ORIGINAL PAPER

A detailed framework for the characterization of rainfall climatology in semiarid watersheds

Pedro R. Mutti[1,2] [ID] · Lizandro P. de Abreu[1,3] · Lara de M. B. Andrade[1,4] · Maria Helena C. Spyrides[1] · Kellen C. Lima[1,5] · Cristiano P. de Oliveira[1,4] · Vincent Dubreuil[2] · Bergson G. Bezerra[1,4]

Abstract

The identification of spatial and temporal rainfall climatology patterns is crucial for hydrometeorological studies over semiarid watersheds, which frequently face water distribution conflicts and socioeconomic issues due to water scarcity. Thus, the objective of this study was to propose a comprehensive approach for the characterization of rainfall climatology over semiarid watersheds. Monthly rainfall time series (1962–2015) with up to 30% of gaps measured in 56 rain gauges in the Piranhas-Açu Watershed—Brazilian semiarid region—were used. Data gaps were filled through a combination of simple spatial interpolation techniques. Principal component analysis and cluster analysis identified two homogeneous rainfall subregions in the basin: C1, in the upper portion, and C2, in the middle and lower portions. Rainfall volumes in C2 were up to 23.5% smaller than those in C1, due to orographic structures which contribute to aridity in this region. Rainfall anomalies were calculated in each cluster through the modified Rainfall Anomaly Index (mRAI) and were associated with the phases of the El Niño–Southern Oscillation (ENSO) and the Atlantic Meridional Mode (AMM). In years when the ENSO (AMM) was in its positive (negative) phase, there was a higher probability of occurrence of months with above-average rainfall, while the opposite was also true. Results showed that the effects of the patterns are mutually influenced, which has been previously found at larger scales. Finally, mRAI trends were identified through the Mann-Kendall test, which indicated significant negative trends in C1 and C2, especially during the wet season.

1 Introduction

Rainfall climatology and its spatial and temporal patterns are of the uttermost importance to hydroclimatological studies at the global, regional, and even local scales (Frazier et al. 2016). Understanding the climatological behavior of rainfall facilitates the identification of changes and deviations from patterns that are normally observed, which in turn usually impact water provisioning, agriculture, ecosystem services, and the general population welfare (Gocic and Trajkovic 2014; Lyra et al. 2014; Rodell et al. 2018). The characterization of rainfall patterns is of an even greater importance when oriented towards regions that suffer from water scarcity, where water conflicts are recurrent and water distribution is a limiting factor to many activities (da Silva et al. 2015; Mutti et al. 2019). Semiarid regions, which occupy approximately 15% of the globe's surface, are usually impacted by this type of conflict because they are extremely vulnerable to human-ecosystem interactions and global climate change (Huang et al. 2016).

The identification of such spatial and temporal rainfall patterns is usually carried out through the use of renowned statistical methods with a good applicability at different scales, such as cluster analysis (Gong and Richman 1995; Lyra et al. 2014; de Oliveira et al. 2017; Rau et al. 2017; Tinôco et al. 2018), principal component analysis (PCA) (Eklundh and Pilesjö 1990; Gocic and Trajkovic 2014; Almazroui et al. 2015; Fazel et al. 2018), and trend analysis through the Mann-Kendall non-parametric test (Mondal et al. 2012; da Silva et al. 2015, 2018; Lacerda et al. 2015; de Oliveira et al. 2017; Bezerra et al. 2018),

✉ Pedro R. Mutti
pedromutti@gmail.com

[1] Programa de Pós-graduação em Ciências Climáticas, Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte 59078-970, Brazil

[2] COSTEL LETG, UMR 6554 CNRS, Université Rennes 2, Rennes, Bretagne, France

[3] Instituto Federal do Piauí (IFPI), Corrente, Piauí, Brazil

[4] Departamento de Ciências Atmosféricas e Climáticas, Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil

[5] Escola de Ciências e Tecnologia, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brazil

among others. Furthermore, the association between rainfall and large-scale circulation mechanisms such as the El Niño–Southern Oscillation (ENSO) and the interhemispheric sea surface temperature gradient in the Tropical Atlantic (Atlantic Meridional Mode (AMM) also allows a better understanding of how rainfall variability is modulated by these mechanisms, aiding in rainfall seasonal forecasts (Rao and Hada 1990; Uvo et al. 1998; Hastenrath 2006; Grimm and Tedeschi 2009; Tedeschi et al. 2016; Timmermann et al. 2018). When combined, the results retrieved by these characterizations provide valuable information on the present and possible future rainfall patterns at different spatial scales. In the case of watersheds located at semiarid environments, this systematic and detailed approach may be considered an important water resources management tool.

Arid and semiarid drylands in Brazil, known as the Brazilian semiarid, are located mainly in the northeast region of the country, between 2.5° S and 16.1° S latitude and 34.8° W and 46.0° W longitude, with a total area of approximately 1,542,000 km², which encompasses 18.3% of the total Brazilian territory (Marengo and Bernasconi 2015). The Brazilian semiarid region is extremely vulnerable to drought and to interannual rainfall variability, which might be aggravated due to climate change projected scenarios which indicate a decrease in precipitated volumes and an increase in aridity until the end of the century (Marengo and Bernasconi 2015; Marengo et al. 2017b, 2018). Furthermore, according to the Intergovernmental Panel on Climate Change (IPCC 2014), the combination of rainfall variability, desertification, land degradation, and low socioeconomic status may increase the vulnerability of the region. Several studies have been carried out in order to identify and characterize spatial and temporal rainfall patterns in the northeast region of Brazil and in its semiarid region, such as drought characterization (Hastenrath 2012; Marengo and Bernasconi 2015; Costa et al. 2016; Marengo et al. 2017b, 2018; Brito et al. 2018), extreme rainfall identification (Correia Filho et al. 2016; da Silva et al. 2018), identification of teleconnection patterns (Rao and Hada 1990; Kane 1997; Uvo et al. 1998; Hastenrath 2006), trend analysis (Lacerda et al. 2015; de Oliveira et al. 2017; Dubreuil et al. 2018; da Silva et al. 2018), and spatial and temporal variability characterization (de Moscati and Gan 2007; Rao et al. 2016; de Oliveira et al. 2017; Tinôco et al. 2018). However, studies in this region which aim to comprehensively identify such characteristics at the watershed scale or at scales smaller than global or regional are still lacking, although some recent efforts can be highlighted, such as the works by da Silva et al. (2009), Lyra et al. (2014), de Andrade et al. (2016), Bezerra et al. (2018), Melo et al. (2018), da Silva et al. (2018), and Mutti et al. (2019).

In this context, one of the most important watersheds which are entirely located in the Brazilian semiarid region is the Piranhas-Açu Watershed (PAW). This water basin, which encompasses the Paraíba (PB) and Rio Grande do Norte (RN) states, is responsible for the domestic water supply of

approximately 1.3 million people and also for providing water to irrigation districts which are strategic for the socioeconomic development of the region (ANA 2014; Mutti et al. 2019). Furthermore, water conflicts have already been assessed in this watershed (de Amorim et al. 2016), which reinforces the relevance of studies on its hydrological processes. However, reliable measured meteorological data in this region are scarce, which hampers the development of comprehensive rainfall characterization studies that end up using interpolated or reanalyzed data (Chen et al. 2008; Wagner et al. 2012; Xavier et al. 2016). This increases computational effort and decreases the reliability of such studies. Still, recent works such as the ones carried out by de Felix (2015) and Mutti et al. (2019) proposed to characterize, although preliminarily, rainfall patterns in the PAW.

Therefore, the objective of this research is to propose a comprehensive approach for the characterization of rainfall climatology over the PAW, which may be replicated in other watersheds, particularly those located in semiarid regions. It also aims to show that even with a reliable but gapped database, one can obtain valuable information on rainfall spatial patterns, teleconnections, anomalies, and trends, with low computational effort and using renowned methods. Understanding rainfall characteristics at the basin scale might also retrieve additional information to complement the results of previous studies carried out at the Brazilian semiarid at regional scale. Furthermore, a detailed characterization of rainfall in watersheds can be used to delineate water resource management policies and to aid in rainfall seasonal forecast models, helping prevent and mitigate the impacts caused by interannual rainfall variability over semiarid regions.

2 Material and methods

2.1 Study area

Located entirely in the Brazilian semiarid region, the PAW comprises part of the RN and PB states (between 38° 75′ W and 36° 17′ W longitude and 5° 06′ S and 7° 83′ S latitude), and it is the most important water basin in both states (ANA 2014). The main water course has its source in the Paraíba state and flows through over 400 km of drylands until reaching its mouth in the city of Macau (RN), totaling approximately 43,682 km² of drainage area. Regarding its hydrology, there are two main water reservoirs worth mentioning: Coremas-Mãe d'Água in PB and Engenheiro Armando Ribeiro Gonçalves in RN, which are responsible for over 70% of all water storage capacity in the watershed (ANA 2014) (Fig. 1). These reservoirs guarantee that the main water course is perennial and therefore are crucial to the establishment of irrigation districts. The most important irrigated areas are located in the upper portions of the basin (near the
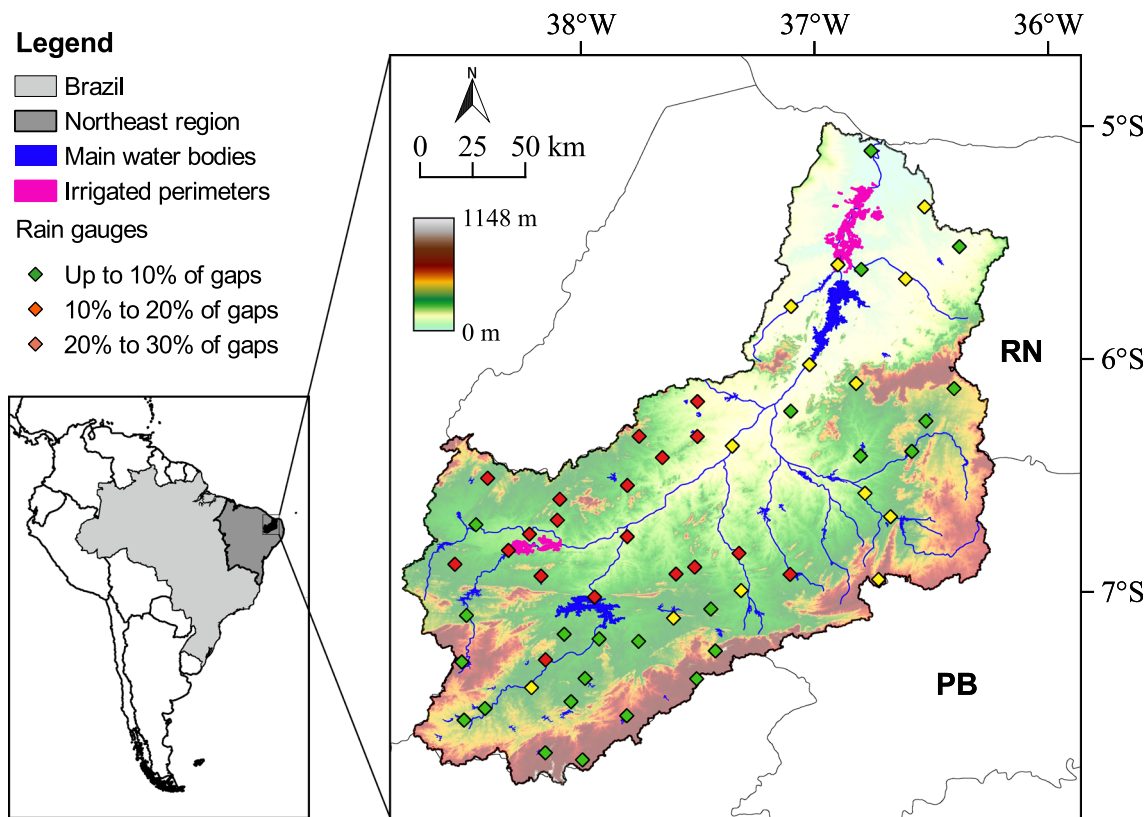
Fig. 1 Location of the Piranhas-Açu Watershed and rain gauges used in the study

Coremas-Mãe d'Água reservoir) and in the lower portions of the river (contiguous to the Armando Ribeiro reservoir), with mainly fruit crops which make these regions particularly relevant in the context of economic activities in the basin.

There are two main climate classifications that occur in the PAW (Alvares et al. 2014). In the upper portion, the climate is tropical with dry summer and annual rainfall reaching up to 1100 mm, while in the rest of the basin, the climate is predominantly arid, with annual rainfall as low as 400 mm. The rainfall regime in the PAW is determined mainly by the equatorial positioning of the Intertropical Convergence Zone (ITCZ) (Marengo et al. 2011; de Oliveira et al. 2017). The wet season occurs from February to May due to the ITCZ assuming its southernmost position (~4° S). It then shifts towards the northern hemisphere (~4 to 5° N) from August to October, which establishes the dry season over the PAW (Hastenrath 2012; de Oliveira et al. 2017). Furthermore, the remarkable interannual variability observed in the region is mainly associated with large-scale circulation mechanisms in the Pacific Ocean (ENSO) and the Atlantic Ocean (AMM), which are the main cause of the alternation between extremely dry years and heavy rainfall years (Marengo et al. 2011, 2017b).

It is important to highlight that a previous understanding of the mechanisms associated with the climate, and, consequently, with the rainfall regime in the studied watershed, is crucial for the adaptation of the proposed methodology to other areas.

For example, other large-scale mechanisms besides ENSO and AMM should be considered in other regions of the globe. In addition, the methodology should also be adapted to different climatological rainfall patterns, which differ from region to considered in other regions of the globe. In addition, the methodology should also be adapted to different climatological rainfall patterns, which differ from region to region, even in semiarid drylands. Particularly in the Brazilian semiarid, there is a high spatial variability of precipitation, and at least four different rainfall patterns (Tinôco et al., 2018), differing according to the proximity of the coast, land cover, topography, and latitudinal position, all of which influence the different atmospheric systems responsible for rain over these areas.

2.2 Data

Rainfall data were obtained from the National Water Resources Information System (SNIRH) through the Hidroweb online platform (http://www.snirh.gov.br/hidroweb/publico/apresentacao.jsf). The SNIRH provides access to a nationwide hydrometeorological database, with data measured in public and private stations and gauges monitored by different regional and federal agencies. In this study, we preselected only rain gauges monitored by the Executive Agency of Water Management of the PB state (AESA) and the Agricultural Research Company of the RN

state (EMPARN). Based on data availability from these preselected gauges, we restricted the studied period to 1962–2015, comprising a total of 54 years of monthly data. One of the steps of the proposed methodology is to evaluate gap filling techniques for monthly rainfall data; therefore, we selected gauges with up to 30% of data gaps in the studied period. From a total of 191 preselected gauges in the PAW, we kept 56 as follows: 24 with up to 10% of gaps, 13 with 10 to 20% of gaps, and 19 with 20 to 30% of gaps (Fig. 1). The remaining 135 stations either had more than 30% of gaps or did not cover the entire 1962–2015 period. Figure 1 shows that gauges with 20 to 30% are located mainly in the western portion of the basin, which indicates that results in this particular region might present larger uncertainties.

2.3 Data quality control

Data homogeneity and consistency was assessed through double-mass curve analysis (Searcy et al. 1960). This analysis is carried out by comparing accumulated rainfall values of a given rain gauge with accumulated rainfall values of a set of reference gauges. To this end, two reference gauges were selected in the PB (São João do Rio do Peixe and Piancó) and in the RN (Caicó and Pedro Avelino), which were considered to have the most reliable data, with less than 5% of monthly gaps in the studied period. Thus, gauges located in the PB were compared with the PB reference gauges, and gauges in the RN were compared with RN reference gauges. When the double-mass curves showed inconsistencies, such as abrupt slope changes or "steps," suspicious data were deleted and marked as missing data (gaps).

After this initial database organization, gaps were filled through spatial interpolation techniques. In this study, we opted to use three spatial interpolation methods based on their simplicity and low computational effort, in such a way that the method with the highest correlation coefficient (r) in each target gauge was selected to fill the gap in that particular gauge and month. The use of multiple spatial interpolation techniques in order to fill monthly rainfall databases has proved to be highly efficient in several previous studies (Eischeid et al. 2000; Wagner et al. 2012; Giambelluca et al. 2013; Frazier et al. 2016). The techniques used in this study were: multiple regression least absolute deviations (MLAD), single best estimator (SBE), and inverse distance weighting (IDW).

2.3.1 Multiple regression least absolute deviations

This criterion consists of a more robust variation of the least squares estimation. Since rainfall data usually do not fit a normal distribution, which has already been verified with measured data in Northeast Brazil (Alvares et al. 2014), the MLAD method has the advantage of not being highly influenced by long tail distributions (Barrodale and Roberts 1973). The method consists of estimating the missing data by

minimizing the sum of the absolute deviations between neighboring rain gauges and the estimated value. In its equation, the regression coefficient β is calculated as to minimize the sum (Eischeid et al. 2000):

$$\sum_i \left| \sum_j X_{ij}\beta_j - Y_i \right| \tag{1}$$

where X_{ij} are the i observations in j neighboring gauges and Y_i are the missing data estimations in the target gauge associated with each i observation. In this study, we chose neighboring gauges for each target gauge considering the geographical aspect as well as the correlation coefficient (r) between gauges. A maximum of four neighbor gauges were selected for each target gauge.

2.3.2 Single best estimator

The SBE method consists of filling missing values with a value observed in the same period in the nearest neighbor gauge. Similarly to what was carried out by Eischeid et al. (2000), we chose the neighbor gauge for the SBE method based on the highest r correlation.

2.3.3 Inverse distance method

The IDW technique consists of filling the gaps in the target gauge with the weighed mean of observed values in neighboring gauges (Chen and Liu 2012):

$$Y_i = \sum_i w_i X_i \tag{2}$$

where the weights (w_i) are attributed according to the distance between gauges:

$$w_i = \frac{l_i^{-\alpha}}{\sum_i l_i^{-\alpha}} \tag{3}$$

in which l_i is the distance between the target gauge and each of the neighboring gauges and α is the power parameter equal to 2 as default. Since the IDW retrieves estimations exclusively where and when there is a data gap, the performance of the method was assessed through leave-one-out cross-validation. In this case, we ran the method k times, where k is the count of observations in the original database of each target gauge, and for each run, we omitted the value of one of the k observations (Lee et al. 1998). In this way, the method can estimate a value for each month, allowing the calculation of the r coefficient for each monthly data.

2.4 Identifying spatial and temporal rainfall patterns

As previously reported, defining groups with homogeneous rainfall spatial patterns at the watershed scale is an important tool for

water resource management. This type of regionalization also allows a more comprehensive assessment of teleconnections and the effects of climate change on water availability in water basins. An important characteristic to be taken into consideration in the specific case of arid and semiarid watersheds is rainfall seasonality. In the northern portion of the Brazilian semiarid, for example, dry months usually register the complete absence of rainfall, with a mean value of approximately 10 mm in these months in the entire region (de Oliveira et al. 2017). In the wet season, however, monthly rainfall spatial variability is much more noticeable. Because of this, in order to identify temporal and spatial rainfall patterns in the PAW, we used an approach which combined PCA and cluster analysis.

2.4.1 Principal component analysis

This method is mainly used in order to remove correlation between variables or reduce the size of the database by creating a new dataset composed of linearly independent components and ordered according to the amount of variance captured by them (Daultrey 1976; Stone and Auliciems 1992). These components are obtained through the determination of the eigenvectors and eigenvalues of the correlation matrix between data, where normalized eigenvectors (called loadings) represent the correlation between the original data and the estimated components (Gocic and Trajkovic 2014). A complete description of the method can be found in Daultrey (1976). The PCA has been used in several climatological studies for the identification of temporal and spatial rainfall patterns and even for the delimitation of homogeneous rainfall groups (Eklundh and Pilesjö 1990; Rodriguez-Puebla et al. 1998; Singh 2006; Almazroui et al. 2015; Fazel et al. 2018). However, in the present study, PCA was not used as a clustering method but to prepare data to be used in cluster analysis. The main objective of using the PCA was to remove correlation between data, which is particularly high in dry season months, restricting the analysis to the components represented by the months which explain most of the data variance. Once calculated, the principal components were rotated by the varimax orthogonal method in order to facilitate data analysis by maximizing high loadings values and minimizing low loadings values (Fazel et al. 2018).

2.4.2 Cluster analysis

The new database generated by the PCA was subjected to cluster analysis in order to define homogeneous rainfall regions in the PAW. Cluster analysis identifies data agglomerations in a way that each group is composed of similar data within each group but heterogeneous data between each other group. In other words, this method seeks to minimize the variance between data clusters. Similarly to the PCA, cluster analysis has also been frequently used in climatological studies in order to define homogeneous rainfall regions in

different spatial scales, including in Northeast Brazil (Gadgil and Iyengar 1980; Gong and Richman 1995; Lyra et al. 2014; de Oliveira et al. 2017; da Silva et al. 2018; Tinôco et al. 2018). A full description of the method can be found in Anderberg (1973). In the present study, we opted to use the Euclidean distance as dissimilarity method, in accordance with other climate data regionalization studies in Brazil (Teixeira and Satyamurty 2011; de Oliveira et al. 2017). The Euclidean distance can be calculated as follows:

$$d_{ij} = \sqrt{\sum_k \left(x_{ik} - \bar{x}_{jk} \right)^2} \tag{4}$$

where k is the number of clusters and x_{ik} and x_{jk} are the observed values in gauges i and j. Furthermore, also in accordance with Teixeira and Satyamurty (2011) and de Oliveira et al. (2017), we used Ward's hierarchical agglomerative method. In this method, each gauge starts off representing one group, and in each subsequent step, one or more elements (groups) are merged according to their similarity until only one group containing all gauges is formed. We then proceed to identify which gauges are part of each group in a given step according to the optimal number of clusters, which will be discussed in the following section. The method aims to minimize the error sum of squares.

2.4.3 Cluster validation

One of the main difficulties in cluster analysis is defining the optimal number of clusters to be formed (Kannan and Ghosh 2011). In this study, we used the NbClust package of the R software (Charrad et al. 2014) which indicates the ideal number of clusters based on a compilation of different indicators. We also considered the silhouette width value for the validation of formed groups as it provides a graphical representation of the results. A complete description for its calculation can be found in Kannan and Ghosh (2011). The silhouette width ranges from -1 to 1, where positive values indicate a good object allocation and negative values indicate a poor object allocation. In this study, all objects with a silhouette width lower than the average were considered for reallocation because, according to Kaufman and Rousseeuw (1990), they have a weak structure.

2.5 Trends and anomalies

For each formed cluster, we created a synthetic monthly rainfall time series with the average of all gauges in each cluster and then we analyzed monthly and annual rainfall anomalies. Currently there are several indices in the scientific literature that can be used to estimate rainfall anomalies, and the Standardized Precipitation Index (SPI) is the standard index for the determination of drought according to the World

Meteorological Organization (WMO) (Hayes et al. 2011). However, in the present study, we opted for the modified version of the Rainfall Anomaly Index (mRAI), developed by Hänsel et al. (2016) and based on the original index by van Rooy (1965). The authors showed that the mRAI is highly correlated with the SPI when considering monthly rainfall anomalies while demanding less computational effort. Since this study advocates the use of simpler techniques, the mRAI suits our needs because it can be calculated at multiple time-scales and the only input data is rainfall.

The mRAI can be calculated as follows:

$$mRAI = 1.7 \left[\frac{\left(N - \overline{N}\right)}{\left(\overline{M} - \overline{N}\right)} \right], \text{ for positive anomalies} \quad (5)$$

$$mRAI = -1.7 \left[\frac{\left(N - \overline{N}\right)}{\left(\overline{m} - \overline{N}\right)} \right], \text{ for negative anomalies} \quad (6)$$

where N is the observed rainfall in the target month, $\overline{N}$ is the median of the complete time series in the target month, $\overline{M}$ is the mean of the 10% highest rainfall values in the target month, and $\overline{m}$ is the mean of the 10% lowest rainfall values in the target month. Equation 5 (or 6) should be used if N is higher (or lower) than the median of the target month. The mRAI identifies each month as being extremely dry ($mRAI \leq -2$), very dry ($-1.99 < mRAI < -1$), dry ($-0.99 < mRAI < -0.5$), normal ($-0.49 < mRAI < 0.49$), wet ($0.5 < mRAI < 0.99$), very wet ($1 < mRAI < 1.99$), and extremely wet ($mRAI \geq 2$), according to a classification adapted from the one initially proposed by Hänsel et al. (2016). Since we aim to evaluate the historical behavior of anomalies, the entire time series (54 years) was used as base period for the calculation of the mRAI.

A well-known issue of using precipitation-based anomaly indices in semiarid regions is the high number of zero rainfall values, especially during the dry season (Kumar et al. 2009; Stagge et al. 2015). In the case of the mRAI, calculating the $\overline{m}$ term for dry months results in an index which is extremely sensitive even to small rainfall volumes. To overcome this problem, we calculated the mRAI for each month considering the 6-month and 12-month aggregated timescales. For that end, we considered a wet 6-month period (January to June—JFMAMJ) and a dry 6-month period (July to December—JASOND). Therefore, all seasonal analyses were carried out considering the 6-month mRAI while annual analyses considered the 12-month mRAI.

Once anomalies were calculated for each month in the time series, we could determine their behavior in relation to the occurrence of phenomena such as El Niño, La Niña, or anomalies in the AMM. In the case of the ENSO, we considered the monthly classification of the phenomenon as El Niño or La

Niña according to the Oceanic Niño Index (ONI) as provided by the National Oceanic and Atmospheric Administration (NOAA) of the USA in the website <http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php>. At the beginning of each period defined as El Niño or La Niña, we considered a minimum of 12 subsequent months in the analysis of the mRAI. For the AMM, we used the Tropical Atlantic sea surface temperature anomaly indices available at <https://www.esrl.noaa.gov/psd/data/climateindices/list/>, in which positive anomalies indicate above average warming of the Northern Tropical Atlantic Ocean and negative anomalies indicate above average warming of the Southern Tropical Atlantic Ocean. Strong AMM anomalous events were defined whenever a period of at least five consecutive months with the 3-month moving average of the anomaly index above 2 or below −2 was observed. Similarly, we analyzed a minimum of 12 subsequent months after the beginning of a strong AMM event. The behavior of rainfall anomalies in each cluster was analyzed according to the frequency distribution of the mRAI in each selected year.

Table 1 shows the years selected as representative of strong positive and negative ENSO and AMM anomalous events. Frequency distribution of the mRAI for each pattern (positive and negative ENSO, positive and negative AMM) was carried out seasonally (6-month mRAI). In order to test if the distributions found were different in each cluster, we used the chi-squared test (Plackett 1983). Furthermore, the 12-month mRAI was used in order to identify the most anomalous years

Table 1 Positive and negative ENSO and AMM years selected for the anomaly analysis

Pattern	Selected years
ENSO + (La Niña)	1964–1965, 1970–1972, 1973–1976, 1983–1985, 1988–1989, 1995–1996, 1998–2001, 2005–2006, 2007–2009, 2010–2012
ENSO − (El Niño)	1963–1964, 1965–1966, 1968–1970, 1972–1973, 1976–1978, 1979–1980, 1982–1983, 1986–1988, 1991–1992, 1994–1995, 1997–1998, 2002–2003, 2004–2005, 2006–2007, 2009–2010, 2014–2015
AMM + (warmer North Atlantic Ocean)	1962–1963, 1969–1970, 1980–1981, 1995–1997, 2001–2002, 2003–2007, 2010–2013
AMM − (warmer South Atlantic Ocean)	1963–1965, 1971–1975, 1977, 1982–1987, 1991–1992, 1993–1994, 2002–2003, 2009, 2014–2015

(in magnitude) and verify which ENSO and AMM patterns were observed in those years, according to Table 1.

Finally, the non-parametric Mann-Kendall test (Kendall 1938; Mann 1945) was used in order to identify linear trends in the seasonal and annual values of the mRAI. This test was reported by Goossens and Berger (1986) as the most adequate to analyze trends in climate variables, and it has been used in several studies worldwide and also in Northeast Brazil (Mondal et al. 2012; da Silva et al. 2015; de Oliveira et al. 2017; Sa'adi et al. 2017; Wu and Qian 2017; Zilli et al. 2017; Bezerra et al. 2018). The method consists of comparing each value in the time series with the remaining subsequent values in order, considering how many times the remaining values are higher or lower than the value being currently analyzed. Thus, we have:

$$S = \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} sgn\left(N_i - N_j\right) \tag{7}$$

where i and j are the years (or seasons) and sgn is defined as follows:

$$sgn = \begin{cases} 1, \text{if } \left(N_i - N_j\right) > 0 \\ 0, \text{if } \left(N_i - N_j\right) = 0 \\ -, \text{if } \left(N_i - N_j\right) < 0 \end{cases} \tag{8}$$

Furthermore, it is known that for databases with $n \geq 8$, the S statistic can be fitted to a normal distribution with a mean of 0 and with variance equal to:

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{q} t_i(t_i-1)(2t_i+5)}{18} \tag{9}$$

where t_i is the number of equal values found until sample i, and finally the Z statistic of the Mann-Kendall test is calculated as:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}}, \text{ if } S > 0 \\ 0, \qquad\quad \text{ if } S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}}, \text{ if } S < 0 \end{cases} \tag{10}$$

3 Results and discussion

3.1 Database gap filling

The overall average performance of the three gap filling methods shows that SBE presented the worst results, especially during the months from August to December (Fig. 2). This result is probably associated with the fact that there are plenty of gauges with 20 to 30% of gaps densely located in the western portion of the basin (as previously seen in Fig. 1).
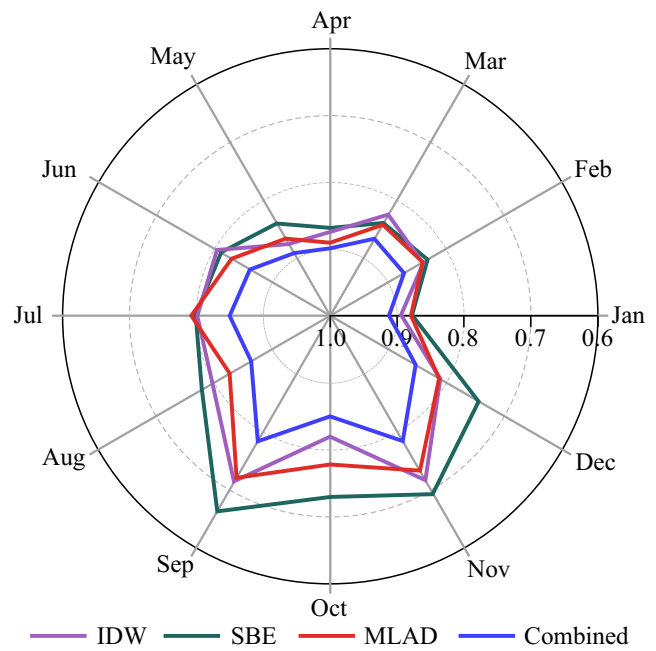

Fig. 2 Monthly distribution of the mean correlation coefficients (r) between observed and estimated rainfall values

Since the SBE uses data from the nearest gauges, during periods with lots of gaps, the overall performance of the method was considerably lower because it ended up using data from gauges geographically farther. Figure 2 also shows that all methods performed well during the months from January to July, with the r coefficient ranging from 0.80 to 0.89. Overall, the MLAD and IDW methods presented better results, and the performance during the driest months (September–December) was relatively worse, with r ranging from 0.72 to 0.86. As previously mentioned, semiarid regions usually register zero rainfall during dry season months. In this case, minor deviations in the interpolated values incur in considerable decreases in the r coefficient, and therefore, the overall performance in the dry season is worse than in the other months.

Figure 2 also shows that the overall performance of interpolation methods for the gap filling of monthly rainfall data greatly benefits from the combination of different techniques, which was also verified in previous studies (Eischeid et al. 2000; Giambelluca et al. 2013). For each month with missing data in each gauge, we chose the method with the highest r coefficient in that month and gauge, guaranteeing that the best performing method would always be chosen and thus increasing the overall skill of the gap filling. The combination of methods increased the gap filling performance from 3.5% in April to as much as 18.2% in September in terms of relative error.

It is important to note that this procedure was carried out after the initial data filtering which consisted of the visual analysis of the double-mass curves. The final gap filled database according to the method chosen can be observed in Fig. 3
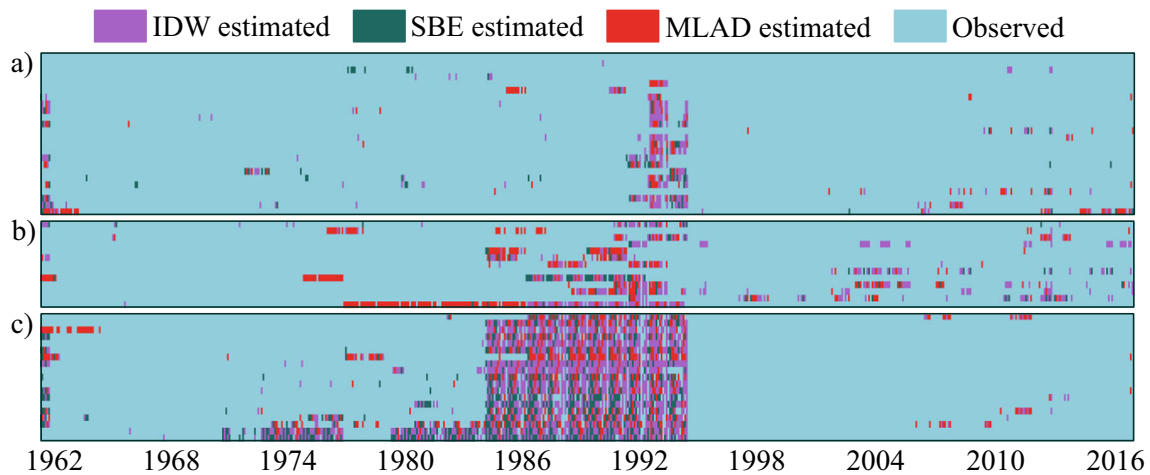
Fig. 3 Data gaps filled by the combination of the inverse distance weighting (IDW), single best estimator (SBE), and multiple regression least absolute deviations (MLAD) methods. (a) Rain gauges with up to 10% of gaps. (b) Rain gauges with 10 to 20% of gaps. (c) Rain gauges with 20 to 30% of gaps

and is further detailed in Table 2. The total number of missing values accounted for 10.2% of total data, and approximately half (49.8%) of these gaps were filled by the IDW method, followed by 28.3% of gaps filled by the MLAD method, and 21.9% filled by the SBE method.

Figure 3 also reveals a recurrent problem in studies using observational data in Brazil, particularly in the northeast region. There are long sequential periods with missing data, particularly during the 1980s and 1990s. Because of that, the analysis of long rainfall time series in the region is a true challenge and reliable gap filling procedures are essential so that larger amounts of consistent observed data can be used. As an alternative, gapped time series could also be merged with gridded reanalyzed data or remotely sensed data, such as what was carried out by Xavier et al. (2016). In the present study, we aimed to use simple spatial interpolation techniques instead of the merging of different databases. However, it is important to note that the fact that most gauges with 20 to 30% of gaps are located in the western portion of the basin combined with a large amount of sequential missing values during the 1980s and 1990s represents a relevant source of uncertainties regarding results found in this particular portion of the basin.

Table 2 Database quantitative description where *n* is the number of months

	Total	Missing data	Gap filling method		
			IDW	SBE	MLAD
n	36,288	3685	1835	807	1043
% of *n*	–	10.2[*]	49.8[**]	21.9[**]	28.3[**]

[*] In relation to the total count

[**] In relation to missing data

3.2 Homogeneous rainfall groups

The remarkable seasonal variability of data can be observed in Fig. 4, where during the wettest months (January to May), data variability is substantial because it is during this period that we can observe the most prominent differences between rainfall volumes throughout the basin. It can also be noted that even during the wet season there were registers of zero rainfall in some gauges, which also contributes to show how susceptible the PAW is to interannual and spatial rainfall variability. Furthermore, the boxplot in Fig. 4 indicates that the data used could benefit from cluster analysis in order to identify groups where such variability would be more homogeneous.

In the first step of the proposed clustering procedure, we used the PCA in our dataset with months organized as
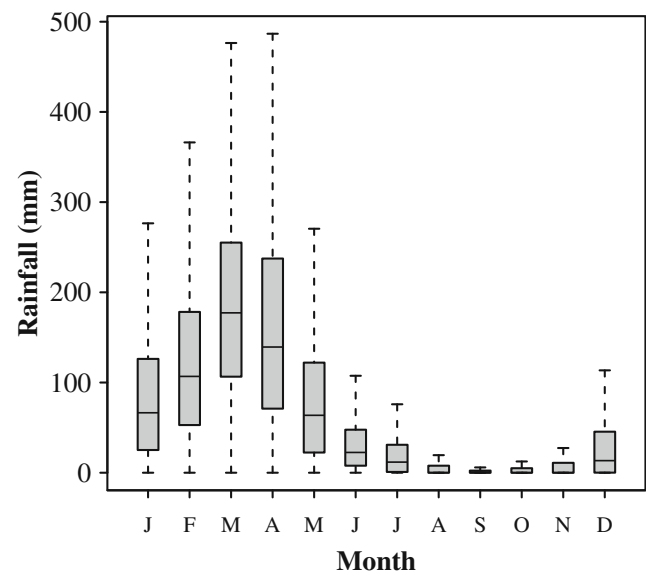

Fig. 4 Monthly boxplot of the gap filled database, comprising 3024 values per month (outliers were omitted)
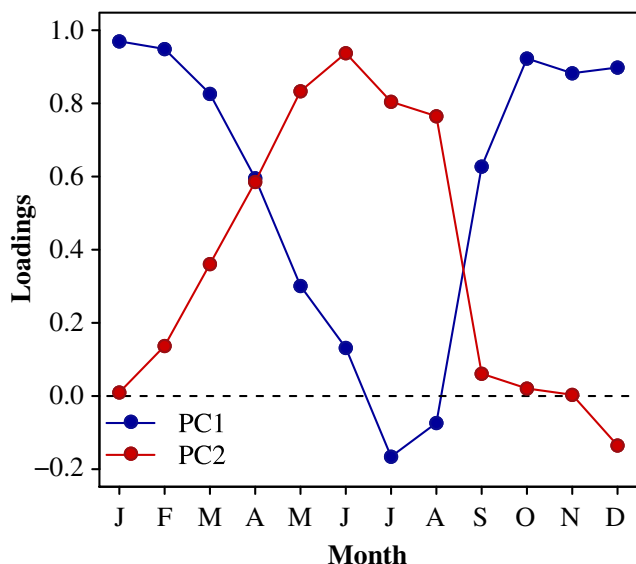
Fig. 5 Rotated principal component loadings

variables in order to remove correlation between months and focus on the components which account for the largest amount of data variability. The PCA indicated that the first component (PC1) explained 79.6% of total data variance while the second component (PC2) explained another 13.5%, totaling 93.1% with only the two first components. Rotated loadings showed that PC1, which explains most of the variation in data, is highly correlated with the months from October to March (Fig. 5). This result shows that rainfall during the transition from the dry season until the peak of the wet season accounts for most of data variability, while rainfall from May to August (transition from wet to dry season)—CP2—is more homogeneous through the watershed.

Thus, in our cluster analysis, we considered a database composed of the PC1 and PC2 scores retrieved from the PCA, which assured that the homogeneous groups would be formed considering temporal data variability, that is, months in which heterogeneity between gauges is more clearly noticeable. The results obtained from the use of the NbClust package retrieved 2 as the optimal number of clusters. This result was reaffirmed by the analysis of the silhouette width (Table 3) considering different numbers of clusters and using the Euclidean distance and Ward's method. The average silhouette width for this number of clusters was 0.61 which represents a reasonable cluster structure (Rau et al. 2017). All other numbers of clusters had a silhouette width lower than 0.60. Additionally, only two gauges presented negative silhouette width, which means no cluster structure and therefore should

be reallocated. Since $k = 2$ retrieved the best combination between average silhouette width and the number of negative silhouette width values, it was chosen as the optimal number of clusters, which is also coherent considering the two climate classifications observed in the watershed.

Figure 6 a shows the two clusters initially formed by the cluster analysis. Data was separated in a coherent spatial pattern, where rain gauges with weaker cluster structure (low silhouette width) are located in the interface between the two clusters. Rain gauges with below-average silhouette width and the two gauges with negative width were considered potential candidates for reallocation. In total, five rain gauges were reallocated from cluster 2 (C2) to cluster 1 (C1): the two negative silhouette width gauges, two gauges with weak structure (width < 0.30), and one gauge that was reallocated in order to maintain geographical coherence (Fig. 6a, b). The two final groups after reallocation are shown in Fig. 6b, in which C1 comprises the upper portion of the watershed and C2 comprises its middle and lower portions. The synthetic monthly rainfall time series for each cluster is shown in the lower right plot in Fig. 6b. The plot shows that, as expected, a difference between clusters cannot be explained by rainfall seasonality, since it is the same in the entire PAW, but by precipitated volume. Annual accumulated rainfall in the C1 equals to 889.8 mm while in C2 it equals to 681.1 mm, representing a difference of 23.5% (208.7 mm). It is also important to notice that because we used PCA prior to cluster analysis, differences between groups are more evident precisely in the months which better represented PC1 (October to April), which explains most of data variations.

There are two possible explanations for the differences observed between precipitated volumes in C1 and C2, although the entire PAW is mainly under the influence of the ITCZ, as previously explained (Mutti et al. 2019). The first is that C1 is in the upper portion of the basin and, as previously observed in Fig. 1, the region is surrounded by hills and mountains ranging from 500 to 1000 m in altitude. This landscape configuration strongly influences the occurrence of orographic rainfall and local convection, which has already proved to be one of the main factors associated with high rainfall rates in certain regions of Northeast Brazil (Lyra et al. 2014; de Andrade et al. 2016). On the other hand, the C2 region is located northwest of a particular mountain formation known as the Borborema Plateau. Since trade winds in this region blow mainly from the southeast (Hastenrath 2012), orographic rainfall occurs upwind of the Borborema hills (outside the PAW), and dry winds descend over the C2 region in the PAW, reducing rainfall

Table 3 Summary of the silhouette width analysis for 2 to 10 clusters formed by Euclidean distance and Ward's method. The best results are in italic

Number of clusters	2	3	4	5	6	7	8	9	10
Average silhouette width	*0.61*	0.51	0.51	0.53	0.54	0.58	0.59	0.56	0.58
Negative silhouette number	*2*	4	3	*2*	3	*2*	3	5	3

Fig. 6 (a) Silhouette width analysis for two clusters. (b) Final clusters' definition and their rainfall patterns
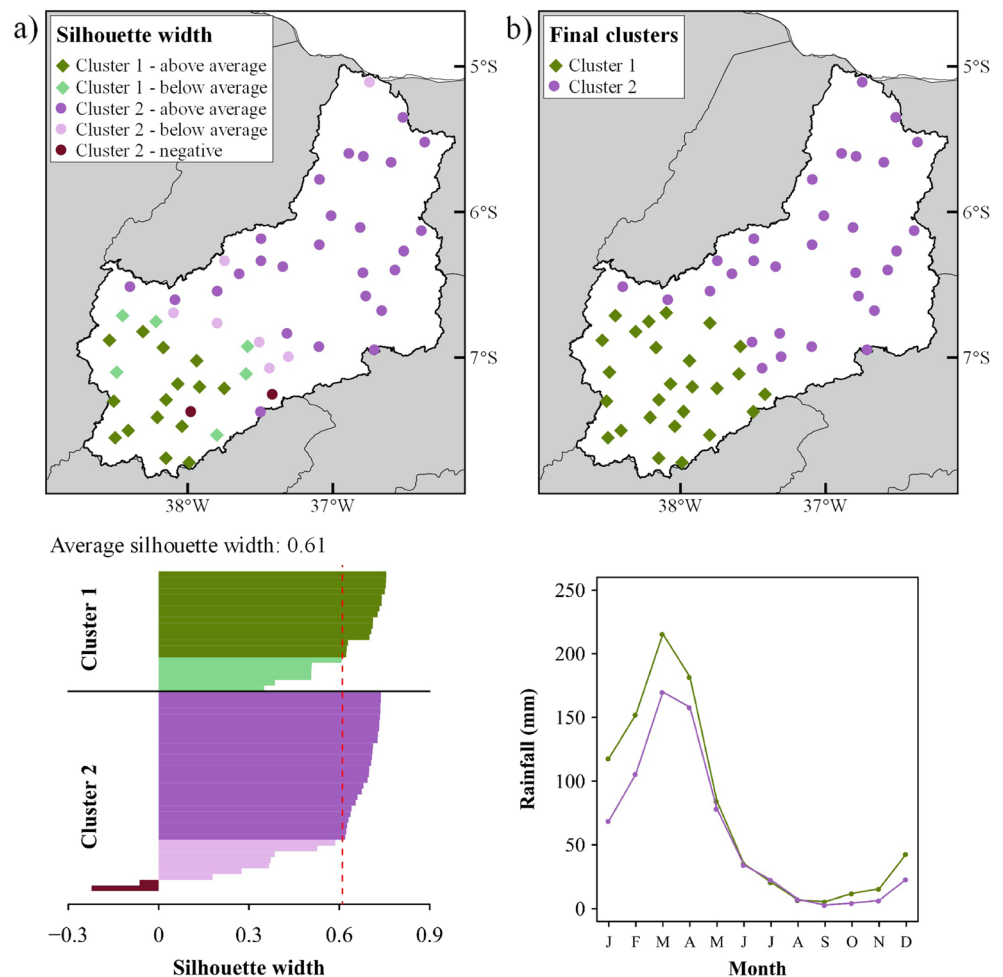
volumes in this region (Correia Filho et al. 2016; Reboita et al. 2016; Mutti et al. 2019).

Through the perspective of water resource management, the cluster configuration and characteristics favor the overall recharge of water storages in the basin. Upper portions receive higher water inputs which favor the recharge of underground and surface water courses in the lower portions. However, as the river flows near its mouth, water input from rainfall is much smaller and the consumption of water in irrigated crops makes the C2 region more vulnerable to extended drought periods, which are recurrent in the region (Marengo et al. 2017b).

3.3 Characterizing trends and anomalies

The time series of both 6-month and 12-month mRAI can be observed in Fig. 7. Remarkably dry and wet periods could be consistently represented in both timescales. The main drought episodes occurred in 1979–1984 (D1), 1990–1993 (D2), 1997–1999 (D3), and more recently in 2012–2015 (D4). D1 and D2 were characterized by their extended duration (5 and 4 years) and peak anomalies being registered by the end of the drought period. D3 was a shorter and less intense episode in

the PAW. D4, on the other hand, which is known to have lasted until 2016, peaked at the beginning of the episode (2012). Afterwards, it retreated in intensity and then rose again by 2015. All this drought episodes have been previously identified in the Brazilian semiarid region, and it has been reported that their impacts were catastrophic, with severe agriculture losses and increasing social conflicts due to water scarcity (Marengo et al. 2017b; Brito et al. 2018). Furthermore, D4 drought impacted all states in Northeast Brazil (Brito et al. 2018) and is known to have been the most severe ever registered when considering the entire Brazilian semiarid region, causing water reservoirs to collapse (Marengo et al. 2017a, 2018). The main difference between drought episodes in C1 and C2 refers to D1 beginning earlier in C2 (early 1979) when compared with C1 (early 1980).

Regarding anomalously wet episodes, from 1962 to 1978, a series of low-intensity, short-duration episodes were identified, with the 1972–1976 episode standing out. Afterwards, another episode was registered between 1985 and 1987 and then a long period without any remarkable positive anomaly was established until 2008–2009. Differences between C1 and C2 positive events were minor. Extremely wet events were

Fig. 7 6-month and 12-month mRAI time series in the C1 and C2 regions
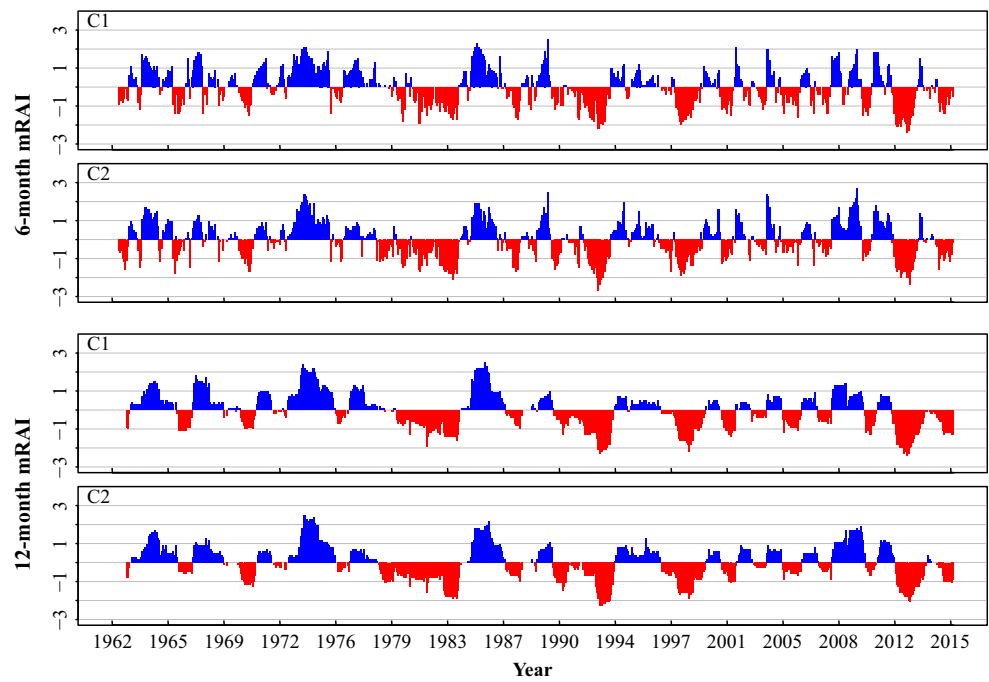
shorter (2 to 3 years) and happened mostly during the first half of the time series. Droughts, on the other hand, lasted longer and increased in frequency after 1979.

In relation to the 12-month mRAI, Table 4 shows the top five wettest and driest years in C1 and C2 and the respective ENSO and AMM phases, with the highest indices being highlighted in italic. One can observe that in years with the most negative mRAI value (5 cases), the ENSO was predominantly in its negative (2 cases) or neutral (2 cases) phase, while in years with the most positive mRAI (7 cases), the ENSO phase was mostly positive (4 cases). No clear pattern

Table 4 Condition of the ENSO and AMM phases (+ positive, − negative, or *n* neutral) during the top 5 positive and negative annual rainfall anomalies (12-month mRAI) in each cluster of the Piranhas-Açu Watershed. Top negative and positive anomalies are in italic

Year	Region	mRAI	ENSO	AMM
1964	C2	+ 1.3	−	−
1967	C1	+ 1.4	*n*	*n*
1974	C1/*C2*	+ 2.2/+ *2.4*	+	−
1977	C1	+ 1.1	−	−
1983	C1/C2	− 1.4/− 1.8	−	−
1985	*C1*/C2	+ *2.5*/+ 1.9	+	−
1990	C1/C2	− 1.3/− 1.5	*n*	*n*
1993	C1/*C2*	− 1.8/− *2.0*	*n*	−
1998	C1/C2	− 1.8/− 1.6	−	*n*
2008	C1/C2	+ 1.3/+ 1.2	+	*n*
2009	C2	+ 1.7	+	−
2012	*C1*/C2	− *2.1*/− 1.7	+	+

for the AMM during years with the most negative mRAI (5 cases) could be identified, although in 2012, its phase was positive despite the establishment of a La Niña in the Pacific. During years with the most positive mRAI (7 cases), the AMM phase was always either negative (5 cases) or neutral (2 cases).

Considering the 54 years in the time series, Figs. 8 and 9 show the frequency distribution of the 6-month mRAI during each ENSO and AMM phase in each cluster and divided by season (wet or dry, as previously defined). In La Niña years—positive ENSO phase (Fig. 8 left panels)—there were positive rainfall anomalies in 60.0% of the cases in C1 and 57.0% of the cases in C2 during the wet season. During these months, negative anomalies were less frequent, being 15.0% and 19.0% of the cases in C1 and C2, respectively. In the dry season of La Niña years, positive anomalies are more frequent in C2 (43%) than in C1 (32%), but negative anomalies are also frequent: 31% in C1 and 26% in C2.

In years when the ENSO phase was negative (El Niño), the mRAI pattern was more negative, but not remarkably more frequent than positive anomalies. Very dry and extremely dry months accounted for 31.0% of the cases in C1 and for 34.0% in C2, while wet or very wet months occurred in 27.0% and 19.0% of the cases in C1 and C2, respectively. In the dry season, however, the frequency of negative anomalies was of 44% (C1) and 42 (C2) while the frequency of positive anomalies was only of 8% (C1) and 15% (C2).

Regarding the AMM, Fig. 9 (left panels) shows the frequency distribution of the mRAI when the AMM phase is positive, that is, when there are warmer sea surface temperatures in the North Tropical Atlantic Ocean. This pattern favors
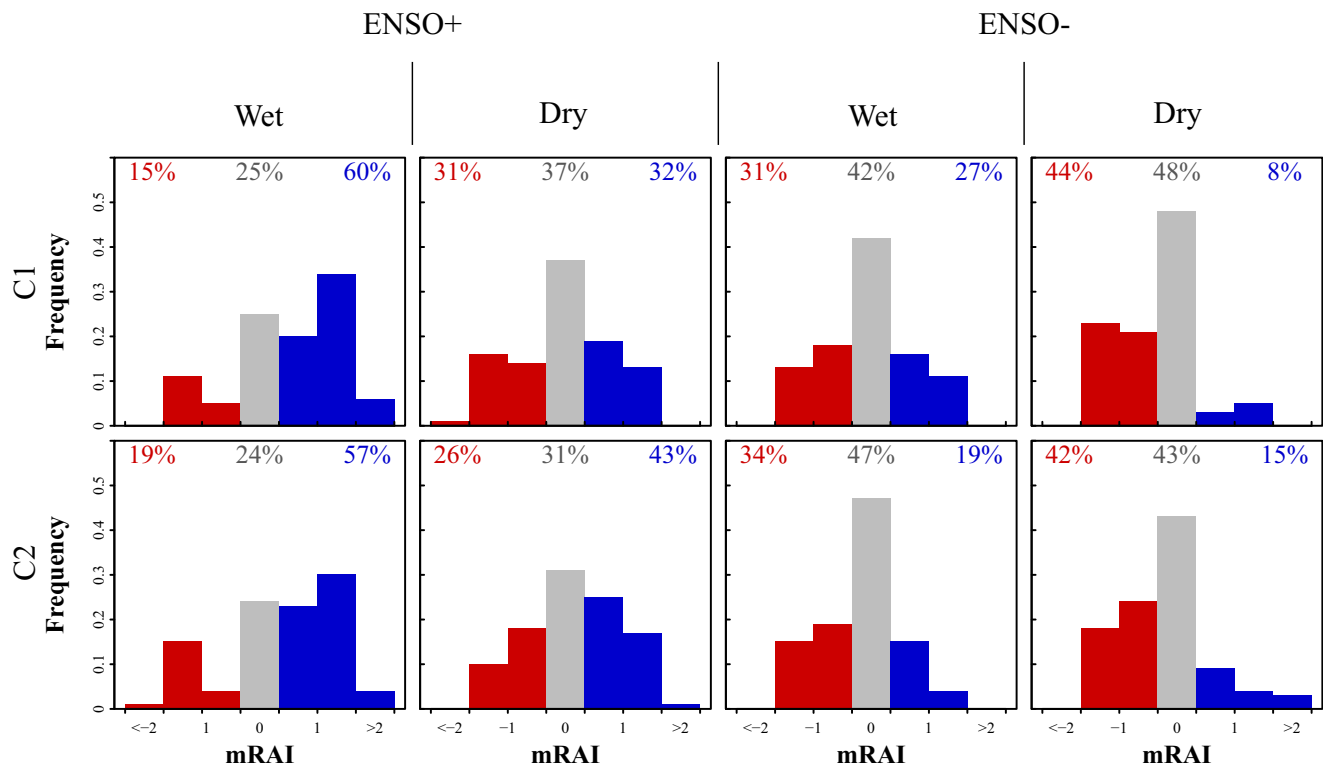
Fig. 8 Frequency distribution of the 6-month mRAI in each cluster and each season during selected ENSO events

the northern displacement of the ITCZ, which reduces rainfall in the PAW region. In these cases, the frequency of occurrence of months with negative mRAI is of 39.0% in C1 and 55.0%

in C2 during the wet season and of 51.0% in C1 and 44% in C2 during the dry season. It is interesting to highlight that extremely dry events occurred in the dry season of both
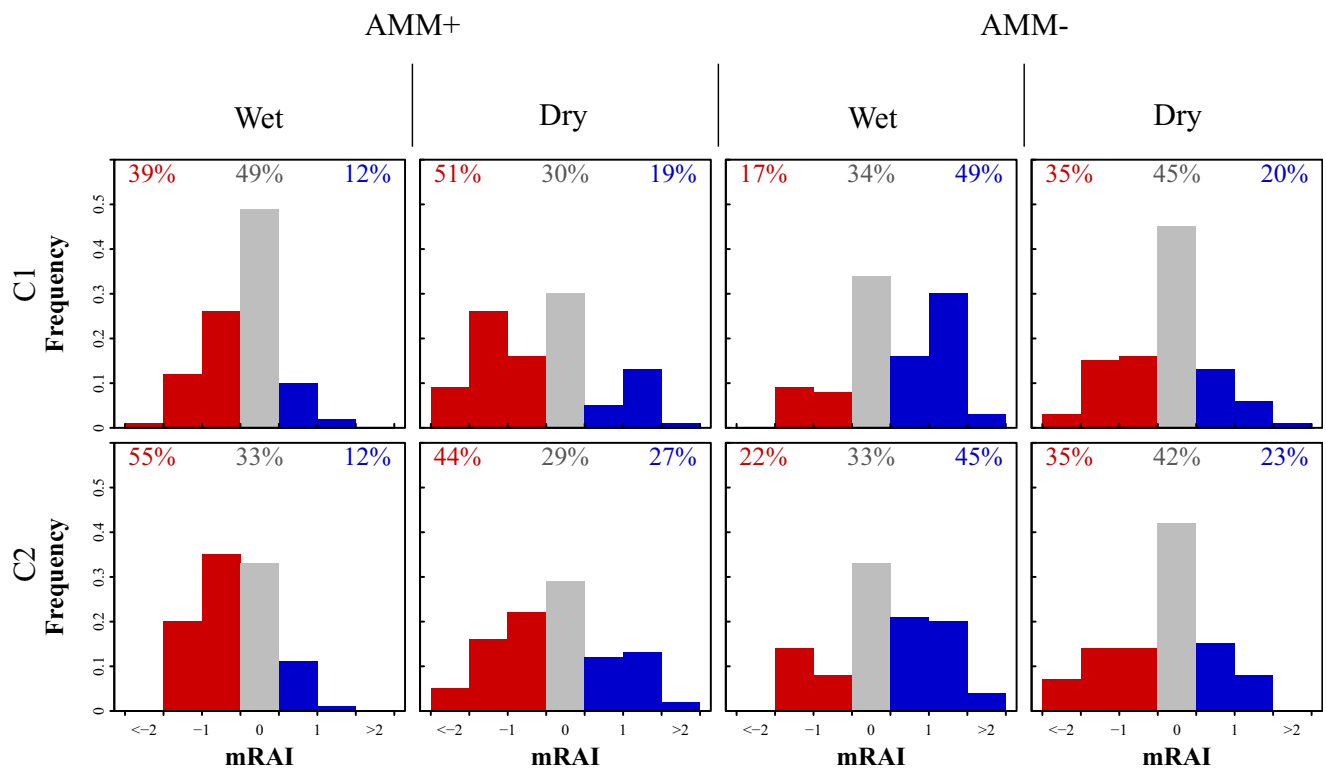

Fig. 9 Frequency distribution of the 6-month mRAI in each cluster and each season during selected AMM events

clusters, which could not be observed in any negative ENSO phase (Fig. 8 right panels). During this AMM phase, the occurrence of positive mRAI range from 12% (C1 wet season) to 27% (C2 dry season).

When the AMM is in its negative phase, sea surface temperatures in the South Tropical Atlantic Ocean are warmer, which favors the southern displacement of the ITCZ and the intensification of rainfall over Northeast Brazil. In these cases, the frequency of occurrence of months with positive mRAI in the wet season was of 49.0% in C1 and 45.0% in C2, and the frequency of months with negative mRAI was of 17.0% and 22.0% (Fig. 9 right panels). During the dry season, however, Fig. 9 does not indicate an evident predominance of wet or very wet months during negative AMM episodes. On the contrary, negative mRAI values were predominant. In Table 4, we noticed that most of the highest positive anomalies happened during negative AMM years coupled with La Niña. Figures 8 and 9 show that the wet season in particular is strongly influenced by these two phases of the ENSO and the AMM.

Results found in this analysis corroborate with several previous studies which assessed the effects of sea surface temperature in the Pacific and Atlantic Oceans on rainfall over Northeast Brazil (Rao and Hada 1990; Uvo et al. 1998; Marengo and Bernasconi 2015; Costa et al. 2016; Rao et al. 2016). These studies reported the occurrence of higher (lower) rainfall volumes in years when the ENSO phase was positive (negative) and the AMM phase was negative (positive). The results of the present study, particularly the ones regarding rainfall anomalies during negative AMM phases, also confirm the conclusions drawn by said authors, which identified that the effects of sea surface temperature anomalies in the Pacific and Atlantic Oceans interact, and therefore, it is extremely difficult to forecast or predict annual rainfall by analyzing only one of the two large-scale mechanisms.

In the case of the PAW, the analysis of mRAI frequency distribution by homogeneous rainfall cluster allows to identify if there is a particular portion of the watershed which is more or less vulnerable to the effects of the ENSO and the AMM. Table 5 shows the results of the chi-squared test in which, for all studied patterns, the p value was higher than 0.05, suggesting the acceptance of the null hypothesis that the frequency distributions are equal. This means that, by accepting the null hypothesis, the occurrence of rainfall anomalies in the entire watershed is equally influenced by large-scale mechanisms.

Positive or negative trends in the mRAI were analyzed annually and by season considering 1%, 5%, and 10% significance levels. Figure 10a shows that there is an annual trend indicating increase in the frequency of negative rainfall anomalies in 24 out of 26 rain gauges in C1. Twenty-five percent of those are negative trends at the 1% level (p value < 0.01), and 71% are non-significant trends (p value > 0.1). Two isolated rain gauges presented positive non-significant (p value > 0.1) trends. Figure 10 b and c show trends when considering only

Table 5 Results of the chi-squared test between the frequency of distribution of the 6-month mRAI in C1 and C2 at a 5% significance level

Pattern		p value
ENSO +		
	Wet season	0.86
	Dry season	0.43
ENSO −		
	Wet season	0.34
	Dry season	0.27
AMM +		
	Wet season	0.23
	Dry season	0.38
AMM −		
	Wet season	0.56
	Dry season	0.88

wet season and dry season months, respectively. In these cases, one can notice that 30.0% of the C1 rain gauges presented significant negative mRAI trends during the wet season. This indicates that negative rainfall anomalies are becoming more frequent in part of the upper PAW, which may represent a risk to the entire watershed because it is its main recharge zone. Results are less conclusive in the dry season (Fig. 10c) when most gauges presented non-significant either positive or negative trends. However, it is important to note that during the dry season, rainfall ranges from 0 to 20 mm per month in average, and therefore, even small deviations from the mean may incur in relevant changes when identifying anomalies through the mRAI, even when using a 6-month aggregated timescale. Thus, although results indicate that some gauges show increase in the frequency of occurrence of positive anomalies during the dry season, it does not mean that annual precipitated volumes will be significantly impacted.

Results were similar in C2, with negative annual mRAI trends in 26 out of 30 rain gauges (out of which 31% are significant—p value < 0.1) and four gauges with non-significant (p value > 0.1) positive trends (Fig. 10a). In the wet season, 33% of total gauges also presented significant (p value < 0.1) negative mRAI trends (Fig. 10b). It is important to note that in C2, although there were relatively less gauges portraying significant trends if compared with C1, they were well distributed throughout the cluster region. This indicates that the increase in the occurrence of negative rainfall anomalies was captured in all the extension of the basin. Regarding the dry season, results were similar to the ones found in C1, with inconclusive results due to the nature of monthly rainfall during these periods.

Trends found in this study reaffirm results found in previous studies in Northeast Brazil, which indicated negative trends in precipitated volumes, including in the semiarid region, particularly during the wet season (de Oliveira et al.
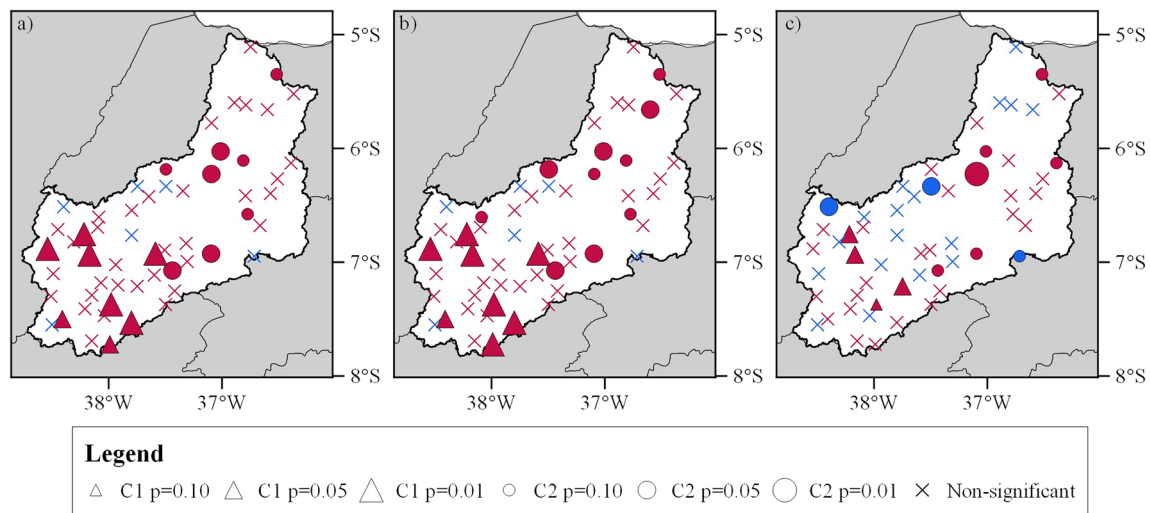
Fig. 10 Trends in the mRAI in C1 and C2 at different significance levels. (a) Annual. (b) Wet season. (c) Dry season. Red symbols represent negative trends and blue symbols represent positive trends

2014, de Oliveira et al. 2017; Lacerda et al. 2015; Marengo and Bernasconi 2015; Marengo et al. 2017a, b; Bezerra et al. 2018; Dubreuil et al. 2018; da Silva et al. 2018). The present study, however, is different in the sense that it found similar results at the scale of watersheds instead of regional scale. In the PAW, although not all rain gauges presented significant trends, the spatial distribution of those that indeed portrayed negative trends was relatively homogeneous. The few stations that presented positive anomalies should be looked into individually, as they are probably being influenced by local factors. Results found in C1 seem to be more conclusive as to the increase of negative rainfall anomalies, especially in the wet season in the southernmost portions of the basin.

4 Conclusion

The objective of this study was to present a comprehensive approach for the characterization of rainfall climatology over watersheds, particularly those located in semiarid regions lacking consistent measured data, with the PAW as an example. This study also advocates the use of renowned but simple techniques in order to gap fill monthly data time series, identify homogeneous rainfall subregions, assess monthly and annual rainfall anomalies (through the mRAI) in relation to teleconnection patterns, and analyze trends in the occurrence of positive and negative rainfall anomalies. Thus, we hope the proposed approach might be replicated for the climatological analysis of rainfall in other watersheds which share similar climate and data availability characteristics.

Gap filling of monthly data with up to 30% of missing data performed better when combining different spatial interpolation techniques. Choosing the best results among the MLAD, SBE, and IDW for each month and each station improved monthly

gap filling performance up to 18.2% if compared with choosing only one method. Cluster analysis allowed the identification of two homogeneous regions with different rainfall patterns. Regarding teleconnection patterns associated with rainfall anomalies (mRAI), results corroborated with previous studies in Northeast Brazil and the semiarid region. In years when the ENSO (AMM) was in its positive (negative) phase, there was a higher probability for the occurrence of months with above-average rainfall, while the opposite was also true. Trend analysis showed that there is evidence of an increase in the occurrence of months with negative mRAI, that is, with below-average rainfall, especially during the wet season. In the C1 region, there were more rain gauges in which significant negative trends were identified if compared with C2, indicating that it might be more vulnerable to potential drastic reductions in rainfall volumes, and therefore, the recharge of water storages in the rest of the watershed might be compromised.

The main limitations of this study are inherent to the database itself, which presents gaps of up to 30%. Although they have been properly and satisfactorily taken care of, important information might have been lost due to incomplete database. For example, despite rainfall anomalies could have been calculated and assessed, extreme event analysis at finer scales would be unfeasible with the used database. Alternatives could have been adopted such as the merging of databases of different types (satellite data or reanalyzed data), although computational effort would greatly increase. Furthermore, the proposed teleconnection patterns analysis is rather simple, and more conclusive results could have been found using more robust methods that, for example, would allow the identification of the combined effect of the two large-scale mechanisms which were considered in this study. It is also important to take note that this methodology should be adapted according to the climatological rainfall and teleconnection patterns of each watershed and region.

References

Almazroui M, Dambul R, Islam MN, Jones PD (2015) Principal components-based regionalization of the Saudi Arabian climate. Int J Climatol 35(9):2555–2573. https://doi.org/10.1002/joc.4139

Alvares CA, Stape JL, Sentelhas PC, de Moraes Gonçalves JL, Sparovek G (2014) Köppen's climate classification map for Brazil. Meteorol Z 22(6):711–728. https://doi.org/10.1127/0941-2948/2013/0507

ANA (2014) Agência Nacional de Águas, Plano de Recursos Hídricos da Bacia Hidrográfica do Rio Piranhas-Açu. Brasília, DF, Brasil. http://piranhasacu.ana.gov.br/.

Anderberg MR (1973) Cluster analysis for applications, 1ed edn. Academic Press

Barrodale I, Roberts FDK (1973) An improved algorithm for discrete L1 linear approximation. SIAM J 10(5):839–848

Bezerra BG et al (2018) Changes of precipitation extremes indices in São Francisco River Basin, Brazil from 1947 to 2012. Theor Appl Climatol:1–12. https://doi.org/10.1007/s00704-018-2396-6

Brito SSB, Cunha APMA, Cunningham CC, Alvalá RC, Marengo JA, Carvalho MA (2018) Frequency, duration and severity of drought in the semiarid Northeast Brazil region. Int J Climatol 38(2):517–529. https://doi.org/10.1002/joc.5225

Charrad M et al (2014) NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61(6):1–36. https://doi.org/10.18637/jss.v061.i06

Chen FW, Liu CW (2012) Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy Water Environ 10(3):209–222. https://doi.org/10.1007/s10333-012-0319-1

Chen M, Shi W, Xie P, Silva VBS, Kousky VE, Wayne Higgins R, Janowiak JE (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. J Geophys Res Atmos 113(4):1–13. https://doi.org/10.1029/2007JD009132

Correia Filho WLF, Lucio PS, Spyrides MHC (2016) Caracterização dos extremos de precipitação diária no nordeste do Brasil. Bol Goiano Geografia 36(3):539–554

Costa DD, da Silva Pereira TA, Fragoso CR, Madani K, Uvo CB (2016) Understanding drought dynamics during dry season in Eastern Northeast Brazil. Front Earth Sci 4(69):1–11. https://doi.org/10.3389/feart.2016.00069

Daultrey S (1976) 'Principal components analysis', Concepts and techniques in modern geography, 930, pp. 527–47. doi: https://doi.org/10.1007/978-1-62703-059-5_22

de Andrade EM et al (2016) Uncertainties of the rainfall regime in a tropical semi-arid region: the case of the State of Ceará. Rev Agro@mbiente On-line 10(2):88–95. https://doi.org/10.18227/1982-8470ragro.v10i2.3500

de Amorim AL, Ribeiro MMR, Braga CFC (2016) Conflitos em bacias hidrográficas compartilhadas: o caso da bacia do rio Piranhas-Açu/PB-RN. Rev Bras Recur Hidr 21(1):36–45

de Oliveira PT, Santos e Silva CM, Lima KC (2014) Linear trend of occurrence and intensity of heavy rainfall events on Northeast Brazil. Atmos Sci Lett 15(3):172–177. https://doi.org/10.1002/asl2.484

de Oliveira PT, Santos e Silva CM, Lima KC (2017) Climatology and trend analysis of extreme precipitation in subregions of Northeast Brazil. Theor Appl Climatol 130(1–2):77–90. https://doi.org/10.1007/s00704-016-1865-z

Dubreuil V, Fante KP, Planchon O, Sant'Anna Neto JL (2018) Climate change evidence in Brazil from Köppen's climate annual types frequency. Int J Climatol 39(3):1446–1456. https://doi.org/10.1002/joc.5893

Eischeid JK, Pasteris PA, Diaz HF, Plantico MS, Lott NJ (2000) Creating a serially complete, national daily time series of temperature and precipitation for the Western United States. J Appl Meteor 39(9):1580–1591. https://doi.org/10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2

Eklundh L, Pilesjö P (1990) Regionalization and spatial estimation of ethiopian mean annual rainfall. Int J Climatol 10(5):473–494. https://doi.org/10.1002/joc.3370100505

Fazel N et al (2018) Regionalization of precipitation characteristics in Iran's Lake Urmia basin. Theor Appl Climatol 132(1–2):363–373. https://doi.org/10.1007/s00704-017-2090-0

de Felix VS (2015) Análise de 40 anos de precipitação pluviométrica da Bacia Hidrográfica do Rio Espinharas-PB. Rev Bras Geogr 8(5):1347–1358

Frazier AG, Giambelluca TW, Diaz HF, Needham HL (2016) Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian Islands. Int J Climatol 36(3):1459–1470. https://doi.org/10.1002/joc.4437

Gadgil S, Iyengar RN (1980) Cluster analysis of rainfall gauges of the Indian peninsula. Q J R Meteorol Soc 106(450):873–886. https://doi.org/10.1002/qj.49710645016

Giambelluca TW, Chen Q, Frazier AG, Price JP, Chen YL, Chu PS, Eischeid JK, Delparte DM (2013) Online rainfall atlas of Hawai'i. Bull Am Meteorol Soc 94(3):313–316. https://doi.org/10.1175/BAMS-D-11-00228.1

Gocic M, Trajkovic S (2014) Spatio-temporal patterns of precipitation in Serbia. Theor Appl Climatol 117(3–4):419–431. https://doi.org/10.1007/s00704-013-1017-7

Gong X, Richman MB (1995) On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J Clim 8:897–931. https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2

Goossens C, Berger A (1986) Annual and seasonal climatic variations over the northern hemisphere and Europe during the last century. Ann Geophys 4(4):385–400

Grimm AM, Tedeschi RG (2009) ENSO and extreme rainfall events in South America. J Clim 22(4):1589–1609. https://doi.org/10.1175/2008JCLI2429.1

Hänsel S, Schucknecht A, Matschullat J (2016) The modified Rainfall Anomaly Index (mRAI) — is this an alternative to the Standardised Precipitation Index (SPI) in evaluating future extreme precipitation characteristics ? Theor Appl Climatol 123:827–844. https://doi.org/10.1007/s00704-015-1389-y

Hastenrath S (2006) Circulation and teleconnection mechanisms of Northeast Brazil droughts. Prog Oceanogr 70(2–4):407–415. https://doi.org/10.1016/j.pocean.2005.07.004

Hastenrath S (2012) Exploring the climate problems of Brazil's Nordeste: a review. Clim Chang 112(2):243–251. https://doi.org/10.1007/s10584-011-0227-1

Hayes M, Svoboda M, Wall N, Widhalm M (2011) The Lincoln declaration on drought indices: universal meteorological drought index recommended. Bull Am Meteorol Soc 92(4):485–488. https://doi.org/10.1175/2010BAMS3103.1

Huang J et al (2016) Global semi-arid climate change over last 60 years. Clim Dyn Springer Berlin Heidelberg 46(3–4):1131–1150. https://doi.org/10.1007/s00382-015-2636-8

IPCC (2014) Central and South America. In: Barros V (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, pp 1499–1566

Kane RP (1997) Prediction of droughts in north-east Brazil: role of ENSO and use of periodicities. Int J Climatol 17:655–665

Kannan S, Ghosh S (2011) Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stoch Env Res Risk A 25(4):457–474. https://doi.org/10.1007/s00477-010-0415-y

Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis, 1ed edn. John Wiley & Sons Inc., Hoboken, New Jersey

Kendall MG (1938) A new measure of rank correlation. Biometrika 30(1/2):81–93

Kumar MN et al (2009) On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. Meteorol Appl 16(April 2009):381–389. https://doi.org/10.1002/met

Lacerda F et al (2015) Long-term temperature and rainfall trends over Northeast Brazil and Cape Verde. Earth Sci Clim Change 6(8):1–8. https://doi.org/10.4172/2157-7617.1000296

Lee S, Cho S, Wong PM (1998) Rainfall prediction using artificial neural networks. J Geogr Inf Decis Anal 2(2):233–242

Lyra GB, Oliveira-Júnior JF, 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

Mann HB (1945) Nonparametric tests against trend. Econometrica 13(3):245–259

Marengo JA, Bernasconi M (2015) Regional differences in aridity / drought conditions over Northeast Brazil : present state and future projections. Clim Chang 129:103–115. https://doi.org/10.1007/s10584-014-1310-1

Marengo JA et al (2011) Variabilidade e mudanças climáticas no semiárido brasileiro. In: Recursos hídricos em regiões áridas e semiáridas. Campina Grande, Instituto Nacional do Semiárido, pp 383–422

Marengo JA et al (2017a) Climatic characteristics of the 2010-2016 drought in the semiarid Northeast Brazil region. An Acad Bras Cienc 90:1–13. https://doi.org/10.1590/0001-3765201720170206

Marengo J, Torres RR, Alves LM (2017b) 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

Marengo JA, Alves LM, Alvala RCS, Cunha AP, Brito S, Moraes OLL (2018) Climatic characteristics of the 2010-2016 drought in the semiarid Northeast Brazil region. An Acad Bras Cienc 90:1973–1985

Melo MMMS, Santos CAC, Olinda RA, Silva MT, Abrahão R, Ruiz-Alvarez O (2018) Trends in temperature and rainfall extremes near the Artifical Sobradinho Lake, Brazil. Rev Bras Meteorol 33(3):426–440

Mondal A, Kundu S, Mukhopadhyay A (2012) Rainfall trend analysis by Mann-Kendall test : a case study of north-eastern part of Cuttack district, Orissa. Int J Geol Earth Environ Sci 2(1):70–78

de Moscati MCL, Gan MA (2007) Rainfall variability in the rainy season of semiarid zone of Northeast Brazil (NEB) and its relation to wind regime. Int J Climatol 27:493–512. https://doi.org/10.1002/joc.1408

Mutti PR, da Silva LL, Medeiros SS, Dubreuil V, Mendes KR, Marques TV, Lúcio PS, Santos e Silva CM, Bezerra BG (2019) Basin scale rainfall-evapotranspiration dynamics in a tropical semiarid environment during dry and wet years. Int J Appl Earth Observ Geoinf Elsevier 75(October 2018):29–43. https://doi.org/10.1016/j.jag.2018.10.007

Plackett RL (1983) Karl Pearson and the chi-squared test. Int Stat Rev 51(1):59–72

Rao VB, Hada K (1990) Characteristics of rainfall over Brazil: annual variations and connections with the Southern Oscillation. Theor Appl Climatol 42(2):81–91. https://doi.org/10.1007/BF00868215

Rao VB, Franchito SH, Santo CME, Gan MA (2016) An update on the rainfall characteristics of Brazil: seasonal variations and trends in 1979-2011. Int J Climatol 36(1):291–302. https://doi.org/10.1002/joc.4345

Rau P, Bourrel L, Labat D, Melo P, Dewitte B, Frappart F, Lavado W, Felipe O (2017) Regionalization of rainfall over the Peruvian Pacific slope and coast. Int J Climatol 37(March 2016):143–158. https://doi.org/10.1002/joc.4693

Reboita MS et al (2016) Causas da semi-aridez do Sertão Nordestino. Rev Bras Climatol 19(12):254–277

Rodell M, Famiglietti JS, Wiese DN, Reager JT, Beaudoing HK, Landerer FW, Lo MH (2018) Emerging trends in global freshwater availability. Nature 557(7707):651–659. https://doi.org/10.1038/s41586-018-0123-1

Rodriguez-Puebla C, Encinas AH, Nieto S, Garmendia J (1998) Spatial and temporal patterns of annual precipitation variability over the Iberian peninsula. Int J Climatol 18:299–316

van Rooy MP (1965) A rainfall anomaly index independent of time and space. Notos 14:43

Sa'adi Z, Shahid S, Ismail T, Chung ES, Wang XJ (2017) Trends analysis of rainfall and rainfall extremes in Sarawak , Malaysia using modified Mann – Kendall test. Meteorol Atmos Phys Springer Vienna (0123456789). https://doi.org/10.1007/s00703-017-0564-3

Searcy, J. K., Hardison, C. H. and Langbein, W. B. (1960) Double-mass curves, Geological Surbey Water Supply Paper 1541-B. Washington, DC

da Silva DF, de Sousa FAS, Kayano MT (2009) Uso de IAC e ondeletas para análise da influência das multi-escalas temporais na precipitação da bacia do Rio Mundaú. Engenharia Ambiental - Espírito Santo do Pinhal 6(1):180–195

da Silva RM, Santos CAG, Moreira M, Corte-Real J, Silva VCL, Medeiros IC (2015) Rainfall and river flow trends using Mann – Kendall and Sen's slope estimator statistical tests in the Cobres River basin. Nat Hazards 77:1205–1221. https://doi.org/10.1007/s11069-015-1644-7

da Silva PE et al (2018) Precipitation and air temperature extremes in the Amazon and Northeast Brazil. Int J Climatol (January):1–17. https://doi.org/10.1002/joc.5829

Singh CV (2006) Pattern characteristics of Indian monsoon rainfall using principal component analysis (PCA). Atmos Res 114(3):317–328. https://doi.org/10.1016/j.atmosres.2005.05.006

Stagge JH, Tallaksen LM, Gudmundsson L, van Loon AF, Stahl K (2015) Candidate distributions for climatological drought indices (SPI and SPEI). Int J Climatol 35(13):4027–4040. https://doi.org/10.1002/joc.4267

Stone R, Auliciems A (1992) SOI phase relationships with rainfall in eastern Australia. Int J Climatol 12(6):625–636. https://doi.org/10.1002/joc.3370120608

Tedeschi RG, Grimm M, Cavalcanti IFA (2016) Influence of Central and East ENSO on precipitation and its extreme events in South America during austral autumn and winter. Int J Climatol 36:4797–4814. https://doi.org/10.1002/joc.4670

Teixeira M da S, Satyamurty P (2011) Trends in the frequency of intense precipitation events in southern and southeastern Brazil during 1960-2004. J Clim 24(7):1913–1921. https://doi.org/10.1175/2011JCLI3511.1

Timmermann A, An SI, Kug JS, Jin FF, Cai W, Capotondi A, Cobb KM, Lengaigne M, McPhaden MJ, Stuecker MF, Stein K, Wittenberg AT, Yun KS, Bayr T, Chen HC, Chikamoto Y, Dewitte B, Dommenget D, Grothe P, Guilyardi E, Ham YG, Hayashi M, Ineson S, Kang D, Kim S, Kim WM, Lee JY, Li T, Luo JJ, McGregor S, Planton Y, Power S, Rashid H, Ren HL, Santoso A, Takahashi K, Todd A, Wang G, Wang G, Xie R, Yang WH, Yeh SW, Yoon J, Zeller E, Zhang X (2018) El Niño – Southern Oscillation complexity. Nat Springer US 559:535–545. https://doi.org/10.1038/s41586-018-0252-6

Tinôco ICM et al (2018) Characterization of rainfall patterns in the semi-arid Brazil. Anuário do Instituto de Geociências - UFRJ 41(2):397–409

Uvo CB, Repelli CA, Zebiak SE, Kushnir Y (1998) The relationships between tropical Pacific and Atlantic SST and Northeast Brazil monthly precipitation. J Clim 11:551–562

Wagner PD et al (2012) Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. J Hydrol Elsevier BV 464–465:388–400. https://doi.org/10.1016/j.jhydrol.2012.07.026

Wu H, Qian H (2017) Innovative trend analysis of annual and seasonal rainfall and extreme values in Shaanxi, China , since the 1950s. Int J Climatol 37(August 2016):2582–2592. https://doi.org/10.1002/joc.4866

Xavier AC, King CW, Scanlon BR (2016) Daily gridded meteorological variables in Brazil (1980-2013). Int J Climatol 2659(October 2015):2644–2659. https://doi.org/10.1002/joc.4518

Zilli MT, Carvalho LMV, Liebmann B, Silva Dias MA (2017) A comprehensive analysis of trends in extreme precipitation over south-eastern coast of Brazil'. Int J Climatol 37(August 2016):2269–2279. https://doi.org/10.1002/joc.4840