Journal of Atmospheric Science Research
https://ojs.bilpublishing.com/index.php/jasr

ARTICLE

Processing of Rainfall Time Series Data in the State of Rio de Janeiro

Givanildo de Gois[1] **José Francisco de Oliveira-Júnior**[2*]

1. Federal University of Rondônia Foundation (UNIR), Porto Velho, Rondônia, 76801-059, Brazil
2. Applied Meteorology and Environment Laboratory (LAMMA), Institute of Atmospheric Sciences (ICAT), Federal University of Alagoas (UFAL), Maceió, Alagoas, 57072-900, Brazil

ARTICLE INFO

Article history
Received: 20 August 2021
Accepted: 11 October 2021
Published: 20 October 2021

Keywords:
Data validation
Parametric tests
Cluster analysis

ABSTRACT

The goal was to perform the filling, consistency and processing of the rainfall time series data from 1943 to 2013 in five regions of the state. Data were obtained from several sources (ANA, CPRM, INMET, SERLA and LIGHT), totaling 23 stations. The time series (raw data) showed failures that were filled with data from TRMM satellite via 3B43 product, and with the climatological normal from INMET. The 3B43 product was used from 1998 to 2013 and the climatological normal over the 1947-1997 period. Data were submitted to descriptive and exploratory analysis, parametric tests (Shapiro-Wilks and Bartlett), cluster analysis (CA), and data processing (Box Cox) in the 23 stations. Descriptive analysis of the raw data consistency showed a probability of occurrence above 75% (high time variability). Through the CA, two homogeneous rainfall groups (G1 and G2) were defined. The group G1 and G2 represent 77.01% and 22.99% of the rainfall occurring in SRJ, respectively. Box Cox Processing was effective in stabilizing the normality of the residuals and homogeneity of variance of the monthly rainfall time series of the five regions of the state. Data from 3B43 product and the climatological normal can be used as an alternative source of quality data for gap filling.

1. Introduction

Rainfall is the climatic variable with high space-time variability; it can cause significant damage to society, and affect many human activities around the world [1,2]. Rainfall interferes with several socials productive sectors: such as use and power generation, tourism, agriculture, industrial production, building, aviation, and population's health problems, among others [1,3-5].

State of Rio de Janeiro (SRJ) has a high annual rainfall rate and a complex topography [6]. However, it has a heterogeneous rainfall spatial distribution, because of the interaction with the topography with coastal environment and weather systems that influence this heterogeneity [7]. Most of the studies on the rainfall in SRJ have been restricted to its time variability. There are few analyses that approach rainfall time trends without identifying the main weather systems [8,9]. The main weather systems that operate in the SRJ range from synoptic scale to site, and are classified as producers and inhibitors of rainfall. These include: Frontal Systems (FS), Mesoscale Convective Systems (MCS),

*Corresponding Author:
José Francisco de Oliveira-Júnior,
Applied Meteorology and Environment Laboratory (LAMMA), Institute of Atmospheric Sciences (ICAT), Federal University of Alagoas (UFAL), Maceió, Alagoas, 57072-900, Brazil;
Email: jose.junior@icat.ufal.br

DOI: https://doi.org/10.30564/jasr.v4i4.3603

South Atlantic Convergence Zone (SACZ), South Atlantic Subtropical High (SASH), Atmospheric Blocking (AB), wind systems (valley/ mountain, lake, bay and marine/ terrestrial), Instability Lines (IL), Mesoscale Convective Complexes (MCC), orographic rainfall, convective storms, and others [4,7-12].

These systems cause varying intensities of rainfall, depending on the location and topography of the region, or inhibit or cause strong droughts and dry spells in SRJ [4,13]. The presence of the Maciços de Pedra Branca, Tijuca and Gericinó, which comprises the Metropolitana region of Rio de Janeiro (MRJR), together with *Serra da Mantiqueira* (SW) and *Serra do Mar* (coastal) provides a barrier to air displacement at low levels of the atmosphere. This presence result in changes in the flow structure and in the weather conditions from the site and/or adjacent regions, together with Bays de Sepetiba and Guanabara, which dramatically interferes in rainfall patterns [14,15]. Seasonal and annual patterns of weather systems that operate in the SRJ may be influenced by climate variability modes, such as El Niño-Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) [11,16].

SRJ has an irrigated area of 36 thousand hectares, which is small compared to other states of the Southeast region. Although the SRJ does not stand out in the Brazilian agricultural scenario, there was an increase in irrigated agricultural production, especially in regard to the fruit growing in the Norte Fluminense and Nordeste Fluminense regions [13]. Therefore, it is necessary to identify time (seasonal, interannual and decadal) and spatial (regional and large scale) patterns to subsidize activities in the agricultural and forest areas in SRJ [7,16].

To understand the rainfall patterns in a region previous knowledge is needed of the several factors that affect, for example, physiographic or dynamic, but all acting simultaneously in SRJ [7,9]. However, studies conducted over the last decades in SRJ [9,17-20] did not identify the producers and inhibitors of rainfall systems. Followed by the action of climate variability modes, as well as a treatment involving the rainfall time series of the Government regions, which were restricted to short time series with gaps.

There are few studies based on a careful analysis of rainfall time series in SRJ, most of them focusing on the gap filling and consistency of the time series. Therefore, the goal is to perform the filling, consistency and processing of the rainfall time series data from 1943 to 2013 in five government regions of the state of Rio de Janeiro by the application of descriptive and exploratory analysis, parametric tests, cluster analysis and the data processing technique.

2. Materials and Methods

2.1 Study Area

SRJ is located in Southeastern Brazil, between the latitudes 20º 45' 54 and 23º 21' 57" S and the longitudes 40º 57' 59" and 44º 53' 18" W, with an area of e 43,696,054 km². It borders to the northeast (NE) with the Espírito Santo, north and northwest (N-NW) with Minas Gerais, southwest (SW) with São Paulo and with the Atlantic Ocean to south and east (S-E). It has extensive coastline, with about 635 km long, is bathed by the Atlantic Ocean.

Currently, the SRJ is geopolitically divided into 92 municipalities [21], inserted in eight Government regions: *Metropolitana*, *Noroeste Fluminense*, *Norte Fluminense*, *Baixadas Litorâneas*, *Serrana*, *Centro-Sul Fluminense*, *Médio Paraíba* and *Costa Verde* (Figure 1). According to [22] the SRJ has a landscape with high cliffs, to sea and in the interior; hills; hills and valleys; with varied rock formations in bays with different forms of encounter between the sea and the coast; natural tropical forests followed by large areas of the plateau, which stretches to the west of the state. It stands out among the remaining the peak of Agulhas Negras with altitude of 2,787 m, in the Serra da Mantiqueira region. Serra da Mantiqueira is an important transition area of the Southeast region directed to the valley of the Paraíba do Sul River, which has the lowest height of 250 m, crossing the States of São Paulo, Rio de Janeiro and Minas Gerais.

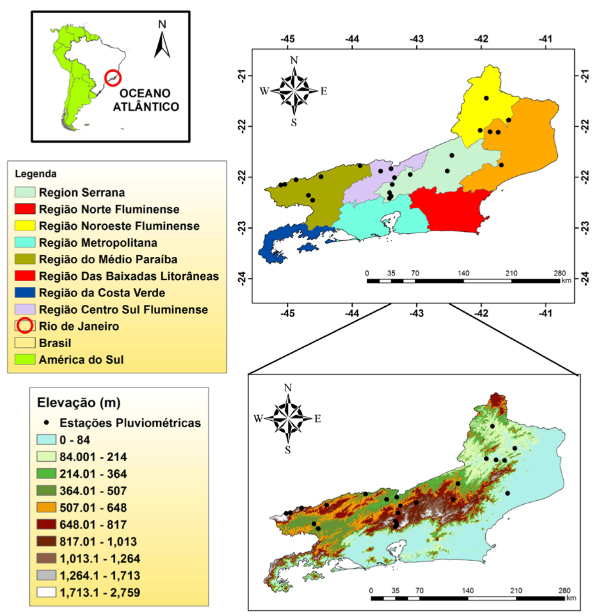


Figure 1. Location of the 23 stations distributed in five Government regions (Norte Fluminense, Noroeste Fluminense, Serrana, Centro Sul Fluminense and Médio Paraíba) of the State of Rio de Janeiro (SRJ).

2.2 Time Series of Rainfall Data 1943-2013

We used rainfall data (mm) from montlhy time series (1943 to2013) of 23 stations, being divided into: pluviometric and automatic and conventional meteorological distributed in five Government regions (*Noroeste Fluminense*, *Norte Fluminense*, *Serrana*, *Centro-Sul Fluminense* and *Médio Paraíba*) of the SRJ - (Table 1). Series was composed by data from ANA (National Water Agency), CPRM (Mineral Resources Research Company), INMET (National Institute of Meteorology), SERLA (State Superintendence of Rivers and Lakes Foundation) and LIGHT (Light Electricity Services S/A). The regions *Baixadas Litorâneas*, *Costa Verde* and *Metropolitana* did not show stations with series from 71 years and, therefore, not entered the study.

Table 1. Identification of the 23 stations distributed in five Government regions (Norte Fluminense, Noroeste Fluminense, Serrana, Centro Sul Fluminense and Médio Paraíba) of the State of Rio de Janeiro (SRJ).

N	Stations	Lat (°)	Long (°)	Altitude (m)
1	São Fidelis	-21.65	-41.75	10.00
2	Cardoso Moreira	-21.49	-41.61	20.00
3	Dois Rios	-21.64	-41.86	50.00
4	Macabuzinho	-22.08	-41.71	19.00
5	Itaperuna	-21.21	-41.91	110.00
6	Três Irmãos	-21.63	-41.99	42.00
7	Aldeia	-21.95	-42.36	376.00
8	Bom Jardim	-22.16	-42.42	530.00
9	Areal	-22.24	-43.10	450.00
10	Moura Brasil	-22.13	-43.15	270.00
11	Paraíba do Sul	-22.16	-43.29	300.00
12	Fazenda Sobradinho	-22.20	-42.90	650.00
13	Itamarati	-22.49	-43.15	1025.00
14	Pedro do Rio	-22.33	-43.14	645.00
15	Petrópolis	-22.51	-43.17	890.00
16	Rio da Cidade	-22.44	-43.17	704.00
17	Barra Mansa	-22.54	-44.18	376.00
18	Fazenda Agulhas Negras	-22.34	-44.59	1460.00
19	Ponte do Souza	-22.27	-44.39	950.00
20	Ribeirão de São Joaquim	-22.47	-44.23	620.00
21	Visconde de Mauá	-22.33	-44.54	1030.00
22	Manuel Duarte	-22.09	-43.56	396.00
23	Santa Isabel do Rio Preto	-22.23	-44.06	544.00

2.3 Gap Filling of Monthly Rainfall Time Series 1943-2013

Monthly rainfall time series that showed gaps were filled with data from TRMM (Tropical Rainfall Measuring Mission) satellite via 3B43 product and the normal climatological coming from INMET (Instituto Nacional de Meteorologia). The 3B43 product was used during the period from 1998 to 2013 and the normal climatological from INMET during 1947 to 1997. The 3B43 product was obtained in NetCDF format at the site: www.mirador.gsfc.nasa.gov/collections/TRMM_3B43_007.shtml. The product provides data with spatial from about 30 km and monthly time resolution.

The 3B43 product was converted into ArcGIS software version 10.1®. Conversion of data, originally in $mm.h^{-1}$ to accumulated monthly ($mm.monthly^{-1}$)-[4]. Selecting the TRMM points was made based on proximity of the stations within the study area. In ArcGIS software version 10.1 was used the ArcToolbox-Multidimension Tools, and the conversion tools Make NetCDF Raster Layer and Make NetCDF Table View for the procedure.

2.4 Consistency and Processing of the Rainfall Temporal Series Data 1943-2013

After the gap filling from the data, a raw time series data was built and, at lastly, we performed an exploratory and descriptive data analysis with the assistance of R software version 3.1.1 [23]. In the descriptive analysis we calculated the mean, median, maximum values, total amplitude, the lower and upper limits, the coefficients of variation (CV, %), asymmetry and kurtosis (K), standard deviation (S), lowest (LQ) and upper (UQ) quartile and interquartile amplitude (IQA). Exploratory analysis was based on boxplot and consists in identifying the outliers, which are values that are three times beyond the boxplot interquartile amplitude.

Shapiro-Wilk (*SW*) and Bartlett (*B*) tests were applied, at 0.05 probability level, for the 23 stations for testing the normality of the residues and homogeneity of data variance. SW test is used when the sample size to be tested is lower than 2,000 observations [24]. If there is no normality of the residues and homogeneity of raw data variance, the variance of the series should be stabilized before any procedure.

SW test was applied to the rainfall time series according to a normal probability distribution. It consists of the ratio from two different variance estimators. The estimator in the numerator based on a linear combination of amounts related to statistics of order from the normal distribution. The estimator in the denominator was obtained in a conventional way.

Statistic from *SW* test, *W*, is defined by Equation 1, given by:

$$W = \frac{\left[\sum_{i=1}^{k} a_{n-i+1}\left(y_{n-i+1} - y_i\right)\right]^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{\left[\sum_{i=1}^{k} a_i\, y_{(i)}\right]^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (1)$$

Wherein, i = 1, 2,....n, is the sample size; y_i= Sample measurement value under analysis ordered from lowest to highest for the value ; $\bar{y}$ = mean value of measuring; *ai* = coefficient generated from the middle, variance and covariance of statistical order from a sample with n size and a normal distribution.

Wherein *X* is a feature of study, we formulate the hypothesis:

H_0: Rainfall data from the stations show residues with normal distribution (Gaussian);

H_1: Rainfall data from the stations do not show residues with normal distribution (Gaussian).

The conditions so that the data of the stations are distributed according to a normal distribution at probability level α were that:

For $W_{cal} \leq W_{tab}$ we rejected H_0 for P-value α < 0.05;

For $W_{cal} \geq W_{tab}$ we accept H_0 for P-value α > 0.05.

[25] test proposed by [26], was employed to verify the assumption that *K* samples from a population shows equal variances, i.e., homogeneity of variance. Statistic from Bartlett's test, *B0* is determined by the following equations:

$$N = \sum_{j=1}^{n} n_j \quad (2)$$

$$S_i^2 = \sum_{j=1}^{n_i} \frac{\left(y_j - \overline{y_i}\right)^2}{n_i - 1} \quad (3)$$

$$S_p^2 = \frac{1}{N-k}\sum_{i=1}^{k}\left(n_i - 1\right)S_i^2 \quad (4)$$

$$q = \left(N-k\right)\text{h}\ S_p^2 - \sum_{i=1}^{k}\left[\left(n_i - 1\right)\text{h}\ S_i^2\right] \quad (5)$$

$$c = 1 + \frac{1}{3\left(k-1\right)}\left(\sum_{i=1}^{n}\frac{1}{n_i - 1} - \frac{1}{N-k}\right) \quad (6)$$

Wherein B_0 is defined as:

$$B_0 = \frac{q}{c} \quad (7)$$

B_0 on the hypothesis $H_0 \approx \chi^2_{k-1}$

Wherein, N = number of observations, *ni* and k = number of observations within groups, S_i^2 = sample variance, S_p^2 = population variance, q= coefficient of numerator, c = coefficient of denominator, χ^2_{k-1} = chi-square distribution , a= significance level and B_0 = statistic from Bartlett's test.

Wherein *X* is a feature of study, we formulate the following hypotheses:

H_0: Rainfall data from the stations show homogeneous variances.

H_1: Rainfall data from the stations do not show homogeneous variances.

The conditions so that the data from the statins show homogeneity or heterogeneity of variances at probability level α were that:

$B_0 \geq \chi^2_{(1-\alpha,\,k-1)}$ we reject H_0 for P-value α < 0.05;

$B_0 \leq \chi^2_{(1-\alpha,\,k-1)}$ we accept H_0 for P-value α > 0.05.

is common applying a processing to the data set. Thus, in this situation was used the [27] processing. The method consists on estimating multiple values for lambda parameter (λ). In the study the quadratic processing was used in the raw time series, given by Equation 8:

$$y(\lambda) = \frac{\left(x^{\lambda} - 1\right)}{\lambda} \qquad \lambda \neq 0 \quad (8)$$

wherein, x = raw data and λ= lambda.

Later it was performed the cluster analysis of time series data processed by software environment R version 3.1.1 [23]. Thus, we determined the respective numbers of groups and the dendrogram. Number of groups adopted and stratification of stations was based on Ward's hierarchical agglomerative method [28] through the dissimilarity measure from Euclidean distance [29,30].

Euclidean distance is given by:

$$d_E = \sqrt{\sum_{j=1}^{p}(x_j - x_k)^2} \quad (9)$$

wherein, dE = Euclidean distance; x_{ij} and x_{kj} = quantitative parameters j from p and k individuals, respectively.

On [29] the distance between two clusters is the sum of the squares among the two clusters made on all variables. In this method, we minimize the dissimilarity or minimize the total of the sums of squares within groups, i.e., occurs by the homogeneity within each group and heterogeneity out of each group [31].

$$W = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 \quad (10)$$

wherein, W = within group homogeneity and heterogeneity by the sum the squares of the deviations; n = number of analyzed values; xi = i-th element of the cluster.

According to [32], the method reveals itself as one of the most suitable on the cluster analysis, wherein the time series rainfall data are arranged in array form $P_{(n\,x\,p)}$ where P_{ij} represents the i-th variable value (site) of the j-th individual (monthly). Each row vector represents the rainfall within the year and each column vector represents the season.

After the cluster analysis, we separate between the five study regions the stations that showed better performance according to the SW and B parametric tests, and were subsequently made to analyze the position and dispersion measures, in order to verify how the empirical distribution approaches from the normal, which is statistically evidenced by non-parametric test.

Considering that the position and dispersion measurements are influenced by the presence of discrepant values (outliers) or extreme, an exploratory data analysis was performed to detect the presence of outliers using descriptive statistics previously cited, followed by the boxplot and normal probability graph. Again, we used the software R version 3.1.1 [23], for calculating the all indexes and statistical graphs.

After applying the Box Cox processing, the stations that had better performance according to *SW* and *B* test were separated. Subsequently, the analysis of the position and dispersion measurements was performed in order to verify how the empirical distribution approaches from the normal, which is evidenced statistically by non-parametric tests.

3. Results

3.1 Descriptive Statistics and Exploratory Data Analysis of Raw Rainfall Data

Time series monthly rainfall in the State Government regions with a probability of occurrence above 75% showed high temporal variability, mainly in the maximum parameter ranging among 397.00 to 1277.90 mm, followed by total amplitudes in the same proportion (Tables 2 to 6). Sample CV for all existing stations in five Government regions was higher than 70%. This indicates a significant dispersion and heterogeneity of raw rainfall data, allowing highlight important features in the Government regions, regarding to producers and inhibitors of rainfall systems, in the SRJ. However, *Moura Brasil* (217.66%) and *Paraíba do Sul* (105.40%,) stations stood out over the others with the highest sample CV, respectively (Table 4).

In the analysis of the lower and upper limits as outlier's delimiters, we observed values higher than the 75th percentile from the time series in all seasons outside of this range, confirmed by standard deviation and the IQA (Figure 2). The highest standard deviations about the mean were: in Norte Fluminense region, at *São Fidelis* (80.03 mm), *Dois Rios* (81.04) and *Macabuzinho* (94.77) regions; in *Noroeste Fluminense*, at *Itaperuna* (90.95 mm) and *Três Irmãos* (88.45 mm); in *Centro-Sul Fluminense*, at Areal (103.78 mm) and *Paraíba do Sul* (106.54 mm). In *Serrana* region, at the *Aldeia* (102.42 mm), *Bom Jardim* (112.37 mm) and *Fazenda Sobradinho* (109.44 mm) stations; in the *Médio Paraíba* regions, at the *Barra Mansa* (99.63 mm) and *Manuel Duarte* (106.73 mm) stations.

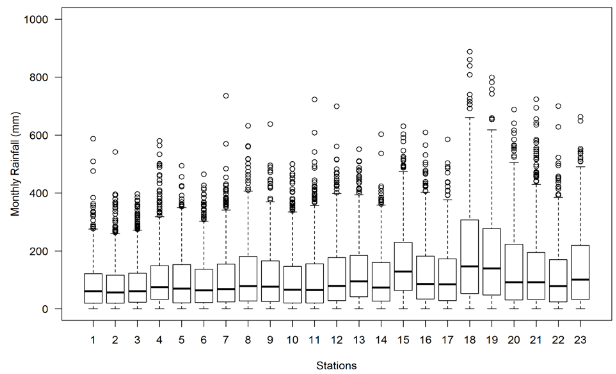
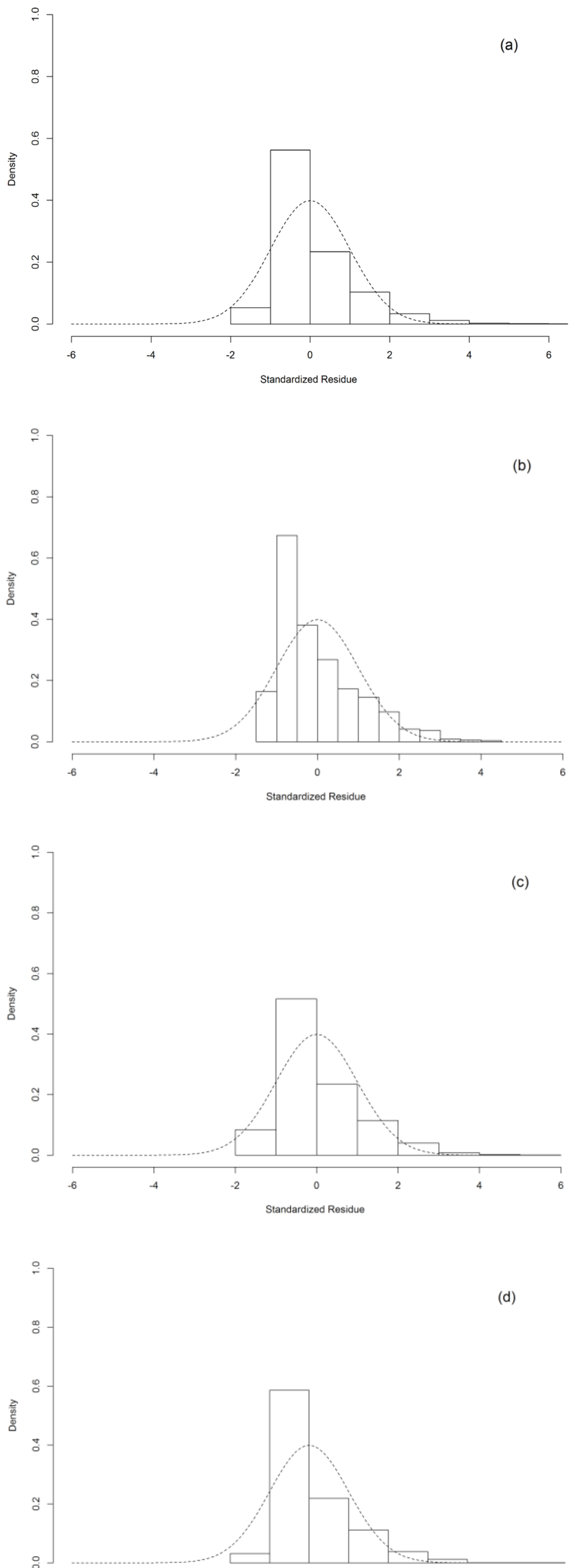


Figure 2. Boxplot of the monthly rainfall time series of raw data from Norte Fluminense, Noroeste Fluminense, Serrana, Centro Sul Fluminense and Médio Paraíba regions of the State of Rio de Janeiro (SRJ).

Through the boxplot and descriptive analysis of the monthly raw rainfall data from the regions of SRJ we verified a similarity of the median parameter in all seasons. The values were lower than the means, being the means influenced by high rainfall values (outliers), which indicates that distributions for the five regions of the state are skewed (asymmetric) to the right (Figure 3). However, we observed values equal to median (61.15 mm) at *São Fidelis* and *Dois Rios* stations (Table 2). *Macabuzinho's* station recorded the highest value for mean (102.72 mm) and median (75.30 mm), respectively. *Nordeste Fluminense, Centro-Sul Fluminense, Serrana* and *Médio Paraíba* regions followed the same behavior towards higher values for the mean and median, in the respective, *Itaperuna, Areal, Petrópolis* and *Fazenda Agulhas Negras* stations (Tables 2 to 6).



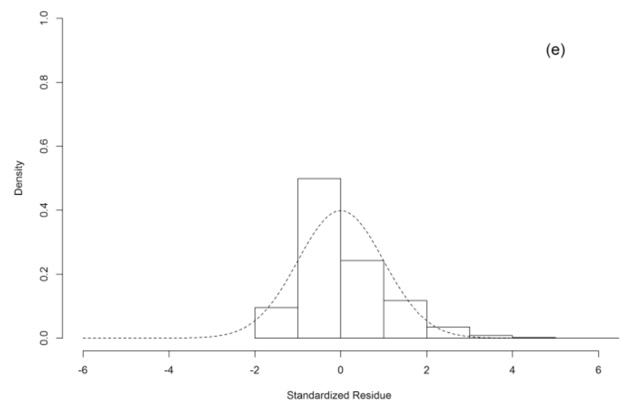


Figure 3. Histograms of the monthly rainfall series of raw data from Norte Fluminense (a), Noroeste Fluminense (b), Serrana (c), Centro Sul Fluminense (d) and Médio Paraíba (e) regions of the State of Rio de Janeiro (SRJ).

Positive asymmetry coefficients (*K*) ranged from 1.16 to 3.60 for the five regions from the State Government. The *K* coefficient characterizes how the data distribution deviates from symmetrical condition. In this case, the tail on the right is more elongated than the tail to the left to middle > 0, which indicates the occurrence of high values with a low frequency - (Figure 3). The *K* coefficient in the *North* region for the *São Fidelis* and *Macabuzinho* stations show that the form of the data distribution is leptokurtic (*K* > 3, more tapered curve), and for the *Cardoso Moreira* and *Dois Rios* stations is platykurtic (*K* < 3, flatter curve). Although the distribution form for the *Cardoso Moreira* station is close to mesokurtic form (*K* = 3, curve neither too flattened nor too tapered), feature of normal distribution (Table 2).

In *Nordeste* region, there is a predominance of a flatter frequency curve (*K* < 3), whose raw data distribution form for the stations from *Itaperuna* and *Dois Rios* is platykurtic (Table 3). The results obtained in *Serrana* region, for the *K* coefficient show that the level of flattening of the frequency curve for all seven stations were platykurtic. However, *Itamarati* station (2.97) obtained *K* coefficient close to 3, where the distribution may be considered in mesokurtic form (Table 5).

The *K* coefficients for the *Centro-Sul* region suggest that the degree of flattening of the frequency curves of the *Areal, Paraíba do Sul* and *Moura Brasil* stations are platykurtic and leptokurtic respectively. However, we observe that in Paraíba station, the *K* coefficient is close to 3, providing mesokurtic distribution (Table 4). The K coefficient for the Médio Paraíba region indicates the platykurtic predominance for *Barra Mansa, Fazenda Agulhas Negras, Ponte do Souza, Visconde de Mauá* and *Manuel Duarte* stations, and leptokurtic for *Ribeirão de São Joaquim* and *Santa Isabel do Rio Preto* (Table 6).

Table 2. Descriptive statistics of the monthly rainfall data (mm) in Norte Fluminense region, State of Rio de Janeiro (SRJ).

Stations	Mean (mm)	Median (mm)	Value	Total Amplitude (mm)	Limit	
			Maximum (mm)		Lower (mm)	Upper (mm)
São Fidelis	82.49	61.15	587.80	587.80	-134.51	275.78
Cardoso Moreira	79.78	56.44	542.00	542.00	-125.20	270.36
Dois Rios	85.16	61.15	397.00	397.00	-126.37	276.94
Macabuzinho	102.72	75.30	580.20	580.20	-140.35	302.71

Stations	Coefficient			Sample Standard Deviation (mm)	Quartile		Interquartile amplitude (mm)
	Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
São Fidelis	97.03	1.54	3.60	80.03	19.35	121.92	102.57
Cardoso Moreira	97.36	1.51	2.82	77.67	19.82	116.50	96.68
Dois Rios	95.17	1.36	1.54	81.04	23.30	123.08	99.78
Macabuzinho	92.26	1.65	3.48	94.77	33.17	148.85	115.68

Table 3. Descriptive statistics for the monthly rainfall data (mm) in *Noroeste* region, State of Rio de Janeiro.

Stations	Mean (mm)	Median (mm)	Value	Total Amplitude (mm)	Limit	
			Maximum (mm)		Lower (mm)	Upper (mm)
Itaperuna	96.01	69.75	494.80	494.80	-176.36	306.43
Três Irmãos	91.66	63.70	464.90	464.90	-149.73	290.01

Stations	Coefficient			Sample standard deviation (mm)	Quartile		Interquartile Amplitude (mm)
	Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
Itaperuna	94.73	1.16	0.95	90.95	21.00	152.57	131.57
Três Irmãos	96.50	1.27	1.22	88.45	21.80	136.15	114.35

Table 4. Descriptive statistics for the monthly rainfall data (mm) in *Centro Sul* region, State of Rio de Janeiro.

Stations	Mean (mm)	Median (mm)	Value	Total Amplitude (mm)	Limit	
			Maximum (mm)		Lower (mm)	Upper (mm)
Areal	109.04	76.20	638.30	638.30	-183.55	319.11
Moura Brasil	96.56	65.88	500.00	500.00	-169.43	300.11
Paraíba do Sul	101.08	64.55	723.00	723.00	-183.10	309.26

Stations	Coefficient			Sample standard deviation (mm)	Quartile		Interquartile Amplitude (mm)
	Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
Areal	95.17	1.25	1.42	103.78	25.73	165.25	139.52
Moura Brasil	217.66	2.87	6.55	39.20	19.98	146.25	126.27
Paraíba do Sul	105.40	1.56	2.85	106.54	20.00	155.40	135.40

Table 5. Descriptive statistics for the monthly rainfall data (mm) in *Serrana* region, State of Rio de Janeiro.

Stations	Mean (mm)	Median (mm)	Value	Total	Limit	
			Maximum (mm)	Amplitude (mm)	Lower (mm)	Upper (mm)
Aldeia	102.59	68.56	735.60	735.60	-171.58	307.71
Bom Jardim	116.20	78.80	632.00	632.00	-201.10	334.26
Fazenda Sobradinho	114.33	79.50	699.00	699.00	-195.05	330.86
Itamarati	126.20	95.30	551.20	551.20	-171.53	337.96
Pedro do Rio	104.70	73.50	603.50	603.50	-173.55	314.06
Petrópolis	158.50	129.00	630.20	630.20	-186.40	383.46
Rio da Cidade	121.04	85.92	609.40	609.40	-187.63	336.06

Stations	Coefficient			Sample standard deviation (mm)	Quartile		Interquartile Amplitude (mm)
	Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
Aldeia	99.83	1.43	2.50	102.42	23.68	153.85	130.17
Bom Jardim	96.73	1.26	1.31	112.37	27.80	180.40	152.60
Fazenda Sobradinho	95.72	1.30	1.70	109.44	28.18	177.00	148.82
Itamarati	84.76	1.12	2.97	7.90	41.85	184.10	142.25
Pedro do Rio	93.88	1.16	1.11	98.32	26.70	160.20	133.50
Petrópolis	77.30	3.52	1.98	9.06	63.20	229.60	166.40
Rio da Cidade	90.83	1.18	1.08	8.12	34.27	182.20	147.93

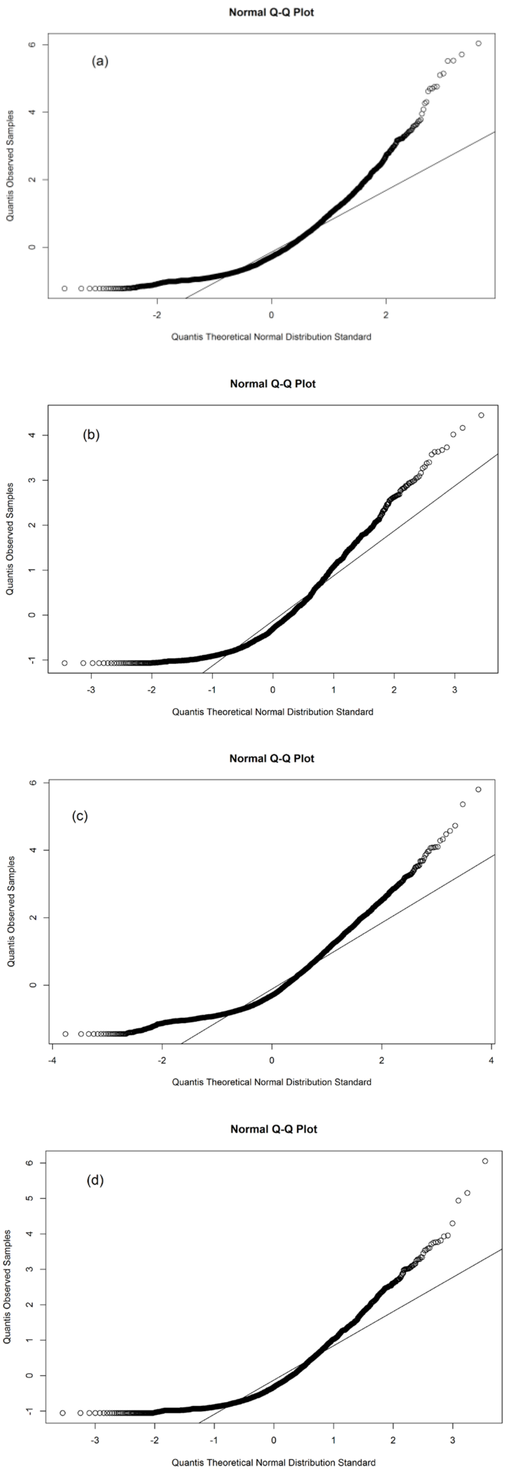
Table 6. Descriptive statistics for the monthly rainfall data (mm) in *Médio Paraíba* region, State of Rio de Janeiro.

Stations	Mean (mm)	Median (mm)	Value	Total	Limit	
			Maximum (mm)	Amplitude (mm)	Lower (mm)	Upper (mm)
Barra Mansa	110.10	84.90	585.60	585.60	-186.78	326.06
Fazenda Agulhas Negras	195.92	147.05	888.20	888.20	-328.72	460.79
Ponte do Souza	175.82	139.40	799.50	799.50	-295.63	430.56
Ribeirão de São Joaquim	141.76	91.85	1243.00	1243.00	-256.91	376.43
Visconde de Mauá	133.50	91.80	724.00	724.00	-208.15	347.46
Manuel Duarte	111.30	78.40	700.00	700.00	-195.33	323.66
Santa Isabel do Rio Preto	141.15	101.09	1277.90	1277.90	-247.05	372.81

Stations	Coefficient			Sample standard deviation (mm)	Quartile		Interquartile Amplitude (mm)
	Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
Barra Mansa	90.50	1.12	1.12	99.63	28.61	172.20	143.59
Fazenda Agulhas Negras	86.81	3.60	1.79	12.57	52.67	306.93	254.26
Ponte do Souza	85.81	3.58	2.17	11.15	47.77	276.70	228.93
Ribeirão de São Joaquim	99.79	1.57	4.55	10.45	30.78	222.57	191.79
Visconde de Mauá	99.03	1.43	1.87	9.77	32.90	193.60	160.70
Manuel Duarte	95.90	1.29	2.01	106.73	23.75	169.80	146.05
Santa Isabel do Rio Preto	95.54	1.60	5.92	9.97	32.55	218.95	186.40

3.2 Processing of Time Series Rainfall Data

Values from *SW* and *B* tests and Q-Q Plt Normal probabilistic graphs (Figure 4) applied to the raw rainfall data series considering 0.05 probability level, indicates the non-normality of residues and heterogeneity of variances in all seasons, being, therefore, verified the non-stability of data variance (Table 7).



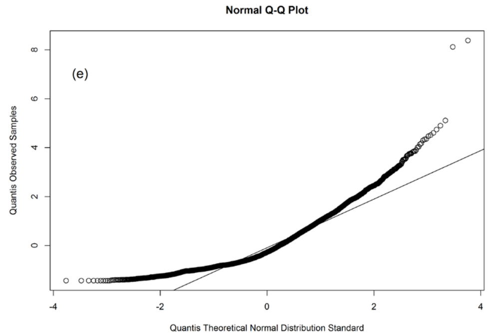


Figure 4. Quartile-Quantile Plot graphs from monthly rainfall series of raw data from Norte Fluminense (a), Noroeste Fluminense (b), Serrana (c), Centro Sul Fluminense (d) and Médio Paraíba (e) regions of the State of Rio de Janeiro (SRJ).

From this, it was necessary to stabilize the variance from the time series by applying the Box-Cox processing, where the λ values ranged from 0.33 to 0.44, (Table 7). Processed data from the *Norte Fluminense, Nordeste Fluminense, Centro-Sul Fluminense, Serrana* and *Médio Paraíba* regions recorded the highest p-value (α > 0.05) for SW and B tests, mainly in Dois Rios (0.05 and 0.159), *Macabuzinho* (0.06 and 0.09), *Itaperuna* (0.37 and 0.23), *Três Irmãos* (0.11 and 0.06), *Paraíba do Sul* (0.28 and 0.27), *Bom Jardim* (0.14 and 0.32) and *Fazenda Sobradinho* (0.43 and 0.27) stations. This shows that after processing of rainfall data there was adjustment to normal distribution (Figure 5 and Table 7).

However, *Médio Paraíba* and *Serrana* regions stood out among the five regions of the state, such as those that record the largest number of stations with the lowest values for p-value ($\alpha < 5\%$) in the stations: *Fazenda Agulhas Negras* (0.01 and 0.001), *Ribeirão de São Joaquim* (0.01 and 0.001), *Manuel Duarte* (0.03 and 0.001), *Santa Isabel do Rio Preto* (0.00 and 0.001) and *Petrópolis* (0.00 and 0.001). Some of the stations from *Norte* and *Centro-Sul Fluminense* regions did not show adjustment of the residues to normal distribution, as *Cardoso Moreira* (0.03 and 0.010) and *Areal* (0.02 and 0.020).

Based on cluster analysis technique (CA) were set two rainfall homogeneous groups (G_1 and G_2) for the 23 stations (Figure 6). The stations belonging to G_1 are located on the slope of the Serra do Mar directed to the interior (20 stations) and the others (3 stations) on the slope of the Serra do Mar directed to the coastal environment (Figure 1). The stations belonging to the group G_2, *Fazenda Agulhas Negras* and *Ponte do Souza* (Médio Paraíba) and *Petropólis* (Serrana) show elevations above 850 m (Table 1). It is important to note the difference in the rainfall regime between the G_1 and G_2, where the first stands out for

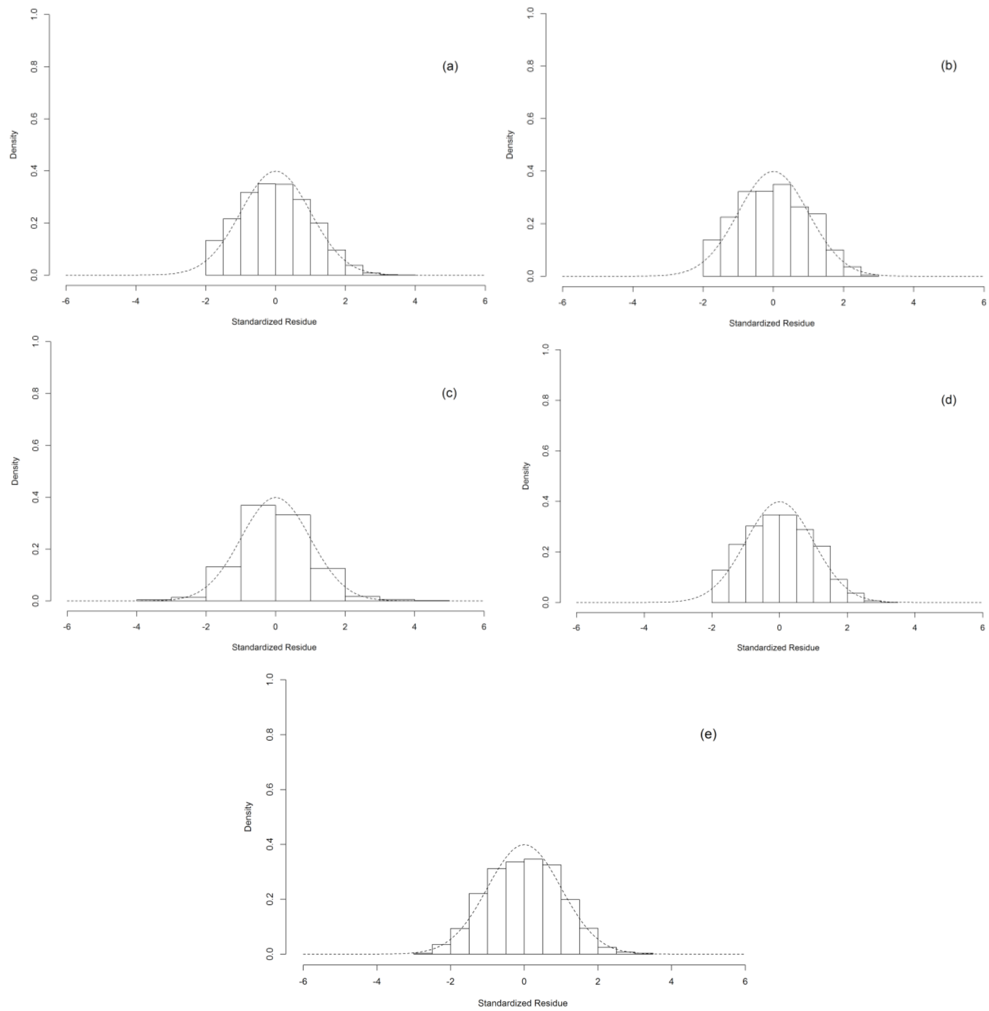


Figure 5. Histograms of the processed data of monthly rainfall series from Norte Fluminense (a), Noroeste Fluminense (b), Serrana (c), Centro Sul Fluminense (d) and Médio Paraíba (e) regions of the State of Rio de Janeiro (SRJ).

its largest annual accumulated 1164.95 mm.year^{-1} compared to group G_2 with an accumulated rainfall of 347.86 mm.year^{-1}, representing, respectively 77.01% and 22.99% of the rainfall in five SRJ Government regions (Table 8).

Box Cox proceeding was effective as processing of the time-series data in accordance with the *SW* and *B* tests, showing the non-violation of normality and homogeneity of variance premises. The best performance based on the CA technique was observed in the group G_1.

Results concerning the descriptive and exploratory analysis of the processed data showed that the *Box-Cox* processing stabilized the data variance (outliers) in all 23 stations. It is important to note the similarity regarding the mean behavior that approached the median values in all regions of the state, compared to the raw data from the time series, previously discussed time series. The distribution form of the monthly rainfall time series for the five Government regions approaches a symmetrical distribution (mean and median practically equal), proven by low asymmetry coefficient values (0.02 to 0.22), indicating that the means of station of the groups G_1 and $G_2$are close to the median (Table 8). The results of the *K* coefficients from the stations belonging to the groups G_1 and G_2reveal that the flattening level of the frequency curve for all stations are platykurtic ($K < 3$), as shown in Figure 5 and Table 8. The obtained results were corroborated by the values from descriptive analysis, which indicates a low temporal variability of the data, with maximum values ranging from 24.10 mm to 31.78 mm in G_1 followed by total amplitudes in the same ratio and upper and lower limits on the order of -4.69 mm to 34.37 mm.

CV analysis for the groups G_1 and G_2 shows that all stations had values above 30%, indicating high dispersion of rainfall data for the whole study area. This situation is corroborated by higher standard deviation and upper IQA, which indicate a high data variability level around the

Table 7. Normality and homogeneity of variance test for the raw and processed rainfall data of the State of Rio de Janeiro (SRJ).

Government regions	Stations	Raw data		Processed data		lambda (λ)
		Shapiro	Bartllet	Shapiro	Bartllet	
		p-value	p-value	p-value	p-value	
NORTE	São Fidelis	2.48×10^{-7}	2.48×10^{-7}	0.01	0.189	0.425
	Cardoso Moreira	2.48×10^{-7}	2.48×10^{-7}	0.03	0.010	0.417
	Dois Rios	2.48×10^{-7}	2.48×10^{-7}	0.05	0.159	0.433
	Macabuzinho	2.48×10^{-7}	2.48×10^{-7}	0.06	0.092	0.401
NOROESTE	Itaperuna	2.48×10^{-7}	2.48×10^{-7}	0.37	0.233	0.432
	Três Irmãos	8.99×10^{-7}	2.48×10^{-7}	0.11	0.065	0.393
CENTRO SUL	Areal	2.48×10^{-7}	2.48×10^{-7}	0.02	0.020	0.425
	Moura Brasil	2.48×10^{-7}	2.48×10^{-7}	0.05	0.000	0.409
	Paraíba do Sul	2.48×10^{-7}	2.48×10^{-7}	0.28	0.273	0.335
SERRANA	Aldeia	2.48×10^{-7}	2.48×10^{-7}	0.02	0.141	0.385
	Bom Jardim	2.48×10^{-7}	2.48×10^{-7}	0.14	0.322	0.409
	Fazenda Sobradinho	2.48×10^{-7}	2.48×10^{-7}	0.43	0.274	0.385
	Itamarati	1.22×10^{-6}	2.48×10^{-7}	0.82	0.001	0.443
	Pedro do Rio	2.48×10^{-7}	2.48×10^{-7}	0.06	0.001	0.411
	Petrópolis	6.55×10^{-6}	2.48×10^{-7}	0.00	0.001	0.565
	Rio da Cidade	2.48×10^{-7}	2.48×10^{-7}	0.45	0.000	0.402
MÉDIO PARAÍBA	Barra Mansa	2.48×10^{-7}	2.48×10^{-7}	0.06	0.005	0.452
	Fazenda Agulhas Negras	2.48×10^{-7}	2.48×10^{-7}	0.01	0.001	0.431
	Ponte do Souza	2.48×10^{-7}	2.48×10^{-7}	0.13	0.000	0.437
	Ribeirão de São Joaquim	2.48×10^{-7}	2.48×10^{-7}	0.01	0.001	0.361
	Visconde de Mauá	2.48×10^{-7}	2.48×10^{-7}	0.49	0.000	0.352
	Manuel Duarte	2.48×10^{-7}	2.48×10^{-7}	0.03	0.001	0.326
	Santa Isabel do Rio Preto	2.48×10^{-7}	2.48×10^{-7}	0.00	0.001	0.326

mean, at *Itaperuna* (7.14 mm-11.57 mm) in G_1, *Petrópolis* (13.83 mm-19.69 mm) and *Fazenda Agulhas Negras* (8.92 mm-14.51 mm) in G_2. Comparatively, *Macabuzinho* (5.78 mm-8.31 mm), *Três Irmãos* (5.87 mm-8.91 mm) and *Paraíba do Sul* (5.23 mm-7.94 mm) stations, inserted in the group G_1 and presented the smallest variations in the sample standard deviation and IQA. However, it was observed that the *Petrópolis* station (group G_2), despite to meet the assessed premises, has a mean higher than the median, followed by higher maximum values, total amplitude (65.85 mm) and lower (-12.72 mm) and upper (66.05 mm) limits (Table 7). Thus, the distribution form of the monthly rainfall time series for *Petrópolis* station is biased to the right, and the same is influenced by high rainfall values (Figure 7).

Table 8. Summary of descriptive statistics of processed data from monthly rainfall time series (mm) by Box Cox Processing for the regions of the State of Rio de Janeiro (SRJ).

Homogeneous Groups	Stations	Mean (mm)	Median (mm)	Value		Limit	
				Maximum (mm)	Total Amplitude (mm)	Lower (mm)	Upper (mm)
G_1	Dois Rios	11.76	11.50	28.54	28.54	-7.26	30.46
G_1	Macabuzinho	11.94	11.69	29.52	29.52	-4.69	28.57
G_1	Itaperuna	12.32	12.26	31.48	31.48	-10.87	35.41
G_1	Três Irmãos	10.73	10.56	25.92	25.92	-7.21	28.41
G_1	Bom Jardim	12.60	12.22	31.78	31.78	-9.07	34.37
G_1	Paraíba do Sul	9.14	9.13	24.10	24.10	-6.62	25.14
G_2	Petrópolis	26.82	25.92	65.85	65.85	-12.72	66.05
G_2	Fazenda Agulhas Negras	17.90	17.68	40.99	40.99	-11.18	46.87

Homogeneous Groups	Stations	Coefficient			Sample standard deviation (mm)	Quartile		Interquartile Amplitude (mm)
		Sample variation (%)	Asymmetry	Kurtosis (K)		Lower (mm)	Upper (mm)	
G_1	Dois Rios	55.70	0.22	-0.56	6.55	6.88	16.32	9.43
G_1	Macabuzinho	48.37	0.22	-0.27	5.78	7.78	16.10	8.31
G_1	Itaperuna	57.78	0.13	-0.82	7.14	6.48	18.06	11.57
G_1	Três Irmãos	34.50	0.15	-0.77	5.87	6.15	15.06	8.91
G_1	Bom Jardim	55.42	0.16	-0.78	6.98	7.22	18.08	10.86
G_1	Paraíba do Sul	57.27	0.03	-0.74	5.23	5.29	13.23	7.94
G_2	Petrópolis	51.52	0.19	-0.52	13.83	16.82	36.51	19.69
G_2	Fazenda Agulhas Negras	49.79	0.08	-0.92	8.92	10.59	25.10	14.51

Rainfall (mm)	Homogeneous Groups	
	G_1	G_2
Accumulated annual	1164.95	347.86
Percentage (%) in the homogeneous groups	77.01	22.99

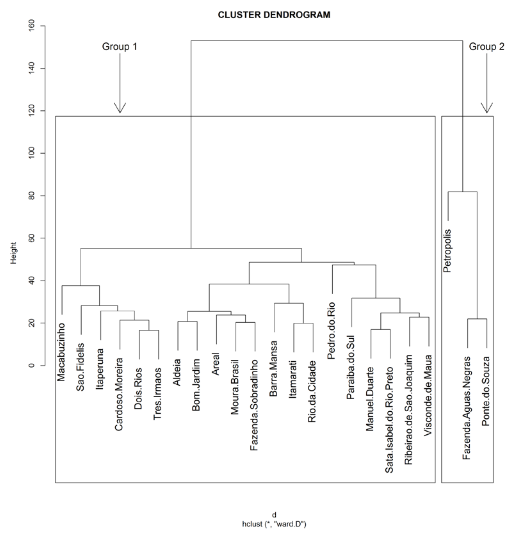


Figure 6. Dendrogram of cluster analysis of monthly rainfall time series of processed data for Norte Fluminense, Noroeste Fluminense, Serrana, Centro Sul Fluminense and Médio Paraíba regions of the State of Rio de Janeiro (SRJ).

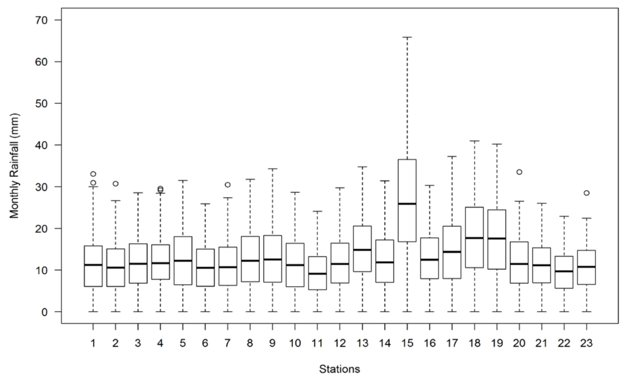


Figure 7. Boxplot of the monthly rainfall time series of the data processed by Box Cox for Norte Fluminense, Noroeste Fluminense, Serrana, Centro Sul Fluminense and Médio Paraíba regions of the State of Rio de Janeiro (SRJ).

Table 7. Normality and homogeneity of variance test for the raw and processed rainfall data of the State of Rio de Janeiro (SRJ).

Government regions	Stations	Raw data		Processed data		lambda (λ)
		Shapiro	Bartllet	Shapiro	Bartllet	
		P-value	P-value	P-value	P-value	
NORTE	São Fidelis	2.48×10^{-7}	2.48×10^{-7}	0.01	0.189	0.425
	Cardoso Moreira	2.48×10^{-7}	2.48×10^{-7}	0.03	0.010	0.417
	Dois Rios	2.48×10^{-7}	2.48×10^{-7}	0.05	0.159	0.433
	Macabuzinho	2.48×10^{-7}	2.48×10^{-7}	0.06	0.092	0.401
NOROESTE	Itaperuna	2.48×10^{-7}	2.48×10^{-7}	0.37	0.233	0.432
	Três Irmãos	8.99×10^{-7}	2.48×10^{-7}	0.11	0.065	0.393
CENTRO SUL	Areal	2.48×10^{-7}	2.48×10^{-7}	0.02	0.020	0.425
	Moura Brasil	2.48×10^{-7}	2.48×10^{-7}	0.05	0.000	0.409
	Paraíba do Sul	2.48×10^{-7}	2.48×10^{-7}	0.28	0.273	0.335
SERRANA	Aldeia	2.48×10^{-7}	2.48×10^{-7}	0.02	0.141	0.385
	Bom Jardim	2.48×10^{-7}	2.48×10^{-7}	0.14	0.322	0.409
	Fazenda Sobradinho	2.48×10^{-7}	2.48×10^{-7}	0.43	0.274	0.385
	Itamarati	1.22×10^{-6}	2.48×10^{-7}	0.82	0.001	0.443
	Pedro do Rio	2.48×10^{-7}	2.48×10^{-7}	0.06	0.001	0.411
	Petrópolis	6.55×10^{-6}	2.48×10^{-7}	0.00	0.001	0.565
	Rio da Cidade	2.48×10^{-7}	2.48×10^{-7}	0.45	0.000	0.402
MÉDIO PARAÍBA	Barra Mansa	2.48×10^{-7}	2.48×10^{-7}	0.06	0.005	0.452
	Fazenda Agulhas Negras	2.48×10^{-7}	2.48×10^{-7}	0.01	0.001	0.431
	Ponte do Souza	2.48×10^{-7}	2.48×10^{-7}	0.13	0.000	0.437
	Ribeirão de São Joaquim	2.48×10^{-7}	2.48×10^{-7}	0.01	0.001	0.361
	Visconde de Mauá	2.48×10^{-7}	2.48×10^{-7}	0.49	0.000	0.352
	Manuel Duarte	2.48×10^{-7}	2.48×10^{-7}	0.03	0.001	0.326
	Santa Isabel do Rio Preto	2.48×10^{-7}	2.48×10^{-7}	0.00	0.001	0.326

4. Discussion

The stations belonging to *Centro-Sul Fluminense*, *Serrana* and *Médio Paraíba* regions are influenced by systems ranging from local to synoptic scale, such as valley/mountain breeze circulation, the local convection, MCC, FS, SACZ, SASH and AB [7,9,12]. Displacement of synoptic systems occurs in the side from Serra do Mar directed toward the interior of the state with North-Northeast direction (N/NE). *Médio Paraíba* region is influenced by the same systems, but the displacement occurs on the side of the *Mantiqueira* directed toward Southwest-Northwest portions (SW/NW) [7,33,34].

IOR analysis also confirmed the high level of variability of raw data regarding the presence of outliers in the studied stations, being *Bom Jardim* (152.60 mm) and *Fazenda Sobradinho* (148.82 mm) those that presented the highest values. *Serrana* and *Médio Paraíba* regions showed IQA higher than the highest rainfall means, and the highest and lowest standard deviations, followed by the *Norte*, *Nordeste* and *Centro-Sul Fluminense* regions, which also showed some stations with the highest and lowest standard deviations and IQA (Tables 2 to 6).

According to [35] and [36], during winter season, there is an increased frequency of FS passages in the Southeast Brazil (SEB). However, these systems move with greater speed and find more difficultly to reach the interior of the state. Thus, the rainfall regime is limited to coast and *Serrana* and *Médio Paraíba* regions, where is located part of the state's stations [9,12]. However, stations in different regions of the state differ in relation to systems and preferred displacement as, for example, stations located at *Serrana* and *Centro-Sul Fluminense* regions, which are influenced by valley/mountain breeze circulation, local convection, FS, SACZ, SASH and AB [4,7,11]. The stations located in the *Norte* and *Nordeste Fluminense* are influenced by the same local and synoptic scale systems and, except for maritime/terrestrial breeze circulation [7,16,34]. The displacement of the systems at *Serrana* and *Centro-Sul Fluminense* regions occurs in the direction North/Northeast (N/NE) to the side of the Serra do Mar directed to the continent, while the flow for the stations from *Norte* and *Nordeste Fluminense* occurs preferably from North/Northeast/Southeast (N/NE/SE) [7].

As previously mentioned, the stations from *Norte Fluminense, Centro-Sul Fluminense, Serrana* and *Médio Paraíba*. It are highly influenced by the proximity to the coastal environment, especially the maritime/terrestrial breeze circulation which intensifies the FS, most frequently in winter, carrying moisture from the Atlantic Ocean to the continent, in addition to IL and MCC (synoptic and mesoscale) [7,20]. Stations from *Itaperuna* and *Três Irmãos* in *Nordeste Fluminense* region are away from the coast. [37] claim that the SRJ as the autumn-winter period progresses, the state's coastline starts to present the highest rainfall values.

Studies carried out by [38] with 15 normality tests using simulations with Monte Carlo method, demonstrate that the power, ease of implementation and test choice depends on many factors, including the type of distribution under alternative hypothesis, the sample size and values critics. The authors affirm that the SW test, among the used tests, provided an overall indicative of non-normality on several alternatives, symmetrical or non-symmetrical, heavy or light tails, and at lastly, on all sample sizes used. [39] compared 33 normality tests and did not mention a single efficient test compared to others in all evaluated cases. They classified tests according common features in each group, and recommend the SW among the most powerful tests for asymmetric distributions, followed by distributions which are mixtures of normal, or normal with presence of outliers when the kind of the non-normality is not known a priori. According to [40], in some situations parametric tests may disagree with the statistical decision by presenting a high sensitivity to normality violation. One of the classic examples of this sensitivity to violation is the Bartlett test for homogeneity of variance, which is seriously, affected by the non-normality, since it present a low power, and in many situations where data are not originated from a normal distribution, these tests reject the null hypothesis.

According to [7] the groups (G_1 and G_2) existing in the SRJ are formed due not only by the influence of the main synoptic systems that act in the SEB region, but mainly because of the proximity to the coastal environment and the complex topography. These local features induce to formation of orographic and convective rainfall and the rainfall regime produced by valley/mountain, lake/bay and maritime/terrestrial breeze circulation, in addition to intensify some systems, for example, IL, FS and MCC. This is due to the low frequency of outlier occurrence, or extreme values in the processed series [41].

5. Conclusions

Data from 3B43 product of the TRMM satellite together with the climatological normal from INMET can be used for gap filling and building of a time series and, especially in the study of rainfall pattern in the State of Rio de Janeiro. The descriptive analysis on the data consistency of the monthly rainfall time series data (raw) in the regions Norte Fluminense, Noroeste Fluminense, Centro-Sul Fluminense, Serrana and Médio Paraíba of the state of Rio de Janeiro revealed that the data with a probability of occurrence above 75% show high temporal variability. Based on cluster analysis, we defined two rainfall homogeneous groups representing rainfall in the five regions of SRJ Government, respectively. Shapiro-Wilk and Bartllet tests applied to the raw data show that the data do not exhibit normality and homogeneity of variance. Box Cox processing is effective for stabilizing the homogeneity of variances and normality of the residues from monthly rainfall time series in the five regions of the State of Rio de Janeiro. Applicating the statistics tools and parametric tests are efficient in terms of data consistency from monthly rainfall time series, and on understanding of the rainfall patterns in the State of Rio de Janeiro.

Author Contributions

The first author wrote the article and made the figures. The second author reviewed the manuscript and contributed to the writing and discussion of results and conclusions.

Conflict of Interest

None.

Acknowledgments

The authors acknowledge the Agência Nacional de Águas (ANA), Companhia de Pesquisa de Recursos Minerais (CPRM), Instituto Nacional de Meteorologia (INMET), SERLA Fundação Superintendência Estadual de Rios e Lagoas and Light Serviços de Eletricidade S/A by gently give in the data for composing the rainfall time series, and the CAPES for granting Doctorate scholarship. The second author thanks Brazilian National Council for Scientific and Technological Development (CNPq) for granting the Research Productivity Fellowship level 2 (309681/2019-7).

References

[1] IPCC - Intergovernmental Panel on Climate Change (2007). Climate Change 2007: The Physical Science Basis. Summary for Policymakers Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 18 pp.

[2] Tammets, T., Jaagus J. Climatology of precipitation extremes in Estonia using the method of moving precipitation totals. Theoretical and Applied Climatology, 111, 3-4, 623-639, 2013. https://doi.org/10.1007/s00704-012-0691-1.

[3] Nadarajah, S., Choi, D. Maximum daily rainfall in South Korea. Journal of Earth System Science, 116, 4, 311-320, 2007. https://doi.org/10.1007/s12040-007-0028-0.

[4] Oliveira-Júnior, J.F., Delgado, R.C., Gois, G., Lannes, A., Dias, F.O., Souza, J.C.S., Souza, M. Análise da Precipitação e sua Relação com Sistemas Meteorológicos em Seropédica, Rio de Janeiro. Floresta e Ambiente, 21, 2, 140-149, 2014. 10.4322/floram.2014.030.

[5] Westra, S., Alexaqnder, L.V., Zwiers, F.W. Global increasing trends in annual maximum daily precipitation. Journal of Climate, 26, 11, 3903-3918, 2013. https://doi.org/10.1175/JCLI-D-12-00502.1.

[6] Reboita, M.S., Gan, M. A, Rocha, R.P., Ambrizzi, T. Regimes de Precipitação na América do Sul: Uma Revisão Bibliográfica. Revista Brasileira de Meteorologia, 25, 2, 185-204, 2010. https://doi.org/10.1590/S0102-77862010000200004.

[7] Brito, T.T., Oliveira-Júnior, J.F., Lyra, G.B., Gois, G., Zeri, M. Multivariate analysis applied to monthly rainfall over Rio de Janeiro state, Brazil. Meteorology and Atmospheric Physics, 129, 2, 469-478, 2016. https://doi.org/10.1007/s00703-016-0481-x.

[8] Dereczynski, C.P., Oliveira, J.S., Machado, C.O. Climatologia da Precipitação no Município do Rio de Janeiro. Revista Brasileira de Meteorologia, 24, 1, 24-38, 2009. https://doi.org/10.1590/S0102-77862009000100003.

[9] André, R.G.B., Marques, V.S., Pinheiro, F.M.A., Ferraudo, A.S. Identificação de regiões pluviometricamente homogêneas no estado do Rio de Janeiro, utilizando-se valores mensais. Revista Brasileira de Meteorologia, 23, 501-509, 2008. https://doi.org/10.1590/S0102-77862008000400009.

[10] Satyamurty, P., Mattos, L.F., Nobre, C.A., Silva Dias, P.L. Tropics - South America. In: Meteorology of the Southern Hemisphere. Ed. Kauly, D. J. and Vincent, D. G., Meteorological Monograph. American Meteorological Society, Boston, 119-139, 1998.

[11] Cataldi, M., Assad, L.P.F., Torres Júnior, A.R, Alves, J.L.D. Estudo da influência das anomalias da TSM do Atlântico Sul extratropical na região da Confluência Brasil-Malvinas no regime hidrometeorológico de verão do Sul e Sudeste do Brasil. Revista Brasileira de Meteorologia, 25, 513-524, 2010. https://doi.org/10.1590/S0102-77862010000400010.

[12] Lima, A.O., Lyra, G.B., Moraes, M., Oliveira-Júnior, J.F., Zeri, M., Cunha-Zeri, G. Extreme rainfall events over Rio de Janeiro State, Brazil: characterization using probability distribution functions and clustering analysis. Atmospheric Research, 247,

1-105221, 2021. https://doi.org/10.1016/j.atmosres.2020.105221.

[13] Cruz, E., Carvalho, D.F., Ceddia, M.B., Antunes, M.A.H., Aquino, R.M. Ocorrência de Veranicos no Estado do Rio de Janeiro. Revista de Engenharia Agrícola de Jaboticabal, 24, 1, 68-79, 2004. https://doi.org/10.1590/S0100-69162004000100009.

[14] Moraes, N.O., Pimentel, L.C.G., Marton, E. Simulações Numéricas da Formação de Ilha de Calor na Região Metropolitana do Rio de Janeiro. Anuário do Instituto de Geociências (Rio de Janeiro), 28, 2, 116-138, 2005. https://doi.org/10.11137/2005_2_116-138.

[15] Zeri, M., Oliveira-Júnior, J.F., Lyra, G.B. Spatiotemporal analysis of particulate matter, sulfur dioxide and carbon monoxide concentrations over the city of Rio de Janeiro, Brazil. Meteorology and Atmospheric Physics, 113: 139-152, 2011. https://doi.org/10.1007/s00703-011-0153-9.

[16] Oliveira-Júnior, J.F., Gois, G, Terassi, P.M.B., Silva Junior, C.A., Blanco, C.J.C., Sobral, B.S., Gasparini, K.A.C. Drought severity based on the SPI index and its relation to the ENSO and PDO climatic variability modes in the regions North and Northwest of the State of Rio de Janeiro - Brazil. Atmospheric Research, 212, 91-105, 2018. https://doi.org/10.1016/j.atmosres.2018.04.022.

[17] Serra, A.B., Ratisbonna, L. Clima do Rio de Janeiro. Ministério da Agricultura. Serviço de meteorologia, 2ª edição. Boletim Geográfico, Fundação Instituto Brasileiro de Geografia (IBGE), nº 214, p. 135, 1957.

[18] Strang, D. Considerações sobre as chuvas de janeiro de 1962 no Estado da Guanabara. Escritório de Meteorologia, 16p, 1962.

[19] Serra, A. B. Clima da Guanabara. Boletim Geográfico, Fundação Instituto Brasileiro de Geografia (IBGE), nº 214, p. 35, 1970.

[20] Machado, R.L., Ceddia, M.B., Carvalho, D.F., Cruz, E. S., Francelino, M.R. Spatial variability of maximum annual daily rain under different return periods at the Rio de Janeiro state, Brazil. Bragantia, 69, suppl., 77-84, 2010. https://doi.org/10.1590/S0006-87052010000500009.

[21] IBGE-Instituto Brasileiro de Geografia e Estatística. IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios Contínua 2014. Available in: < http://www.ibge.gov.br/estadosat/perfil.php?sigla=rj>. Accessed in 20 jul. 2015.

[22] Costa, C.D. Variabilidade Intranual da Precipitação Pluvial Mensal no Estado do Rio de Janeiro. Monografia (Graduação em Engenharia Florestal). Curso de Engenharia de Florestal, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 42p., 2010.

[23] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. 2005.

[24] Schlotzhaver, S.D., Littell R.C. SAS System for elementary statistical analysis. 2ª ed. Cary, NC: SAS Institute Inc., p.456, 1999.

[25] Bartlett, M.S. Properties of sufficiency and statistical tests. Processings of the Royal Society, Series A, 160: 268–282, 1937. jstor.org/stable/96803.

[26] Snedecor, G., Cochran, W., William, G. Statistical Methods (8th Ed.). Ames, Iowa: Blackwell Publishing Professional. ISBN 978-0-8138-1561-9, 1989. Retrieved 2011-08-05.

[27] Box, G.E.P., Cox D. R. An analysis of processings. Journal of the Royal Statistical Society, Series B, 26: 211-252, 1964.

[28] Shapiro, S.S., Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). Biometrika, 52, 1, 591-611, 1965.

[29] Ward, J.H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, 1, 236-244, 1963. https://doi.org/10.1080/01621459.1963.10500845.

[30] Everitt, B.S., Dunn, G. Applied multivariate analysis. London: Edward Arnold, 400p, 1991.

[31] Lyra, G.B., Oliveira-Júnior, J.F., Zeri, M. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. International Journal of Climatology, 34, 13, 3546-3558, 2014. https://doi.org/10.1002/joc.3926.

[32] Kubrusly, L.S. Um procedimento para calcular índices a partir de uma base de dados multivariados. Pesquisa Operacional, 21, 1, 107-117, 2001. https://doi.org/10.1590/S0101-74382001000100007.

[33] Atlas Eólico do Rio de Janeiro. Governo do Estado do Rio de Janeiro, Secretaria de Estado de Energia, da Indústria Naval e do Petróleo, 2002.

[34] Correia Filho, W.L.F., Souza, P.H.A., Oliveira-Júnior, J.F., Terassi, P.M.B., Gois, G., Silva Junior, C.A., Sobral, B.S., Rangel, R.H.O., Pimentel, L.C.G. Investigating the characteristics and predictability of measured wind speed data over Rio de Janeiro, Brazil. Pure And Applied Geophysics, 1, 1-23, 2021. https://doi.org/10.1007/s00024-021-02751-w.

[35] Lemos, C. F., Calbete, N. Sistemas Frontais que atuaram no litoral de 1987-1995. Climanalise Especial-Edição comemorativa de 10 anos. MCT/INPE/

CPTEC, 1996.

[36] Ferreira, N. J., Sanches, M., Silva Dias, M.A.F. Composição da Zona de Convergência do Atlântico Sul em Períodos de El Niño e La Niña. Revista Brasileira de Meteorologia, 19, 89-98, 2004.

[37] Mello, C.R., Viola, M.R, Curi, N., Silva, A.M. Distribuição Espacial da Precipitação e da Erosividade da Chuva Mensal e Anual no Estado do Espírito Santo. Revista Brasileira de Ciência do Solo, 36, 6, 1878-1891, 2012. https://doi.org/10.1590/S0100-06832012000600022.

[38] Yazici, B., Yolacan, S. A comparison of various tests of normality. Journal of Statistical Computation and Simulation, 77, 2, 175-183, 2007. https://doi.org/10.1080/10629360600678310.

[39] Romão, X., Delgado, R., Costa, A. An empirical power comparison of univariate goodness-of-fit of normality. Journal of Statistical Computation and Simulation, 80, 5, 545-591, 2010. https://doi.org/10.1080/00949650902740824.

[40] Torman, V.B.L., Coster, R., Riboldi, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. Revista Hospital das Clínicas Paraná, 32, 2, 227-234, 2012.

[41] Pereira, M.A.F., Kobiyama, M. Análise de variâncias pluviométricas na região da bacia hidrográfica do Cubatão do Sul (SC). Ambiência (Online), 9, 1, 95-111, 2013. https://doi.org/10.5777/ambiencia.2013.01.07.