlation coefficient *c* for the time series of state, duration, and severity for each linkage method (simple, complete and average) at multiple time scales over Paraíba State (1998–2017).


State
Duration
Severity
SPI-1
SPI-3
SPI-6
SPI-9
SPI-12
SPI-18
SPI-24
SPI-48
Short-term drought
Medium-term drought
Long-term drought
Correlation distance (Z), silhouette (Si) and Calinski-Harabasz (CH) indices
0.2
0.4
0.6
0.8
2
4
6
8
10
Number of clusters
Rain gauge CH
TRMM CH
Rain gauge Si
TRMM Si
Rain gauge Z
TRMM Z

long-term droughts. On the other hand, when evaluating the duration and severity time series, the sensitivity of *c* is evident as a function of the change in the temporal scale, such that the results are better as the time scale increases. For long-term droughts, the values are more notable and indicate greater consistency in the analysis of clusters, while the results of short- and medium-term droughts are slightly worse.

In relation to the two rainfall datasets, it can be seen that making a direct comparison, i.e., same time scale, type of time series and linkage method, the *c* values obtained from TRMM-estimated rainfall data are predominantly higher than those calculated based on gauge-measured rainfall data. Furthermore, it indicates that the cluster analyses developed based on the TRMM estimates are more consistent than the analyses developed based on the rain gauge data. Finally, regarding the variation of coefficient *c* according to the linkage method, it is noted that the average linkage method presented the best performance.

In general, the results based on the complete linkage method performed moderately, while the poorest values occurred when using the single linkage method. It can be highlighted that there is a variation between the values of these two methods (i.e., single and complete), and these values depend on the combination of the series, time scale, or database used. In addition, it is noteworthy that the results based on the average linkage method were not as sensitive to these combinations. In other words, from the single linkage method, the coefficient *c* values calculated based on the gauge-measured rainfall data, for example, are 0.285, 0.535 and 0.535 for the SPI-9 state, duration, and severity time series. However, when evaluating this result considering the SPI-48, the values exceed the order of 0.800, which shows considerable variability.

Contrary to what occurred when using the single and complete linkage method, there is consistency in the correlation coefficient values between the time scales and types of drought time series when using the average linkage method. These results corroborate with Unal et al. (2003), who concluded that the average linkage method could fill gaps of other methods as it can minimize the variance within the series of the same cluster and maximize the variance between the different clusters. Several studies have been carried out based on this linkage method, and the results have been extremely satisfactory, although the purpose in these instances was to regionalize different areas based on the precipitation regime (Santos et al., 2019a) or according to the drought pattern (McGree et al., 2016; Shiau and Lin, 2016; Yang et al., 2017).

3.2. Definition of number of clusters

Based on the average linkage method, Fig. 5 shows the relationship between the correlation distance between clusters, silhouette method and the Calinski-Harabasz criterion with the number of clusters for drought state, duration, and severity time series over Paraíba State (1998–2017). These results help to define the optimal number of clusters to develop an efficient cluster analysis for the region. The results based on satellite-estimated data presented shorter Pearson correlation distances, indicating greater similarity between these time series. The variation curve of the distances between the clusters by the number of clusters related to TRMM is, in most cases, below that obtained from the gauge-measured rainfall data. This difference is smaller for short-term droughts but increases when assessing medium- and long-term droughts and the duration and severity time series.

Calinski-Harabasz values tend to be higher when using TRMM-estimated data, regardless of the type of drought time series, time scale, or the number of clusters, and this scenario only changes in some cases, e.g., in the case of SPI-9 for drought state, duration and severity time series. In addition, there is a similarity between the silhouette's values with those of *CH,* and these values were predominantly higher when using TRMM-estimated data, especially for medium- and long-term droughts. It is noteworthy that the best results for *CH* and *Si* were obtained for less than five clusters.

Comparing the three drought time series results, the distances between the clusters are shorter for the drought state time series and longer for the duration and severity time series. For SPI-1, when evaluating four clusters, the distance between groups is about 0.45 for the drought state time series, but for the duration and severity time series, the distance is 0.55. For the SPI-12, the pattern was even more evident because after grouping the drought state time series into four clusters, with amplitudes ranged 0.25, 0.60 and 0.60 for the state, and duration and severity time series, respectively. The results show that regardless of the time scale or database, the drought state time series can be considered more homogeneous with each other than the duration and severity time series.

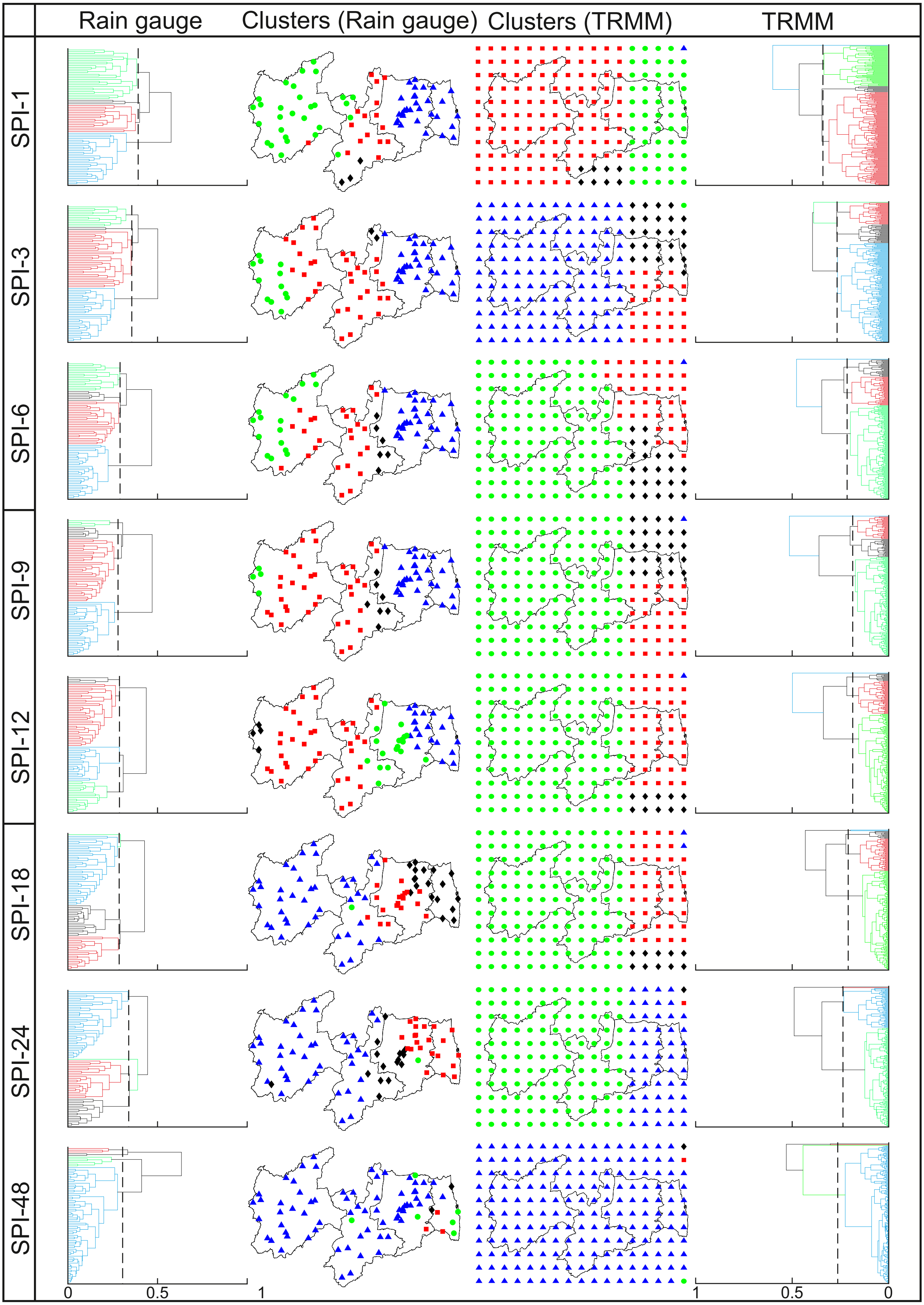
Concerning *CH* values, the results indicate that the values tended to be higher for the drought state time series for short- and medium-term droughts. Also, the *CH* values obtained from the TRMM-estimated data had greater variability between the types of drought time series compared to the results obtained from the rain gauge-measured data. Regarding the silhouette method, there is no significant variation between the results for the same time scale, except for what was obtained when evaluating long-term droughts. In this case, the duration and severity time series values were higher than those found when evaluating the drought state time series, as found for the *CH* (Fig. 5).

For the drought state time series, there is a kind of stability in the distance values between the clusters for a small number of groups, mainly for short- and medium-term droughts. This means that the curves in the figure relate the distance between the clusters, and their number becomes less steep (almost constant) from four clusters, indicating that it is unnecessary to divide the TRMM cells grid or rain gauges into more groups. For example, for the SPI-3 state time series, the distances between four clusters are 0.25 based on TRMM-estimated data and 0.35 based on rain gauge-measured data (Fig. 5), whereas for ten clusters, the distances are almost the same, i.e., 0.30 based on TRMM-estimated data and 0.20 based on rain gauge-measured data.

Therefore, as the distance between the clusters has been subtly altered, it is irrelevant to divide the TRMM cells grid and the rain gauges into more groups. In contrast, when developing the same analysis for the SPI-48, the distances between the clusters using four and ten groups already vary from 0.40 (TRMM) to 0.20 (rain gauges), which shows a greater jump in the curve compared to the results presented for short- and medium-term droughts. For duration and severity time series, this pattern was maintained, and for short- and medium-term droughts, the values of correlation distance stabilized with a smaller number of groups. For long-term droughts, there was a greater distinction between the series, and the stability of the curve was found with a higher number of clusters.

Using these three methods (i.e., variation curve, *CH* and *Si*) to define the number of clusters, it is noted that using few clusters is the adequate alternative in most cases. In this sense, we adopted four as the number of clusters to perform the analysis, which is also the number of territorial divisions over Paraíba State. For this quantity, the values of the silhouette and the *CH* criterion are expressive, as well as the variation curves tend to be less steep, which makes a choice effective from the perspective of different methods. Although there are cases in which this quantity is not so appropriate, it should be noted that the choice was based on the results of the 48 cluster analyses. In addition, the capacity of TRMM-estimated data to reproduce the same pattern of results as rain gauge-measured data stands out.

Fig. 5. Relationship between the correlation distance between clusters (*Z*), the silhouette criterion (*Si*) and the Calinski-Harabasz criterion (*CH*) with the number of clusters for the time series of state, duration, and severity of droughts over Paraíba State (1998–2017).


Rain gauge
Clusters (Rain gauge)
Clusters (TRMM)
TRMM
SPI-1
SPI-3
SPI-6
SPI-9
SPI-12
SPI-18
SPI-24
SPI-48
0
0.5
1
1
0.5
0

3.3. Analysis of drought state time series

After defining the dissimilarity metric (i.e., Pearson correlation coefficient), the linkage method (i.e., average linkage method), and the optimal number of clusters (i.e., four), the distribution process of the clusters was carried out based on the time series of state, duration, and severity of droughts over the region. Initially, Fig. 6 shows the results of the hierarchical cluster analysis developed for the drought state time series based on gauge-measured and TRMM-estimated rainfall data for different time scales. A variation between the results obtained when considering the different SPI indices and the rainfall datasets was observed.

For short-term droughts, there is a correspondence between the spatial distribution of clusters and the mesoregions of Paraíba State (Fig. 1). Based on gauge-measured rainfall data, the results indicate that at a distance of 0.50 between clusters, a group encompasses the mesoregions of Sertão Paraibano and Borborema (♦ ■ ●), while Agreste Paraibano and Mata Paraibana are covered by another cluster (▲). In the case of SPI-3, Sertão is bisected into an area to the west (●) and another to the east (■), with the latter extending towards the coast of Paraíba and covering the entire Borborema to the west of Agreste Paraibano. In the case of SPI-6, the difference in the behavior of the clusters on the border between Borborema and Agreste Paraibano becomes clearer (♦).

When assessing the dendrograms, the pattern in the inland of Paraíba State is more heterogeneous than in the regions closest to the coast. In SPI-3 and SPI-6, the rain gauges located in Agreste Paraibano and Mata Paraibana only differ from each other at a distance of less than 0.25, while at a distance of 0.30, there are already two clusters dividing the inland of Paraíba State. Based on TRMM-estimated data, the time series tend to be more homogeneous when compared to the results obtained from the rain gauge-measured data and the results show a good correspondence between the spatial distribution of the clusters and the mesoregions of Paraíba (Fig. 1).

Still based on TRMM-estimated data, it is noteworthy that, except for the SPI-1 results, where there was a distinction between time series in Sertão Paraibano (■) and south of Borborema (♦), these two mesoregions were always grouped in a cluster, differently from the results found based on rain gauge- measured data. These results indicate that there is greater variability between the drought state in the mesoregions of Agreste Paraibano and Mata Paraibana. It is noted that although the TRMM satellite has shown an accuracy in separating the regions of Sertão and Borborema from the mesoregions of Agreste and Mata Paraibana, there was a certain inaccuracy when estimating which of these two mesoregions were more homogeneous with each other.

For medium-term droughts, the distances between four clusters were the shortest among the time scales, indicating greater similarity between the series over the region. Based on rain gauge-measured data, the division of Paraíba State between the mesoregions of Sertão Paraibano and Borborema, and the region of Agreste Paraibano and Mata Paraibana became increasingly evident. A cluster is formed at the western Sertão Paraibano, and the other part of this mesoregion is covered by another cluster (■) that extends to Agreste Paraibano. In Mata Paraibana, the behavior is the same for SPI-9 and SPI-12, and the existence of only one cluster over the entire area is noted (▲).

Based on TRMM-estimated data, the distances found between the clusters are smaller than those found based on rain gauge-measured data. At a correlation distance of 0.40, the TRMM series are grouped into a large group that covers the entire state, which differs from the results based on rain gauge-measured data. Despite the differences, the results obtained from the TRMM-estimated data demarcate the division of Paraíba State into two major regions: one located in the interior and formed by Sertão and Borborema (●), and the other on the coast of Paraíba State, standing out that the region close to the coast was divided into two zones: one located in the north and the other in the south.

It is noteworthy that TRMM estimates did not identify the particularities of western Sertão Paraibano and Agreste Paraibano and Mata Paraibana. However, just as for short-term droughts, it identified that clusters located in the inland of the state are more homogeneous than the clusters near the coast. Finally, the results show a high variation for long-term droughts in relation to the results of short- and medium-term droughts, especially in the case of SPI-48. Based on rain gauge-measured data, the spatial distribution of the clusters for the SPI-18 and SPI-24 indices are similar, but the results of the SPI-48 have a more particular pattern.

For SPI-18 and SPI-24, the regions of Borborema and Sertão Paraibano showed less similarity to each other despite being mostly covered by a cluster (▲). From the border between Borborema and Agreste Paraibano to the central portion of this mesoregion, there is a cluster, while from the center of Agreste to the coast, there is another. For the SPI-48, the behavior is more intriguing, and the reason is that there is a predominance of a group (▲) in all mesoregions of the state, covering from the Sertão Paraibano to the coast. In Mata Paraibana, four different clusters could be found, highlighting the variability of the drought pattern in this region.

Based on the TRMM-estimated data, the division of Paraíba State based on SPI-18 and SPI-24 into two regions is clearer: one formed by Sertão Paraibano and Borborema (●) and the other by Agreste and Mata Paraibana (♦ ■ ▲). From the dendrograms, the clusters are more similar to each other when dealing with the Sertão Paraibano and Borborema and that these will only start to differentiate at a correlation distance of 0.15. When evaluating the SPI-48, almost all TRMM grids are grouped in a cluster (▲), as well as in the results obtained based on rain gauge-measured data. This indicates that for the long-term drought state time series, there is a high similarity pattern between the series over the region.

One of the possible explanations for the distances between the clusters in the case of short-term droughts to be so high is that as the behavior of these SPI series is very variable, it is expected that the time series have less similarity. In other words, any disturbance in the precipitation time series can cause an extremely dry or wet SPI value to appear, and this can lead to differences in the similarity between the time series. Evaluating medium- and long-term droughts, except for rare cases, the time series tend to behave in the same way due to the accumulation of precipitation over time, which makes the series have high similarity.

The pronounced similarity of the drought state time series based on TRMM-estimated data can be linked to the algorithm employed by the mission, which can tend to compensate for the precipitation values between the regions. In this case, as the drought state time series are grouped according to the similarity of the SPI variation over time, the compensation may have made the series more similar to each other. When using rain gauge-measured data, on the other hand, point variations are captured in a more particular way, and this increases the dissimilarity between the state time series from the rain gauges.

3.4. Analysis of drought duration time series

Fig. 7 shows the result of the hierarchical cluster analysis for the drought duration time series based on gauge-measured and TRMM-estimated rainfall data for different time scales. It is noteworthy that the spatial distribution of the clusters based on the drought duration time series differs in some situations from the configuration obtained when evaluating the drought state time series (Fig. 6). This result is relevant because it shows that a given rain gauge (or TRMM grid) may be highly similar to another concerning the drought state time series but

Fig. 6. Analysis of hierarchical cluster and its dendrograms using four clusters based on the time series of drought state over Paraíba State (1998–2017).

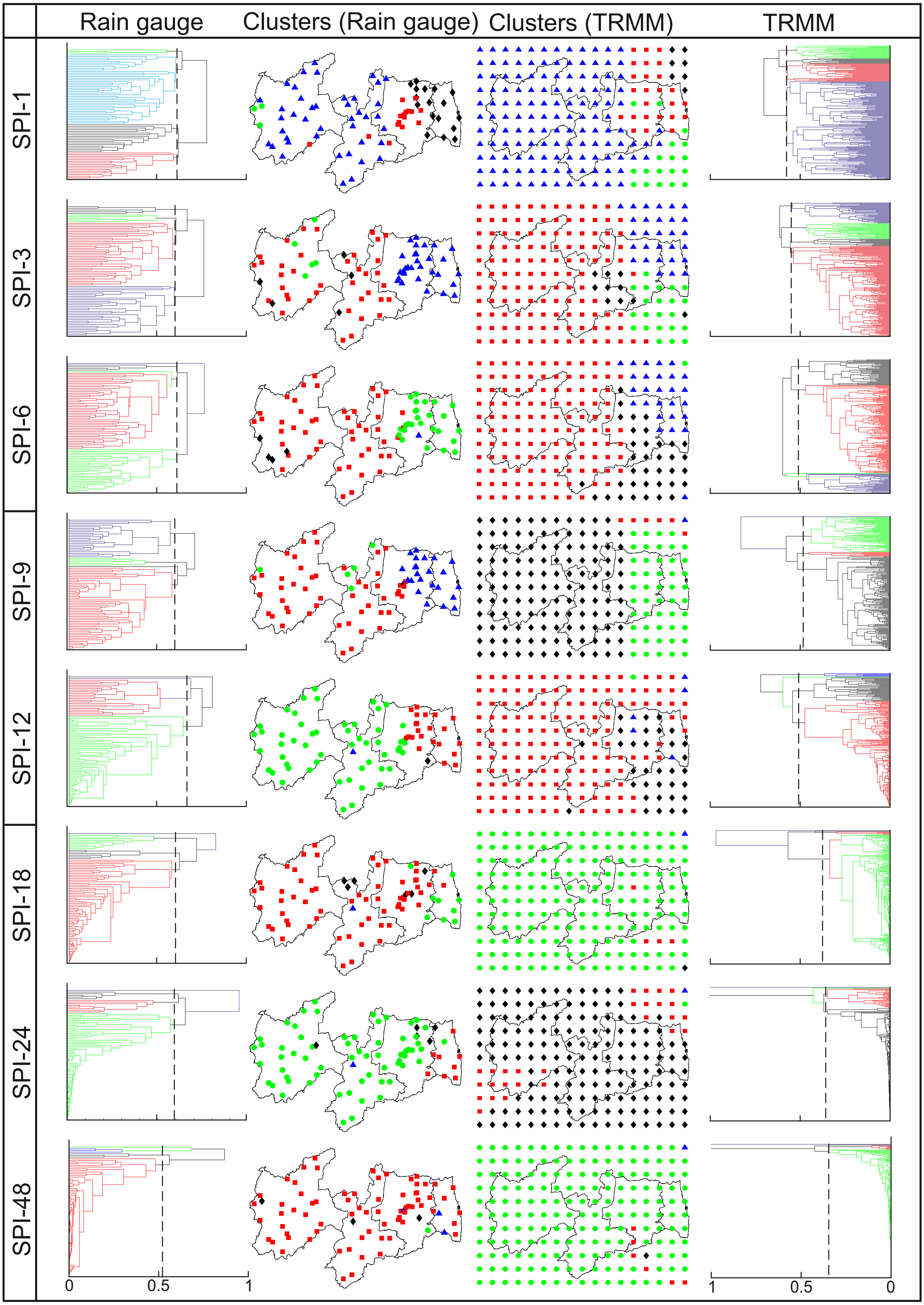


Fig. 7. Analysis of hierarchical cluster and its dendrograms using four clusters based on the time series of drought duration over Paraíba State (1998–2017).

differ when assessing the drought duration time series. For short-term droughts, this is more evident when considering the gauge-measured data. For SPI-1, Sertão Paraibano and Borborema are now mostly covered by only one cluster (▲), while Mata Paraibana and Agreste Paraibano are formed by two distinct zones, one to the east (♦) and the other to the west (■).

For SPI-6, it can be noted that the Sertão Paraibano and Borborema were grouped in a cluster (■), while Mata Paraibana and Agreste Paraibano were formed mostly by another one (●). For SPI-3, the distribution of clusters is more irregular compared to the pattern of SPI-1 and SPI-6: the Sertão is divided into three regions, one in the southwest (♦), the other in the center (■), and another in the northeast (●), while Agreste and Mata Paraibana are composed by single cluster (▲). This result differs from that found in Fig. 6, where the grouping was more consistent with the limits of the mesoregions of Paraíba State, especially in the case of the Sertão and Borborema mesoregions.

However, it is interesting to note that there are indications that the regions of Mata and Agreste Paraibano continue to have homogeneous behavior between them, while Sertão and Borborema have a pattern of higher dissimilarity. Moreover, it is important to emphasize that the mesoregions of Mata Paraibana and Agreste Paraibano are no longer so similar between themselves when compared to the similarity that exists between the Sertão Paraibano and Borborema, a fact that differs from the pattern that was obtained when evaluating the drought state time series.

From the TRMM-estimated data, changes were noted regarding the spatial distribution of clusters over Paraíba State, and in general, what is noticeable is that the regions of the Sertão and Borborema tend to have a large part of their territory composed of a cluster. In addition, in comparison to the cluster analysis of the drought state time series, it is clear that the spatial distribution of the clusters was maintained, especially in the case of SPI-6. In the case of SPI-1, there was a greater variation in the distribution of clusters in Agreste and Mata Paraibana, which started to be divided into three clusters (♦ ■ ●); in the case of SPI-3, a cluster appeared between Agreste Paraibano and Borborema (♦).

For the SPI-6, it is worth noting that there is a greater similarity between the clusters that cover the mesoregions of Sertão, Borborema and Agreste (♦ ■). The other cluster covers the north of Agreste and a large part of Mata Paraibana (▲) and has a unique behavior compared to the others. This result shows that, in contrast to the state time series of short-term droughts, for the duration series of the SPI-3 and SPI-6 indices, there is evidence to zone most of Paraíba State in a single cluster, which is what occurs when evaluating the long-term drought state time series.

For medium-term droughts and based on rain gauge-measured data, there is a similarity regarding the spatial distribution of the clusters over Paraíba State compared to the results in Fig. 6, but some differences should be noted. For SPI-9, the region of Sertão Paraibano and Borborema start to be composed by a single cluster (■), and this cluster crosses the border between these mesoregions and extends to the center of Agreste Paraibano. From this region to the coast, another cluster (▲) is formed that covers Agreste Paraibano and Mata Paraibana. For SPI-12, the change in the spatial distribution of the clusters intensifies such that from the Sertão to the central portion of Agreste, there is a more representative cluster (●), and from Agreste to the coast, there is another cluster (■), as well as in the case of SPI-9.

Thus, differently from the results shown in Fig. 6, Paraíba State started to be divided into less representative clusters. However, from the dendrograms, it can be seen that the interior mesoregions showed greater dissimilarity between them when compared to the pattern of the mesoregions closer to the coast, as found for the drought state time series. Based on TRMM-estimated data, the division of Paraíba State into two major regions is even more evident, especially when evaluating SPI-9. In a way, these results corroborate those found by using rain gauge-measured data, considering that the formation of two main clusters on Paraíba State was identified.

The results indicate that the Sertão Paraibano and Borborema are more homogeneous among themselves than the regions of Agreste and Mata Paraibana, as obtained in the analysis of the drought state time series. Moreover, the dendrogramic distances between the clusters over Paraíba State have high stickiness when evaluating long-term droughts, but these are the smallest when only four clusters are evaluated. This implies that when considering a distance between the clusters of 0.80, the short- and medium-term drought duration time series would be grouped into a single cluster, while when evaluating SPI-18, SPI-24 and SPI-48, the time series obtained based on rain gauge-measured data and satellite-estimated data are subdivided into at least two clusters, which highlights the heterogeneity of long-term droughts.

On the other hand, at the level of four clusters, these distances are not pronounced, and this shows that the rain gauges and the TRMM grids have unique behavior, which distinguishes the clusters with a high dissimilarity. Based on rain gauge-measured data, the results obtained for SPI-18 and SPI-24 are more similar, while those for SPI-48 have particularities. In general, for SPI-18, the Sertão, Borborema and Agreste are covered by a cluster (■), while Mata Paraibana is composed of another (●), results similar to those obtained for the SPI-24. For SPI-48, one cluster covered a large part of the state (■), and the others were concentrated in the center-south portion of Agreste, as well as in the case of Fig. 6.

Based on TRMM-estimated data, Paraíba State is covered by the same cluster, and this applies to the SPI-18, SPI-24 and SPI-48 indices. In detail, except a grid in the center of Agreste (■) for SPI-18, from the southwest region of Sertão and north of Mata Paraibana (■) for SPI-24 and from south of Agreste (■) for SPI-48, all regions showed the same pattern of variation as to the drought duration time series over time. Although there are differences, TRMM-estimated data identified that there is basically a cluster over Paraíba State.

3.5. Analysis of drought severity time series

Finally, Fig. 8 shows the result of the hierarchical cluster analysis for the drought severity time series based on gauge-measured and TRMM-estimated rainfall data for the different time scales. In general, there is a high similarity between the results of the cluster analysis of the drought severity and duration time series (Fig. 7). It is worth noting that for short-, medium- and long-term drought severity time series, the distances between the first two clusters based on TRMM-estimated data are greater than those obtained from gauge-measured data, and this result occurred systematically when evaluating SPI-1, SPI-6, SPI-9, SPI-18 and SPI-48.

Based on gauge-measured rainfall data, it is noted that for SPI-1, Paraíba State was divided into two clusters: one of which covered much of the Sertão and Borborema (●) and the other the Mata and Agreste Paraibano (▲). The spatial pattern of the clusters in the case of SPI-3 was similar to SPI-1, and the division of the state into two main regions is evident, with the central region of Borborema (▲) starting to behave a little more distinctly from the interior of the state. As for SPI-6, the center-west of Agreste also started to behave more similarly to Sertão and Borborema (■) and more dissimilar to regions close to the coast. Based on satellite-estimated data, the regions of the Sertão Paraibano and Borborema were covered by a single cluster, i.e., SPI-1 (●), SPI-3 (●), SPI-6 (■), while that Agreste and Mata Paraibana showed greater variability when compared to the results of the drought duration time series.

In the case of medium-term droughts, the distances between the clusters have the same order of magnitude as the drought duration time series. Based on rain gauge-measured data, there is once again a high similarity between the results obtained in Figs. 6 and 8. For SPI-9, there is a small change in the north and central portion of the Agreste, but the same regionalization when evaluating the drought duration and severity time series. For the SPI-12, two main clusters stand out over the region: one located from Sertão to Agreste (■) and another

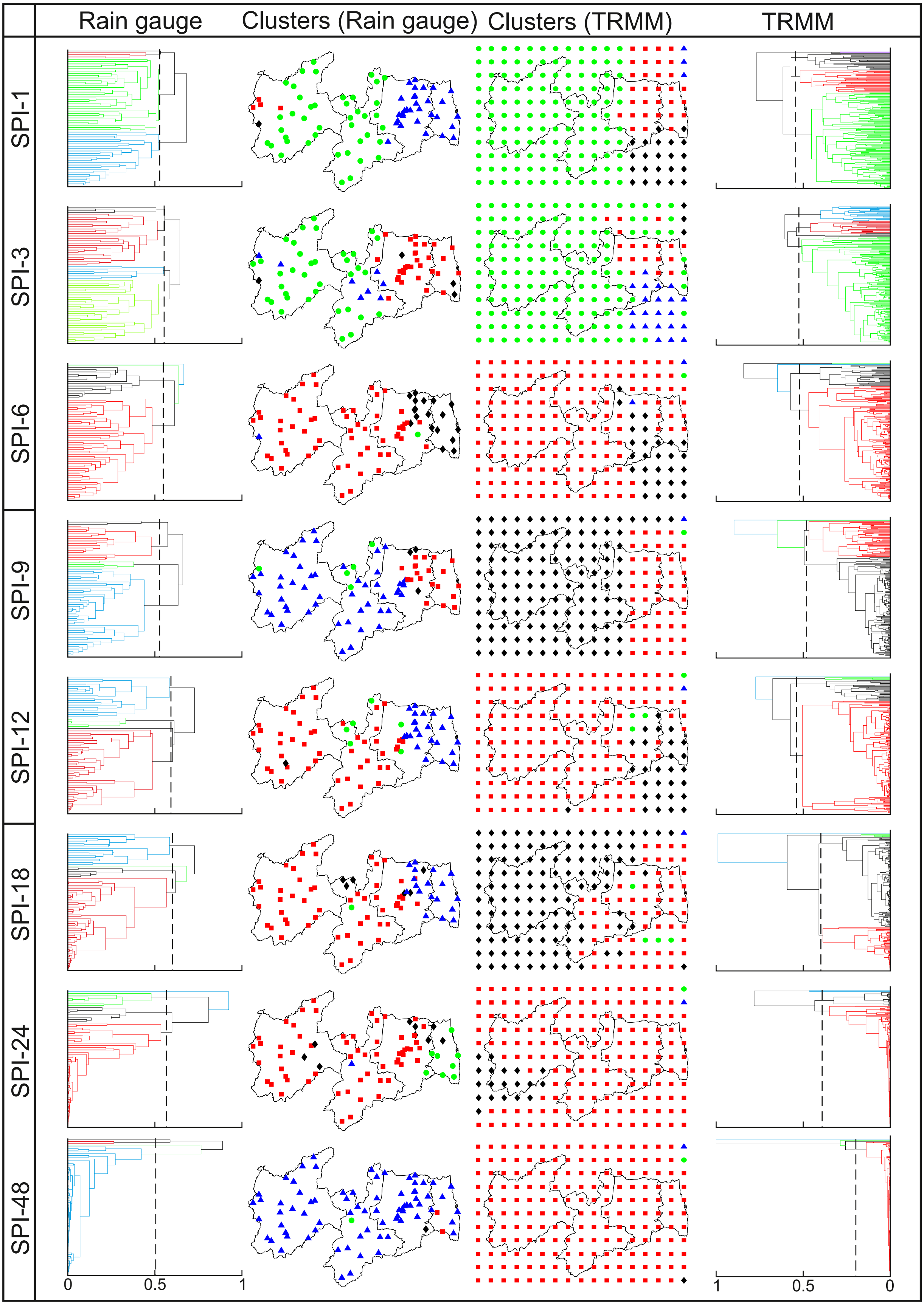


Fig. 8. Analysis of hierarchical cluster and its dendrograms using four clusters based on the time series of drought severity over Paraíba State (1998–2017).

from Agreste to the coast (▲). Regarding the results based on TRMM-estimated data, for both SPI-9 and SPI-12, Paraíba State was divided into two major regions.

The region closest to the coast has greater variability than the interior, especially in the case of SPI-9. When comparing the results between the datasets, clustering using TRMM clearly subdivided Paraíba State into two distinct homogeneous zones for SPI-9. This was less clear from gauged data and the most significant difference within the two datasets was to the north of the Borborema zone, as this zone was incorporated into the cluster that covers the interior of the state (♦). For long-term droughts, it is noted that based on gauge-measured data, for the SPI-18, Paraíba is covered by two clusters, one that runs from the Sertão to the center of Agreste (■) and the other that goes from the east of Agreste to the coast of Paraíba (▲).

For SPI-24, there are two distinct clusters in the vicinity of Mata Paraibana; one is to the north (♦) and another to the south (●). For the SPI-48, as well as for the results of the drought state and duration time series, a cluster covers the entire Paraíba State (▲). In general, the increase in the time scale of the SPI indices causes the entire state to behave in the same way concerning the drought severity time series. Based on the TRMM-estimated data, Paraíba tends to be grouped into a single cluster. For SPI-18, however, this pattern is not so evident, and one can see the existence of a cluster that covers the Sertão, the north and southwest of Borborema (♦), and another that extends from the central portion of Borborema to the coast of Paraíba (■). For SPI-24, except for southwestern Sertão (♦), all regions are members of the same cluster (■), and this is repeated when evaluating SPI-48.

4. Discussion

Earlier research has mapped Paraíba State into different homogeneous regions based on the pluviometric regime, and the results provide interesting points of comparison. Keller Filho et al. (2005), for example, delimited about 25 homogeneous rainfall zones over Brazil, of which ten are in Northeast Brazil, and four are over Paraíba State. The results indicate that one of the zones is located on the coast of Paraíba, another covers a large part of Agreste Paraibano and the others cover the areas of Borborema and Sertão Paraibano. Reboita et al. (2010) suggested that Paraíba State could be categorized into two main regions: one close to the coast and the other in the Northeastern Sertão.

The coastal region, which presents the Intertropical Convergence Zone as its main climatic system, has an average annual rainfall greater than 1500 mm, while the Northeastern Sertão area has rainfall less than 500 mm, which corroborates our study and the results found by Santos et al. (2019b). In Northeast Brazil, Araújo and Souza (2012) identified four different zones regarding precipitation pattern, two of which are distributed over Paraíba State. In such a study, it was noticed that the coastal area has a distinct pattern compared to the interior of the state, which corroborates with our research. The caveat is that the coastal area consists of a less significant area than the other region that covers the interior and much of the state, which indicates that Paraíba tends to behave very homogeneously.

In Paraíba, Macedo et al. (2010) delimited three different zones: one covers Mata Paraibana and the eastern half of Agreste, another is located from the western half of Agreste to Borborema, and the third zone is situated in the Sertão Paraibano. Despite this unconventional division in relation to the separation of the Sertão Paraibano from Borborema, the researchers emphasize that these regions are the most similar to each other and that the behavior found on the coast has more unique characteristics, similar to the results found in this paper. Using the TRMM-estimated data, Santos et al. (2019a) showed an evident differentiation of precipitation in the regions of Agreste and Mata Paraibana from the behavior of the regions of Borborema and Sertão, and that this clustering pattern can be found even when evaluating different time scales.

The conclusions of these studies reaffirm the results found in this research. In general, there is evidence to zone Paraíba State into two regions, one comprising Sertão and Borborema, and the other comprising Agreste and Mata Paraibana. The TRMM-estimated data identified this behavior at multiple scales and for different categories of drought time series. The results may be related to several factors, such as the influence of altitude and the Borborema Plateau, among others. This formation blocks the effects of atmospheric systems and influences precipitation and droughts in the region. In addition, factors such as the proximity of the regions to the ocean or the performance of different climatic systems may have caused this grouping pattern.

This study has illustrated the value of TRMM data in the regional characterization of drought and the utility of using hierarchical cluster analysis to extract important understanding regarding the heterogeneity of drought state, duration, and severity. Although neither TRMM based assessment using gauged observations nor drought monitoring represent novel applications, we consider that the use of TRMM in the context of drought assessment based on multi-scale accumulation is of significant interest and this is even more relevant when we integrate this methodology with the use of hierarchical cluster analysis. Although a limited time series was used, the approach demonstrates the potential for application with other drought indices and remote sensing data products. The obtained results are encouraging and the proposed innovative methodology proved to be adequate for such an analysis, and may be perfectly applicable in other regions, which should present even better results for regions with long complete time series. Finally, future research could focus on the influence of the length of time series, the use of other drought indices and other types of satellite-estimated rainfall data.

5. Conclusions

This study evaluated the performance of the TRMM rainfall product for monitoring drought over Paraíba State using hierarchical cluster analysis to identify areas with homogeneous behaviors, duration and severity of droughts over eight time scales across Paraíba State (1998–2017). For short-term droughts, there is a rationale for dividing the state into two large regions: Mata Paraibana and Agreste Paraibano and another by the Sertão Paraibano and Borborema. For long-term droughts, there is a stronger argument to group the entire state into a single cluster.

The TRMM-estimated time series are more similar to each other and demonstrate that the Sertão and Borborema mesoregions have greater homogeneity between them, while the results obtained from gauge-measured rainfall data have greater variability and show that the Mata Paraibana and Agreste mesoregions are more similar. Factors such as proximity to the ocean, the behavior of macro-, meso- and micro-scale climatic systems, and the configuration of the local relief are potential influencers of the pattern of occurrences of droughts and rains in the region, especially the Planalto da Borborema as a determining agent. It is concluded that the TMPA precipitation estimates of the TRMM are a valuable source of data to regionalize and identify the drought pattern and Paraíba State.

Finally, further studies of this type should be carried out to categorize and monitor these phenomena more accurately from satellite data derived from subsequent missions such as the Global Precipitation Mission (GPM), and in particular, the IMERG precipitation datasets which offer near real time opportunities for analysis and monitoring.

CRediT authorship contribution statement

R.M.B.N. and C.A.G.S. designed the research; R.M.B.N. and C.A.G.S. worked on the visualization and wrote the original draft; R.M.B.N., C.A.G.S., R.M.d.S., C.A.C.d.S, Z.L. and N.W.Q. conducted the manuscript review and editing, and wrote the final paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES)–Finance Code 001 (Pró-Alertas Research Project–Grant No. 88887.091737/2014-01 and 88887.123949/2015-00), the National Council for Scientific and Technological Development, Brazil – CNPq (Grant Nos. 304213/2017-9, 304540/2017-0 and 304493/2019-8) and Federal University of Paraíba (Public call No. 01/2021 Produtividade em Pesquisa PROPESQ/PRPG/UFPB proposal code: PVF14853-2021).

References

Al-Falahi, A.H., Saddique, N., Spank, U., Gebrechorkos, S.H., Bernhofer, C., 2020. Evaluation the performance of several gridded precipitation products over the highland region of Yemen for water resources management. Remote Sens. 12 (18), 2984. https://doi.org/10.3390/rs12182984.

Amini, A., Abdeh Kolahchi, A., Al-Ansari, N., Karami Moghadam, M., Mohammad, T., 2019. Application of TRMM precipitation data to evaluate drought and its effects on water resources instability. Appl. Sci. 9 (24), 5377. https://doi.org/10.3390/app92453770-1714-y.

Araújo, W.S., Souza, F.A.S., 2012. Identificação de regiões pluviometricamente homogêneas no nordeste do brasil usando análise multivariada. Rev. Bras. Clim. 10 (1), 136–152. https://doi.org/10.5380/abclima.v10i1.30600.

Azhdari, Z., Bazrafshan, O., Shekari, M., Zamani, H., 2020. Three-dimensional risk analysis of hydro-meteorological drought using multivariate nonlinear index. Theor. Appl. Climatol. 142, 1311–1327. https://doi.org/10.1007/s00704-020-03365-3.

Bazrafshan, O., Zamani, H., Shekari, M., Singh, V.P., 2020. Regional risk analysis and derivation of copula-based drought for severity-duration curve in arid and semi-arid regions. Theor. Appl. Climatol. 141, 889–905. https://doi.org/10.1007/s00704-020-03217-0.

Brasil Neto, R.M., Santos, C.A.G., Nascimento, T.V.M., Silva, R.M., Santos, C.A.C., 2020. Evaluation of the TRMM product for monitoring drought over Paraíba state, northeastern Brazil: a statistical analysis. Remote Sens. 12, 2184. https://doi.org/10.3390/rs12142184.

Brasil Neto, R.M., Santos, C.A.G., Silva, J.F.C.B.C., Silva, R.M., Santos, C.A.C., Mishra, M., 2021. Evaluation of the TRMM product for monitoring drought over Paraíba state, northeastern Brazil: a trend analysis. Sci. Rep. 11, 1097. https://doi.org/10.1038/s41598-020-80026-5.

Brito, T.T., Oliveira-Júnior, J.F., Lyra, G.B., Gois, G., Zeri, M., 2017. Multivariate analysis applied to monthly rainfall over Rio de Janeiro state, Brazil. Meteorog. Atmos. Phys. 129 (5), 469–478. https://doi.org/10.1007/s00703-016-0481-x.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. Theory Methods 3 (1), 1–27. https://doi.org/10.1080/03610927408827101.

Chen, S., Zhang, L., Zhang, Y., Guo, M., Liu, X., 2020. Evaluation of tropical rainfall measuring Mission (TRMM) satellite precipitation products for drought monitoring over the middle and lower reaches of the Yangtze River basin, China. J. Geogr. Sci. 30, 53–67. https://doi.org/10.1007/s11442-02.

Dantas, J.C., Silva, R.M., Santos, C.A.G., 2020. Drought impacts, social organization and public policies in northeastern Brazil: a case study of the upper Paraíba River basin. Environ. Monit. Assess. 192, 765–785. https://doi.org/10.1007/s10661-020-8219-0.

De Jesús, A., Breña-Naranjo, J., Pedrozo-Acuña, A., Alcocer Yamanaka, V., 2016. The use of TRMM 3B42 product for drought monitoring in Mexico. Water (8), 325 https://doi.org/10.3390/w8080325.

de Medeiros, I.C., Silva, J.F.C.B.C., Silva, R.M., Santos, C.A.G., 2019. Run-off-erosion modelling and water balance in the Epitácio pessoa dam river basin, Paraíba state in Brazil. Int. J. Environ. Sci. Technol. 16, 3035–3048. https://doi.org/10.1007/s13762-018-1940-3.

Ferreira da Silva, G.J., de Oliveira, N.M., Santos, C.A.G., da Silva, R.M., 2020. Spatiotemporal variability of vegetation due to drought dynamics (2012–2017): a case study of the upper Paraíba River basin, Brazil. Nat. Hazards 102, 939–964. https://doi.org/10.1007/s11069-020-03940-x.

Hadria, R., Boudhar, A., Ouatiki, H., Lebrini, Y., Elmansouri, L., Gadouali, F., Lionboui, H., Benabdelouahab, T., 2019. Combining use of TRMM and ground observations of annual precipitations for meteorological drought trends monitoring in Morocco. Am. J. Remote Sens. 7 (2), 25–34. https://doi.org/10.11648/j.ajrs.20190702.111016/j.jhydrol.2020.124916.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Stocker, E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8 (1), 38–55. https://doi.org/10.1175/JHM560.1.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., 2010. The TRMM multi-satellite precipitation analysis (TMPA). Chapter 1 in Satellite Rainfall Applications for Surface Hydrology. https://doi.org/10.1007/978-90-481-2915-7.

IBGE, Instituto Brasileiro de Geografia e Estatística, 2016. Divisão regional do Brasil em mesorregiões e microrregiões geográficas territorial brasileira. Disponível em: thttp://biblioteca.ibge.gov.br/visualizacao/livros/liv2269_1.pdf. Rio de Janeiro. Acesso em: nov. de 2019.

IPCC, 2014. Central and South America. In: Barros, V.R. (Ed.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1499–1566.

Keller Filho, T., Assad, E.D., Lima, P.R.S.R., 2005. Regiões pluviometricamente homogêneas no Brasil. Pesq. Agrop. Brasileira 40 (4), 311–322. https://doi.org/10.1590/S0100-204X2005000400001.

Li, X., Zhou, W., Chen, Y.D., 2015. Assessment of regional drought trend and risk over China: a drought climate division perspective. J. Clim. 28 (18), 7025–7037. https://doi.org/10.1175/jcli-d-14-00403.1.

Li, L., Song, X., Xia, L., Fu, N., Feng, D., Li, H., Li, Y., 2020. Modelling the effects of climate change on transpiration and evaporation in natural and constructed grasslands in the semi-arid loess plateau, China. Agric. Ecosyst. Environ. 302, 107077. https://doi.org/10.1016/j.agee.2020.107077.

Lilliefors, H.W., 1967. On the kolmogorov-smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 62 (318), 399–402.

Liu, C.L., Zhang, Q., Singh, V.P., Cui, Y., 2011. Copula-based evaluations of drought variations in Guangdong, South China. Nat. Hazards 59, 1533–1546. https://doi.org/10.1007/s11069-011-9850-4.

Liu, Z., Ostrenga, D., Teng, W., Kempler, S., 2012. Tropical rainfall measuring Mission (TRMM) precipitation data and Services for Research and Applications. Bull. Am. Meteorol. Soc. 93 (9), 1317–1325. https://doi.org/10.1175/BAMS-D-11-00152.1.

Lyra, G.B., Oliveira-Júnior, J.F., Zeri, M., 2014. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, northeast of Brazil. Int. J. Climatol. 34 (13), 3546–3558. https://doi.org/10.1002/joc.3926.

Macedo, M.J.H., Guedes, R.V.S., Souza, F.A.S., Dantas, F.R.C., 2010. Analysis of the standardized precipitation index for the Paraíba state, Brazil. Ambiente Água 5 (1), 204–214. https://doi.org/10.4136/ambi-agua.130.

Marengo, J.A., Torres, R.R., Alves, L.M., 2017. Drought in Northeast Brazil–past, present, and future. Theor. Appl. Climatol. 129 (3–4), 1189–1200. https://doi.org/10.1007/s00704-016-1840-8.

McGree, S., Schreider, S., Kuleshov, Y., 2016. Trends and variability in droughts in the pacific islands and Northeast Australia. J. Clim. 29 (23), 8377–8397. https://doi.org/10.1175/JCLI-D-16-0332.1.

McKee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. Proceedings of the Eighth Conference on Applied Climatology. American Meteorological Society, pp. 179–184.

Mossad, A., Alazba, A.A., 2018. Determination and prediction of standardized precipitation index (SPI) using TRMM data in arid ecosystems. Arab. J. Geosci. 11, 132. https://doi.org/10.1007/s12517-018-3487-5.

Naumann, G., Barbosa, P., Carrao, H., Singleton, A., Vogt, J., 2012. Monitoring drought conditions and their uncertainties in Africa using TRMM data. J. Appl. Meteorol. Climatol. 51 (10), 1867–1874. https://doi.org/10.1175/JAMC-D-12-0113.1.

Oliveira-Júnior, J.F., Xavier, F.M.G., Teodoro, P.E., Gois, G., Delgado, R.C., 2017. Cluster analysis identified rainfall homogeneous regions in Tocantins State, Brazil. Biosci. J. 33 (2), 333–340. https://doi.org/10.14393/BJ-v33n2-32739.

Prakash, S., Mitra, A.K., Momin, I.M., Rajagopal, E.N., Basu, S., Collins, M., Turner, A.G., Achuta, Rao K., Ashok, K., 2015. Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. Int. J. Climatol. 35 (9), 2326–2338. https://doi.org/10.1002/joc.4129.

Qin, Y., Chen, Z., Shen, Y., Zhang, S., Shi, R., 2014. Evaluation of satellite rainfall estimates over the chinese mainland. Remote Sens. 6 (11), 11649–11672. https://doi.org/10.3390/rs61111649.

Rad, A.M., Khalili, D., 2015. Appropriateness of clustered raingauge stations for spatio-temporal meteorological drought applications. Water Resour. Manag. 29 (11), 4157–4171. https://doi.org/10.1007/s11269-015-1051-6.

RajKhatiwada, K., Pandey, V.P., 2019. Characterization of hydro-meteorological drought in Nepal Himalaya: a case of Karnali River basin. Weather Clim. Extremes 26, 100239. https://doi.org/10.1016/j.wace.2019.100239.

Reboita, M.S., Gan, M.A., Rocha, R.P., Ambrizzi, T., 2010. Regimes de precipitação na América do sul: uma revisão bibliográfica. Rev. Bras. Meteorol. 25 (2), 185–204. https://doi.org/10.1590/S0102-77862010000200004.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Santos, C.A.G., Brasil Neto, R.M., Passos, J.S.A., Silva, R.M., 2017. Drought assessment using a TRMM-derived standardized precipitation index for the upper São Francisco River basin, Brazil. Environ. Monit. Assess. 189 (6), 250–270. https://doi.org/10.1007/s10661-017-5948-9.

Santos, C.A.G., Brasil Neto, R.M., Silva, R.M., Costa, S.G.F., 2019a. Cluster analysis applied to spatiotemporal variability of monthly precipitation over Paraíba state using tropical rainfall measuring Mission (TRMM) data. Remote Sens. 11 (6), 637. https://doi.org/10.3390/rs11060637.

Santos, C.A.G., Brasil Neto, R.M., Silva, R.M., Santos, D.C., 2019b. Innovative approach for geospatial drought severity classification: a case study of Paraíba state, Brazil. Stoch. Env. Res. Risk A. 33 (2), 545–562. https://doi.org/10.1007/s00477-018-1619-9.

Sharma, S., Khadka, N., Hamal, K., Shrestha, D., Talchabhadel, R., Chen, Y., 2020. How accurately can satellite products (TMPA and IMERG) detect precipitation patterns, extremities and drought across the nepalese Himalaya? Earth and space. Science 7 (8). https://doi.org/10.1029/2020EA001315.

Shiau, J.-T., Lin, J.-W., 2016. Clustering quantile regression-based drought trends in Taiwan. Water Resour. Manag. 30 (3), 1053–1069. https://doi.org/10.1007/s11269-015-1210-9.

Silva, R.M., Beltrão, J., Dantas, J.C., Santos, C.A.G., 2018. Hydrological simulation in a tropical humid basin in the cerrado biome using the SWAT model. Hydrol. Res. 49, 908–923. https://doi.org/10.2166/nh.2018.222.

Soares, A.S.D., Da Paz, A.R., Piccilli, D.G.A., 2016. Avaliação das estimativas de chuva do satélite TRMM no estado da Paraíba. Rev. Bras. Recur. Hidr. 21 (2), 288–299. https://doi.org/10.21168/rbrh.v21n2.p288-299.
Suliman, A.H.A., Awchi, T.A., Al-Mola, M., Shahid, S., 2020. Evaluation of remotely sensed precipitation sources for drought assessment in semi-arid Iraq. Atmos. Res. 242, 105007. https://doi.org/10.1016/j.atmosres.2020.105007.
Tan, M.L., 2019. Assessment of TRMM product for precipitation extreme measurement over the Muda River Basin Malaysia. Hydrol. Res. 2 (1), 69–75. https://doi.org/10.1016/j.hydres.2019.11.004.
Tan, M., Duan, Z., 2017. Assessment of GPM and TRMM precipitation products over Singapore. Remote Sens. 9 (7), 720. https://doi.org/10.3390/rs9070720.
Tan, M.L., Tan, K.C., Chua, V.P., Chan, N.W., 2017. Evaluation of TRMM product for monitoring drought in the Kelantan River Basin, Malaysia. Water 9 (1), 57–71. https://doi.org/10.3390/w9010057.
Tao, H., Fischer, T., Zeng, Y., Fraedrich, K., 2017. Evaluation of TRMM 3B43 precipitation data for drought monitoring in Jiangsu Province, China. Water 8 (6), 221. https://doi.org/10.3390/w8060221.
Teodoro, P.E., Oliveira-Júnior, J.F., Cunha, E.R., Correa, C.C.G., Torres, F.E., Bacani, V.M., Gois, G., Ribeiro, L.P., 2016. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Mato Grosso do Sul State, Brazil. Meteorog. Atmos. Phys. 128 (2), 197–209. https://doi.org/10.1007/s00703-015-0408-y.
Unal, Y., Kindap, T., Karaca, M., 2003. Redefining the climate zones of Turkey using cluster analysis. Int. J. Climatol. 23 (9), 1045–1055. https://doi.org/10.1002/joc.910.
Ur Rahman, K., Shang, S., Shahid, M., Wen, Y., 2020. Hydrological evaluation of merged satellite precipitation datasets for streamflow simulation using SWAT: a case study of Potohar Plateau, Pakistan. J. Hydrol. 587, 125040. https://doi.org/10.1016/j.jhydrol.2020.125040.
Vieira, R.M.S.P., Sestini, M.F., Tomasella, J., Marchezini, V., Pereira, G.R., Barbosa, A.A., Santos, F.C., Rodriguez, D.A., Nascimento, F.R., Santana, M.O., Campello, F.C.B., Ometto, J.P.H.B., 2020. Characterizing spatio-temporal patterns of social vulnerability to droughts, degradation and desertification in the Brazilian northeast. Environ. Sustain. Indic. 5, 100016. https://doi.org/10.1016/j.indic.2019.100016.
Wang, X., Shen, H., Zhang, W., Cao, J., Qi, Y., Chen, G., Li, X., 2015. Spatial and temporal characteristics of droughts in the Northeast China transect. Nat. Hazards 76 (1), 601–614. https://doi.org/10.1007/s11069-014-1507-7.
Xia, L., Zhao, F., Mao, K., Yuan, Z., Zuo, Z., Xu, T., 2018. SPI-based analyses of drought changes over the past 60 years in China's major crop-growing areas. Remote Sens. 10 (2), 171–185. https://doi.org/10.3390/rs10020171.
Yang, P., Xia, J., Zhang, Y., Han, J., Wu, X., 2017. Quantile regression and clustering analysis of standardized precipitation index in the Tarim River Basin, Xinjiang, China. Theor. Appl. Climatol. 134 (3–4), 1–12. https://doi.org/10.1007/s00704-017-2313-4.
Yevjevich, V., 1967. An Objective Approach to Definitions and Investigation of Continental Hydrologic Droughts. Hydrology Paper 23. Colorado State U, Fort Collins.
Zamani, H., Bazrafshan, O., 2020. Modeling monthly rainfall data using zero-adjusted models in the semi-arid, arid and extra-arid regions. Meteorog. Atmos. Phys. 132, 239–253. https://doi.org/10.1007/s00703-019-00685-6.
Zeng, H., Li, L., Li, J., 2012. The evaluation of TRMM multisatellite precipitation analysis (TMPA) in drought monitoring in the Lancang River basin. J. Geogr. Sci. 22, 273–282. https://doi.org/10.1007/s11442-012-0926-1.
Zhang, Q., Kong, D., Singh, V.P., Shi, P., 2017. Response of vegetation to different timescales drought across China: spatiotemporal patterns, causes and implications. Glob. Planet. Chang. 152, 1–11. https://doi.org/10.1016/j.gloplacha.2017.02.008.
Zhao, Q., Chen, Q., Jiao, M., Wu, P., Gao, X., Ma, M., Hong, Y., 2018. The temporal-spatial characteristics of drought in the loess plateau using the remote-sensed TRMM precipitation data from 1998 to 2014. Remote Sens. 10 (6), 838. https://doi.org/10.3390/rs1006083.