ECOSYSTEMS

Spatial and Seasonal Assessment of Water Quality in the Lobo Stream River Basin, Brazil Using Multivariate Statistical Techniques

GABRIELA L. NEVES, TAINÁ T. GUIMARÃES, PHELIPE S. ANJINHO, MARIANA A.G.A. BARBOSA, ALLITA R. DOS SANTOS, JORIM S. VIRGENS FILHO & FREDERICO F. MAUAD

Abstract: In view of the complexity of the processes associated to water quality, this study objective is to identify the main factors related to spatial and seasonal variability in water courses of the Lobo Stream River Basin. To this, multivariate statistical techniques were used. Data collection for water quality and streamflow variables were carried out monthly, from May 2018 to April 2019, at 10 monitoring points along basin's tributaries. The results show that, during the dry season, the main causes for water quality decrease are related to erosion process on the river margins, which is intensified by inadequate handling in livestock activities in some monitoring points. In the rainy season, the main causes are related to soil leaching in agricultural areas that increases the nitrogen compounds concentration and reduces water quality. However, in addition to this, it was noted that regardless the environmental conditions, the most impactful factor is the point pollution from the effluent discharge of Itirapina City sewage treatment plant, responsible for nutrient concentration increase, organic contamination, OD reduction, and, consequently, water quality deterioration. With this, the study shows how multivariate statistical analysis enables more relevant evaluation of water quality data variability and supports further studies in the basin.

Key words: anthropic activities, cluster analysis, correlation matrix, hydrological variables, principal component analysis, water pollution.

INTRODUCTION

The maintenance of social, economic and ecological systems depends on the availability of natural resources, especially water, which is essential for the existence of life. Water must be available in sufficient quantity and in adequate quality to meet its multiple uses and maintain the proper functioning of natural ecosystems and associated ecosystem services (Grizzetti et al. 2016, MEA 2005, Pantano et al. 2016).

However, water availability and quality are being threatened due to human activities that generate multiple pressures, such as the dumping of domestic waste, hydromorphological changes, diffuse pollution of agricultural areas and drainage of urban areas that can pollute water bodies and change their ecological status (Grizzetti et al. 2017, Zhao et al. 2018, Mello et al. 2020). The increase in population, human activities, and demand for water, as well as inappropriate land use and climate change, are the main factors responsible for changes in water resources, which can compromise aquatic ecosystems and socioeconomic sectors that depend on services ecosystems provided by them (Tundisi et al. 2015, Bai et al. 2019, Pham et al. 2019).

The influence of environmental and hydrological variables on water quality has been

the subject of several studies, since it support water quality modeling and contribute to the management of river basins (Tilburg et al. 2015, Zhou et al. 2017). The rainfall conditions and the type of land use can affect the flow regime of rivers as well as the transport and dilution of pollutants, which can affect water quality.

The rainfall, and consequently flow, can affect the water quality and the interference can be examined on two aspects. The first is that the higher the flow, the greater the dilution capacity of water bodies. However, in periods of greater precipitation, pollutants are carried to water bodies by percolation of water through soil and by soil leaching, thus increasing the amount of contaminating substances in water bodies and reducing water quality. However, in dry periods, nutrients that can be carried to water bodies remain stored in the soil and are not carried away by runoff. Thus, increased rainfall can reduce water quality, when the effect of dilution is not significant (Menció et al. 2011).

In view of the complexity and the large number of variables associated with water quality, multivariate statistical techniques appear as important tools in monitoring and management of water resources. Techniques such as principal component analysis (PCA), factor analysis (FA) and cluster analysis (CA) have been used in several studies with the aim of reducing the number of data and identifying the main sources or factors that explain the temporal variations and spatial aspects of water quality in river basins (Barakat et al. 2016, Gulgundi & Shetty 2018, Li et al. 2018, Réhman et al. 2018). Study developed by developed by Rocha et al. (2014) for example, proposed a water quality index for the Óros reservoir, located in northeastern Brazil, based on PCA and other statistical techniques. The authors reduced the number of variables from 33 to 12, identifying those most significant for the management of the reservoir's water quality.

The Lobo Stream River Basin (LSRB) has been presenting an increase in nutrients concentration and a reduction in water quality in its reservoir and in its main water courses due to inadequate use of the soil in the region, agricultural activities, increase in tourism and discharge of domestic sewage (Moruzzi et al. 2012, Frascareli et al. 2018, Anjinho et al. 2020, Mizael et al. 2020). Therefore, due to the importance of LSRB for the central-eastern region of the São Paulo State, this study aimed to identify, using multivariate statistical techniques, the main factors related to the spatial and temporal variability of water quality in LSRB.

MATERIALS AND METHODS

Study area

The LSRB, object of study, is located in the central-eastern region of the São Paulo State, between the municipalities of Brotas, Itirapina and São Carlos and has total area of 227.7 km². The hydrographic network is basically formed by four main rivers: Lobo stream, Itaqueri river (with affluents Água Branca stream and Limoeiro stream), Geraldo stream and Perdizes stream (Figure 1), which drain their water towards the Lobo reservoir.

The climate in LSRB, according to the Köppen classification, is Cwa, which is characterized by humid subtropical climate, with dry winter and hot summer. The average annual accumulated precipitation varies from 1300 to 1500 mm, with greater frequency and volume occurring during the summer season (Alvares et al. 2013).

The main land use activities in the basin are reforestation, natural vegetation, pastures, agriculture, urban areas, low-density human occupation, sand mining, industry and water bodies (Anjinho et al. 2021).

Data base

For data collection, monthly measurements and samplings were carried out, from May 2018 to
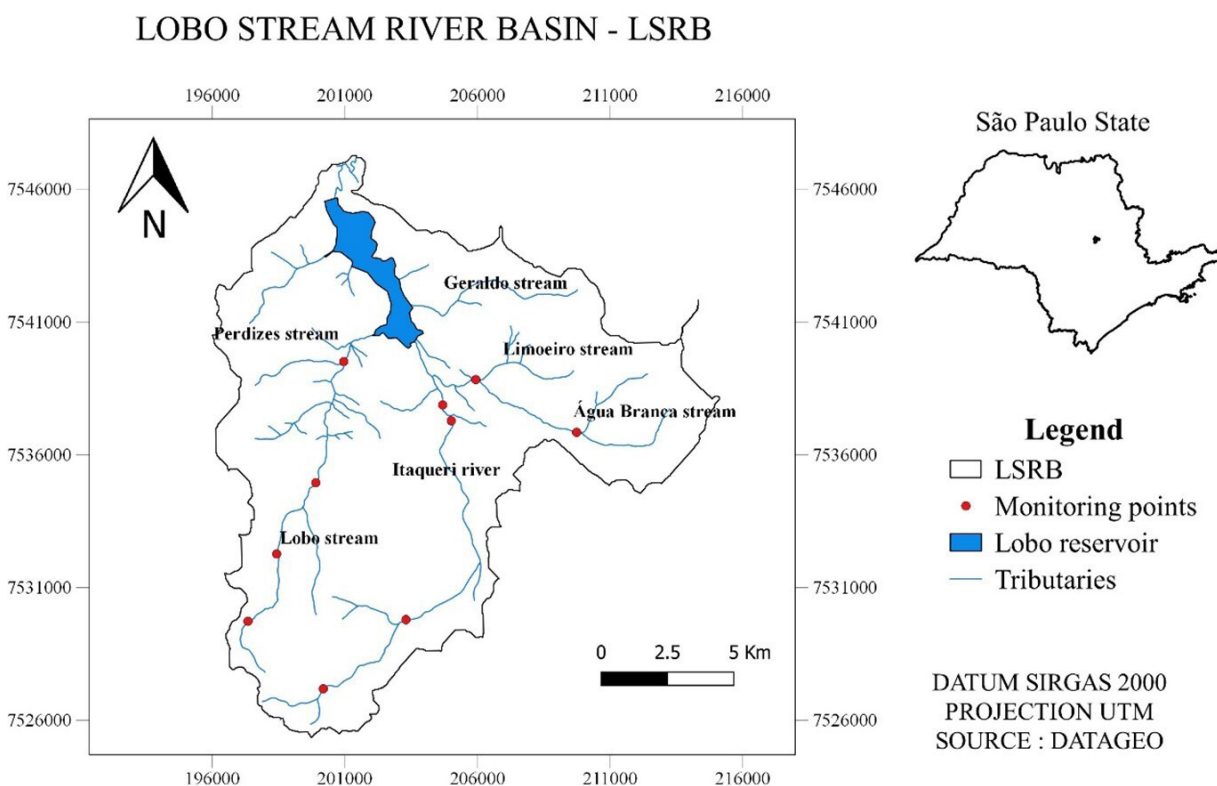
Figure 1. Location and hydrography of LSRB and monitoring points.

April 2019, at 10 monitoring points along LSRB tributaries: Lobo stream (Lb-1, Lb-2, Lb- 3 and Lb-4), Itaqueri river (It-1, It-2, It-3 and It-4), and Água Branca stream (Ab-1 and Ab-2) (Figure 1). The choice to sample these water courses is because they are the ones that most contribute to the transport of pollutants to the Lobo reservoir, according to the literature (Moruzzi et al. 2012, Anjinho et al. 2020).

The points were selected in order to identify the spatial variability of water conditions and land use activities, from the source to the outflow of the water courses into the reservoir. Table I presents a brief description of the monitoring points.

The water quality data, pH, turbidity, electrical conductivity and dissolved oxygen were obtained from in situ measurements using the YSI 6820 Multiparameter Sonde. To obtain the other parameters, surface water samples were collected, which were taken to the laboratory of

the Hydrometry Center at the Center for Water Resources and Environmental Studies, University of São Paulo (CRHEA/USP), where the physical-chemical analysis were performed (Table II).

LSRB rivers are classified, according to CONAMA Resolution 357/2005, as class 2 water bodies, that is, water resources destined for domestic supply after conventional treatment, protection of aquatic communities, recreation of primary contact, irrigation of vegetables, fruits and public areas, aquaculture and fishing activities (Brazil 2005). Therefore, the parameters obtained at the monitoring points were compared to the quality standards indicated by the CONAMA Resolution for class 2 rivers and by the CETESB report- Environmental Company of the State of São Paulo (CETESB 2019).

In order to detect the influence of rainfall conditions on water quality parameters, the accumulated precipitation values for the period of 15 days prior to the dates of data collection

Table I. Description of monitoring points at LSRB tributaries.

Point	Description
Ab-1 **Água Branca stream**	Located upstream of the Sewage Treatment Plant (STP) in the city of Itirapina. Presence of urban areas on the left bank and forestry on the right bank.
Ab-2 **Água Branca stream**	Located downstream from the STP in the city of Itirapina. Presence of riparian forest on the river banks, and silviculture activities and fragments of cerrado near the stretch.
It-1 **Itaqueri river**	Closest point to the source of the Itaqueri river. It presents changes in the river channel and livestock activities in the surroundings. There is no presence of riparian vegetation in this stretch.
It-2 **Itaqueri river**	Located in rural area. It presents riparian forest in the stretch, in addition to forestry and livestock activities nearby.
It-3 **Itaqueri river**	Located in the floodplain area of the Itaqueri river. It has pasture and forestry areas in its surroundings.
It-4 **Itaqueri river**	Closest point to the Lobo reservoir, downstream from the sand mining company and upstream from STP Itirapina.
Lb-1 **Lobo stream**	Closest point to the source of the Lobo stream. Presence of riparian vegetation on the left bank and livestock activities on the right bank.
Lb-2 **Lobo stream**	Located in rural area. Degraded banks due to soil erosion. Presence of livestock activities in the surroundings.
Lb-3 **Lobo stream**	Presence of dense riparian forest around the point.
Lb-4 **Lobo stream**	Closest point to the Lobo reservoir, located inside the Ecological Station of Itirapina.

Source: Adapted from (Anjinho et al. 2020).

were obtained. Daily data were obtained from the Climatological Station of the CRHEA / USP.

For data analysis, it was considered that the precipitation regime in the study area is divided in a rainy season, from October to March, and a dry season, corresponding the months from April to September.

To determine the flow (Q) at each monitoring point, the width between the margins and the depth was measured, thus obtaining the area of the river's cross section. Using a current meter, the water velocity was also obtained. The Q was determined by the product of the wet section area and the average flow velocity of the current.

Statistical analysis

In order to analyze the data sets of the variables and investigate the possible sources of pollution in LSRB, descriptive statistical analysis, Spearman correlation analysis and multivariate statistical analysis were used, which includes principal component analysis (PCA), factor analysis (FA) and cluster analysis (CA), performed in the R software environment.

For distribution assessment of data collections, the descriptive analysis represented in boxplot graphics was used. This methodology allows the visualization of the first quartile (lower limit) that represents the 25% percentile

Table II. Methodologies for physical-chemical parameters determination.

Parameters	Unit	Abbreviation	Methodology
pH	-	pH	Sonde YSI 6820
Temperature	°C	Temp	Sonde YSI 6820
Turbidity	NTU	Turb	Sonde YSI 6820
Electrical conductivity	µS/cm	EC	Sonde YSI 6820
Dissolved oxygen	mg/L	DO	Sonde YSI 6820
Biochemical oxygen demand	mg/L	BOD_5	Incubation Method (20°C, five days)
Total phosphorus	mg/L	TP	Method 4500 B – Persulphate Digestion
Orthophosphate	µg/L	PO_4	Standard Methods – 1985
Nitrate	mg/L	NO_3	Method $4500\text{-}NO_3$ – B (APHA, 2005)
Nitrite	mg/L	NO_2	Mackereth, Heron & Talling, 1978
Chlorophyll	µg/L	Cl-a	Nush et al., 1980
Total suspended solids	mg/L	TSS	Gravimetric method
Fixed suspended solids	mg/L	FSS	Gravimetric method
Volatile suspended solids	mg/L	VSS	Gravimetric method

of the observed data, the third quartile (upper limit) that represents the 75% percentile of the data, as well as the median, average, maximum, minimum, and outliers.

The observation of outliers in monitoring water quality is of importance, since the anomalous values, extremely high or extremely low, which differ numerically from the rest of the data, can contain significant information and indicate possible sources of pollution that must be analyzed (Muñiz et al. 2012).

In order to verify the possibility of applying PCA to the data of this study, Bartlett's test of sphericity, Kaiser-Meyer-Olkin criterion (KMO) and an analysis of the correlation matrix were previously performed.

The Kaiser-Meyer-Olkin (KMO) criterion assesses the correlation of the matrix by comparing simple correlations with partial correlations between variables. This statistic can vary from 0 to 1, with the value closest to "0" indicating that PCA is not adequate (weak correlation) and the value closest to "1" indicates that the use of the analysis is adequate. Values above 0.5 are considered acceptable for the application of PCA (Finkler et al. 2015, Chen et al. 2018).

Bartlett's test of sphericity aims, at a significance level of 0.05, to reject the null hypothesis that says that the correlation matrix is an identity matrix, which would indicate that there is no correlation between the variables and the analysis should not be performed (Chen et al. 2018).

After performing the KMO criterion and Bartlett's test and previously analyzing the correlation matrix, PCA was applied to two data sets (dry and rainy seasons) considering all collections from all monitoring points. The technique converts the original interrelated

variables into a smaller set of new independent and uncorrelated variables, called principal axes or components, which explain the data variation (Hongyu et al. 2015). Thus, the analysis allows to reduce the data set and can identify possible sources of water pollution (Zhao et al. 2011).

The application of PCA and FA occurs in five main stages: (1) standardization of values, for data to have zero mean and unit variation to ensure that all have the same weight in the analysis; (2) calculation of the correlation or variance matrix; (3) determination of eigenvalues; (4) disposal of components with a small proportion of the variation and (5) creation of the factor loading matrix (Finkler et al. 2015).

The CA was used for the purpose of detecting spatial and temporal similarity and separating the groups according to the similarity of water quality between the data collected at the monitoring points. The technique allows to classify similar variables in groups, which are represented by a dendrogram, which shows the proximity or distance between the groups (Hossain et al. 2015). In this study, CA was performed using the Euclidean distance method as a measure of similarity (Gulgundi & Shetty 2018), using the agglomerative hierarchical

method, in which, initially, each observation is considered individually as a group, then pairs are formed according to their similarities, repeating the procedure in new larger groups, reaching a single large group with all samples (Trebuňa & Halčinová 2013, Saxena et al. 2017). In this study, the WARD minimum variance technique was used, which aims to seek the smallest variance within the groups, which is measured by the sum of squares and then unite the groups according to homogeneity (Santos et al. 2017). This method is recommended and has been used in several studies (Carvalho et al. 2017, Pena et al. 2017, Crispim et al. 2020).

RESULTS AND DISCUSSION

Precipitation

Figure 2 shows that, considering the dry months, April had the highest accumulated precipitation, characterizing the transition from rainy to dry season. Among the rainy months, October stands out as the second month with the highest accumulated precipitation.
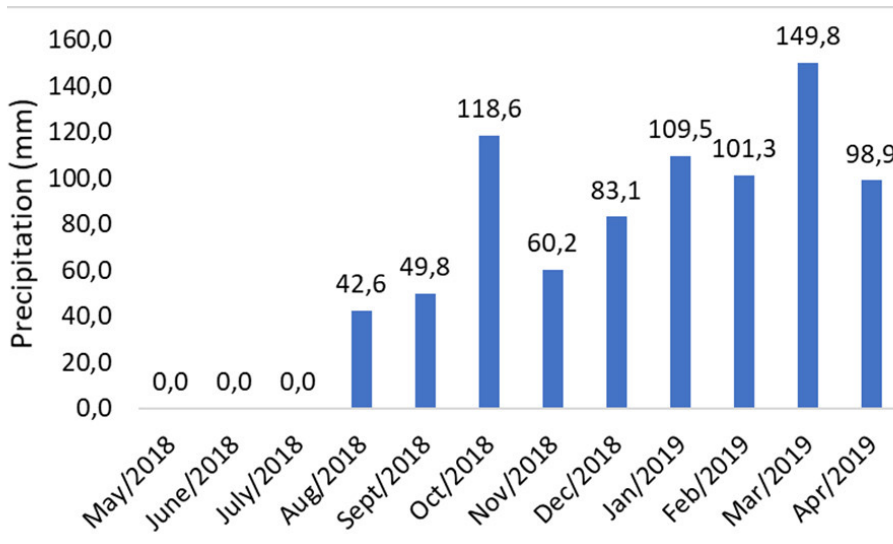

Figure 2. Accumulated precipitation for the period of 15 days before data collection.

General analysis of quality parameters and compatibility with legislation

Figure 3 presents the results of the descriptive analysis of water quality parameters, represented by boxplot graphics. The horizontal lines represent the quartiles, the central horizontal bar represents the median value, the diamond represents the average value of the variable and the points above or below the central vertical lines are the outliers. The dotted lines represent the standards established by CONAMA Resolution 357/2005 for class 2 rivers. Differences in data behavior can be observed between the monitoring points.

The behavior of water temperature can be observed following the seasonal variation of air temperature, where the lowest values occurred in the coldest months and the highest values occurred in the warmest months. It was observed that the points Ab-1, Ab-2 and It-1 presented higher mean temperature values, which may indicate the presence of residues in the water
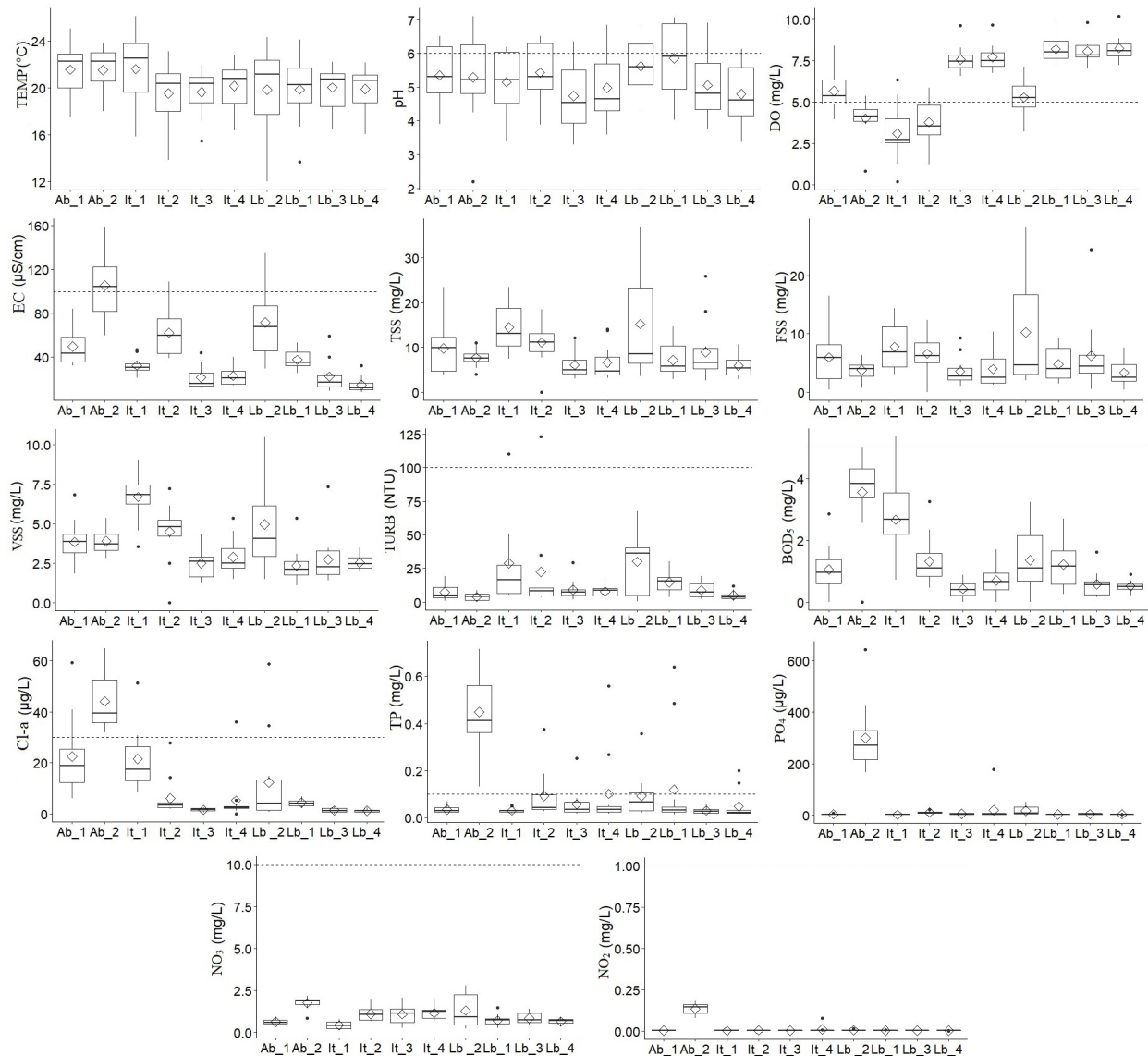

Figure 3. Boxplot graphs of the quality parameters obtained at the monitoring points.

body. The variation in water temperature depends on the climatic regime, geographic conditions and the period of the year, and is also influenced by the presence of effluents and the characteristics of the river channel (CETESB 2019).

All monitoring points showed average values below the ideal range for the pH variable (6.0 to 9.0) indicated by the resolution (Brazil 2005). It is also noted the presence of an outlier at point Ab- 2, a value of 2.20 that occurred during March. Low pH values may indicate the presence of domestic or industrial waste in the water body (Von Sperling 2005) or the presence of materials from riparian vegetation. However, it is also noteworthy that the low pH values found in all monitoring points may be associated with the acidity of LSRB soil, a fact that has already been verified in other studies (Anjinho et al. 2020).

Regarding DO concentrations, it was observed that at points Ab-2, It-1 and It-2, more than 80% of the observed values of DO were below the limit of 5 mg / L established by the resolution (Brazil 2005). Outlier values of 0.82 and 0.21 mg / L were also observed at points Ab-2 and It-1 respectively, which occurred in May.

The variation of DO in water is associated with climate variations, since the oxygen solubility depends on the temperature and also on the characteristics of the river channels, since the presence of rapids and waterfalls promote oxygenation due to water turbulence in these areas, thereby increasing the levels of dissolved oxygen in this environment (Nozaki et al. 2014). The low concentrations of DO at point It-1 may be consequence of a change in the river channel, where the water flow was interrupted by an artificially constructed barrier, making a lentic environment. This modification, together with livestock activity in the surroundings, may be responsible for the increase in organic matter, and the reduction of DO.

However, low concentrations of DO can also indicate pollution due to the presence of organic matter or other compounds in high concentrations that require the consumption of DO to stabilize them. This fact may be responsible for the reduction in DO at points Ab-2 and It-1 and It-2.

It is observed that most of the points presented average values of EC below the level indicated by CETESB (2019) of 100 μs / cm, with the exception of Ab-2. However, it is noted that the point Lb-2 in November and March and the point It-2 in March, presented values above the level indicated by CETESB, indicating the presence of salts and, consequently, water pollutants.

For the suspended solids variables, there is no limit established by CONAMA Resolution 357/2005, however, it was noted that the Lb-2 point showed the greatest temporal variability and the highest values among the monitoring points.

Outlier values were observed for different monitoring points in different months, showing that the variability and increase in suspended solids may be related to specific issues, such as the erosion process that occurs on the river banks and the intensified runoff in the rainy months together with improper land use around the points.

There is certain predominance of FSS compared to VSS in most observations, which indicates the presence of inorganic matter as the main portion among the material suspended in the water bodies.

It is observed that for the turbidity, the points presented average values below the maximum allowed by the CONAMA Resolution, of 100 NTU. However, there are extreme values of 110.1 NTU and 123.10 NTU at points It-1 and It-2, and these values were found in December and October, respectively. The increase in turbidity is

associated with an increase in the concentration of solids in water that occurs intensely during precipitation events.

The BOD$_5$ values, for all points, were below the level of 5 mg / L established by CONAMA Resolution, with the exception of the point It-1, which presented a value of 5.36 mg / L in March. It is noted that most of the points did not present high values of BOD$_5$, which means that they do not have high amount of organic matter that requires consumption of DO, unlike point It-1 which presents conditions of pollution by organic matter.

The average values of Cl-a, from most of the monitoring points, were below the limit of 30 μg / L established by the CONAMA Resolution, with the exception of point Ab-2, which presented an average of 44.13 μg / L, maximum value of 64.82 μg / L and all observations above the limit.

Extreme values of Cl-a can be observed at other points, above the indicated limit, 59.2 μg / L (Ab-1 in November), 51.30 μg / L (It-1 in October), 36.11 μg / L (It-4 in May), 34.63 μg / L and 58.70 μg / L (Lb-2 in June and July). The high observed values of Cl-a represent the high phytoplankton productivity that is stimulated by the high availability of nutrients, thus indicating the trophic level of the water.

In relation to total phosphorus, the points Ab-2, It-4 and Lb-1 showed an average above or equal the limit of 0.1 mg / L established by the CONAMA Resolution, being 0.45, 0.10 and 0.12 mg / L, respectively. Besides the average values, values above the limit are observed at point Ab-2 in all months of collection, with the maximum value being 0.72 mg / L and at points It-2 (0.19 mg / L in December, 0.38 mg / L in January and 0.14 mg / L in February), It-3 (0.25 mg / L in September), It-4 (0.56 mg / L in May and 0.27 mg / L in April), Lb-1 (0.64 mg / L in January and 0.49 mg / L in February), Lb-2 (0.13 mg / L in September, 0.14 mg / L in October and

0.36 mg / L in February), and Lb-4 (0.20 mg / L in December and 0.15 mg / L in January).

The dissolved inorganic phosphorus, or orthophosphate, showed a maximum value of 427.55 μg / L, in January at the point Ab-2, which presented an outlier value of 642.79 μg / L and the highest average among the monitoring points. Also noteworthy is the point It-4, which presented an extreme value of 177.90 μg / L.

The high concentration of phosphate compounds (total phosphorus and orthophosphate) is associated with discharges of domestic and industrial sewage with detergents and agricultural fertilizers. It is important to highlight that phosphorus is an essential nutrient for biological processes and in high concentrations it is responsible for the excessive growth of algae and for the water eutrophication process (Von Sperling 2005).

For most of the observations, the nitrate values were below the limit of 10 mg / L indicated by CONAMA Resolution, with the exception of points It-3 and It-4, which presented extreme values of 11.17 and 11.03 mg / L, respectively. However, it was noted that the Ab-2 point had the highest average among the points (1.98 mg / L), and that all monitoring points had outliers in March.

Nitrite had the highest concentration (0.185 mg / L) in December at the Ab-2 point, with this point having the highest average (0.135 mg / L). Outlier values were observed in May (0.07 at the It-4 point), in October (0.006 at Lb-1 and 0.013 at Lb-2). However, none of the monitoring points showed values above the limit of 1.00 mg / L established by the CONAMA Resolution.

The high concentration of nitrogen components is related to the release of domestic and industrial effluents, fertilizers used in agriculture, and animal excreta. Like phosphate compounds, nitrogen is also an indispensable nutrient for algae growth, and when they are

carried through runoff to water bodies, in high concentrations, they contribute to the eutrophication process. The presence of nitrogen in organic and ammoniac forms, representing the beginning of nitrogen oxidation, indicates the high oxygen consumption that will be used in the transformation of ammoniac nitrogen into nitrite and this into nitrate, which may indicate a degradation environment (Von Sperling 2005).

Correlation matrix

For the construction of the correlation matrix, the Royston test (Royston 1983) was applied, which is based on the Shapiro-Wilk statistics, to test the multivariate normality of the data. The test results for the two data sets (dry and rainy seasons) indicated that the data do not follow a normal distribution (α = 0.05). Thus, for the construction of the correlation matrix, the Spearman coefficient was used, which refers to a non-parametric correlation that measures the strength of the relationship between variables on a scale of -1 to +1.

The correlation between the variables obtained from data collection at all monitoring points, using Spearman's coefficients, is shown in Figure 4a-b. The significant positive correlations are indicated in blue color and the negative in red color. The white color represents statistically non-significant correlations.

For the interpretation of the coefficients was considered: values below 0.1 (trivial); values between 0.1 and 0.3 (small); values between 0.3 and 0.5 (moderate), between 0.5 and 0.7 (large); from 0.7 to 0.9 (very large) and greater than 0.9 (almost perfect), these values being negative or positive.

The FSS variable was excluded from the analysis, as it presents self-correlation with the TSS variable, since they derive from the same analyzed parameter.

Among the significant negative relationships, it is observed that the pH presented moderate negative correlation with the variables Q, Temp and TSS in the dry season. In the rainy season, the pH showed moderate negative correlation with NO_3 and large negative correlation with Q. Similar results were observed by Alvarenga et al. (2016) and Ben-Eledo et al. (2017), who showed that the pH can be influenced by the
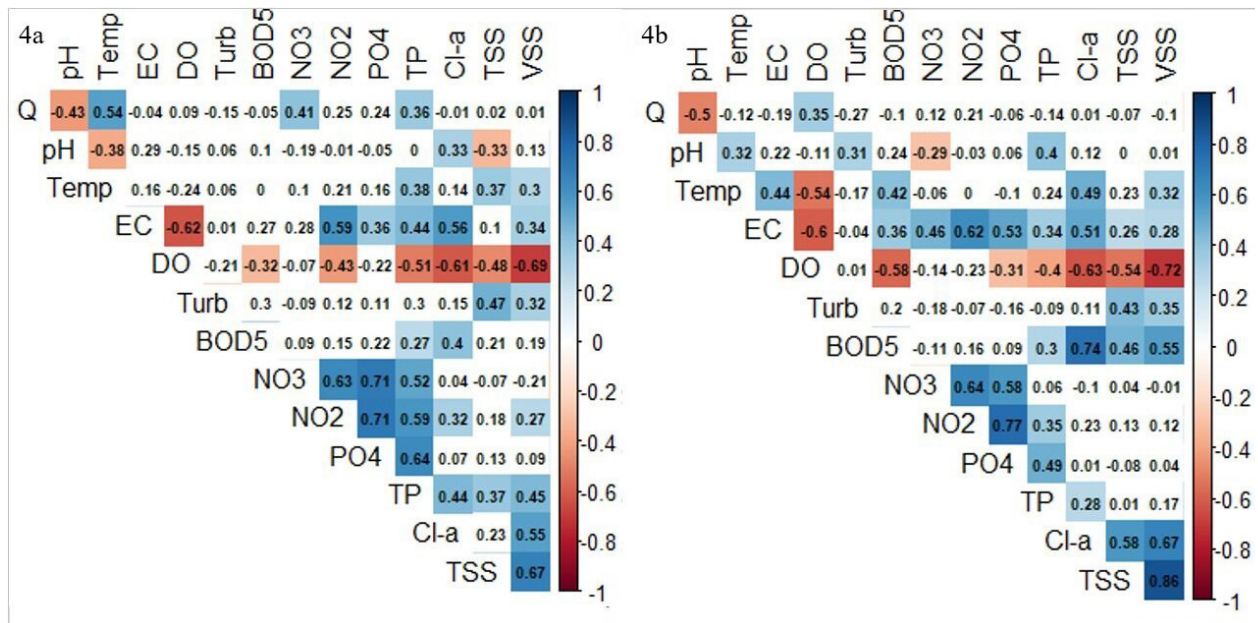

Figure 4. Correlation matrix by the Spearman method for the dry season (a) and the rainy season (b).

rainfall conditions, since in the period of greater precipitation and consequently higher Q, the pH is lower. Mendes & Ferreira (2014) found negative correlation between pH and precipitation and attributed the decrease in pH to the leaching of Cerrado biome soil that are naturally acidic, characteristic that is present in the study area and that can also explain the result found in LSRB.

Still among the significant negative relationships, it is important to note that in the dry season, there are moderate correlations between the DO and the variables DBO_5, NO_2 and TSS and large correlations with the variables EC, TP, Cl-a and VSS. In the rainy season, DO showed moderate negative correlation with TP and PO_4, large with the variables Temp, EC, BOD_5, Cl-a and TSS and very large with VSS. The negative correlations of DO with these variables in the two periods analyzed, indicate that the higher the DO, the lower the concentrations of the other parameters, since organic and inorganic compounds such as NO_2 tend to decrease the levels of DO during the organic matter degradation process and also during the nitrification process. However, it is noted that there is a greater relationship between DO and VSS, since there is high consumption of DO in the decomposition processes of organic compounds. Thus, it is emphasized that DO is a good indicator of water quality and, at low levels, represents a polluted environment, mainly due to high concentrations of organic matter.

Negative correlations of DO with different variables were observed by Matta et al. (2017) and Chen et al. (2018) in different periods. However, (Barakat et al. 2016) highlighted that the increase in temperature is responsible for the reduction of DO solubility in water, thus decreasing its concentration, a fact that explains the negative correlation of DO with Temp during the rainy season, corresponding to the period

with higher air temperatures and consequently higher water temperatures.

With the exception of DO, all other variables showed significant positive correlations with at least one other variable. In the dry season, there is strong positive correlation between the variables Q and Temp, result also found by found by Alvarenga et al. (2012) who pointed out the water temperature as the parameter most influenced by the seasonality of the flow. Also were found in this study large positive correlations between NO_2 and NO_3 / TP / EC, between NO_3 and TP, between TP and PO_4 and between Cl-a and VSS / EC. Also, in the dry season, very large positive correlations can be observed between PO_4 and NO_3 / NO_2.

Significant positive correlations between Cl-a and suspended solids have been found in other studies, where it was observed that the organic part of the solids, which corresponds to VSS, is more associated with Cl-a compared to the inorganic part. This relationship can be explained by the fact that in environments with certain level of eutrophication, the increase in solids concentration is a consequence of the growth of algae and some aquatic plants (Lim & Choi 2015, Zhang et al. 2017, Chen et al. 2018).

In the rainy season, large significant positive correlations were found between EC and the variables PO_4 / NO_3 / NO_2 / Cl-a, between Cl-a and TSS / VSS, between VSS and DBO_5, and very large correlations between variables NO_2 and PO_4, between TSS and VSS and between DBO_5 and Cl-a. The correlation between EC and PO_4 / NO_3 for the rainy season was also observed in Ribeiro et al. (2017) and may be related to leaching of soil and the transport of ions to water bodies, increasing the water conductivity.

Comparing the two periods analyzed (dry and rainy), it was observed that the high correlations between nitrogen and phosphate compounds occur for the two periods, thus

suggesting that high concentrations may be associated with point pollution from domestic sewage, which are responsible for the growth of plant biomass, causing the deterioration of water quality. The correlation with BOD_5 is more evident in the rainy season, which corresponds to the period of higher temperature, which intensifies the degradation process of organic matter.

Principal component analysis

As shown in Table II, the KMO results for the two data sets have values above 0.50, considered satisfactory, indicating the possibility of using PCA (Hair et al. 2014). Bartlett's test presented a p-value <0.05, rejecting the null hypothesis that the correlation matrix is an identity matrix, showing that the variables are correlated and the PCA can be applied to the data sets of this study.

Additionally, the correlation matrix (Figure 4a-b) shows that there is a substantial number of correlations between variables greater than 0.30, which also justifies the application of PCA (Hair et al. 2014).

The eigenvalues were obtained to set the number of principal components and to reduce the number of variables to be analyzed. Components with eigenvalues above 1 were selected for analysis (Mustapha et al. 2012, Barakat et al. 2016, Chen et al. 2018). Thus, for

the dry season, the first three components were selected, which explained 73.75% of the total data variance (Figure 5a). For the rainy season, the first four components were selected, which explained 76.20% of the total data variance (Figure 5b).

The correlation between variables and principal components are represented by the factor loading (Tables III and IV), which were classified as "strong" (> 0.75), "moderate" (0.50 to 0.75) and " weak "(0.30-0.50) (Liu et al. 2003, Barakat et al. 2016). The graphs in Figures 6 and 7 represent the principal components and the contribution of main variables and observations in data variability.

Dry season

The results obtained from the PCA, for the dry season, are shown in Table III and Figure 6a-d.

For the dry season (Table III), the first principal component (C1) was responsible for 38.90% of the total data variance and showed moderate positive correlations with BOD_5 and NO_3, strong positive correlations with EC, NO_2, PO_4, TP and Cl-a, and weak negative correlation with DO. It can be seen in Figure 6a-d greater contribution of nitrogen and phosphate compounds, EC and Cl-a that occur for point Ab-2 in four different months during the analyzed period, with no predominance of any monthly observation.
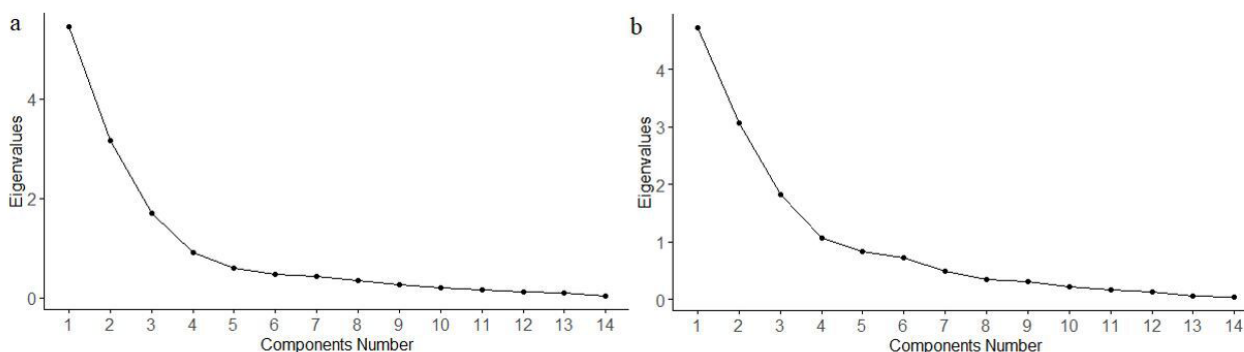

Figure 5. Eigenvalues of the principal components for the dry season (a) and for the rainy season (b).

Table III. KMO criterion and Bartlett's test of sphericity results for the dry and rainy seasons.

Evaluation Method	Dry season	Rainy season
Kaiser-Meyer-Olkin	0.77	0.72
Bartlett's test of sphericity	p = 0.000	p = 0.000

Table IV. Factor loading of variables and principal components for the dry season.

Variables	C 1	C 2	C 3
Q	0.26	**-0.68**	**0.42**
pH	0.13	**0.31**	**-0.8**
Temp	0.29	-0.22	**0.7**
EC	**0.87**	0.02	-0.11
DO	**-0.47**	**-0.58**	-0.16
Turb	-0.02	**0.82**	0
BOD_5	**0.7**	0.26	-0.12
NO_3	**0.75**	**-0.39**	0.01
NO_2	**0.93**	-0.19	-0.08
PO_4	**0.95**	-0.16	-0.12
TP	**0.88**	-0.12	0.02
Cl-a	**0.79**	0.28	-0.07
TSS	0.04	**0.71**	**0.45**
VSS	0.25	**0.82**	**0.36**
Eigenvalues	5.45	3.17	1.71
Explained variance (%)	38.9	22.62	12.24
Accumulated variance (%)	38.9	61.52	73.75

Considering that in the dry season there is less influence from diffuse sources, is implied that C1 is mainly associated with point sources of pollution, mostly from the discharge of effluents from STP Itirapina at point Ab-2, responsible for the increase of nutrients in water, thus causing favorable conditions for eutrophication, increased organic matter and reduced DO.

Similar results were observed in other studies, which associated the first principal component to organic and nutrient pollution from domestic wastewater and water treatment discharges (Shrestha & Kazama 2007, Finkler et al. 2015).

The second principal component (C2), in the dry season, was responsible for 22.62% of the total data variance and showed weak positive correlation with pH, moderate positive correlation with TSS and strong positive correlation with VSS and Turb. Still, it showed weak negative correlation with NO_3 and moderate negative correlations with Q and DO. The greatest contribution of VSS is observed at the Lb-2 point in July.

Considering the negative correlation with Q and the non-influence of diffuse pollution, C2 can be attributed to the point pollution associated with the erosion process on the margins of the monitoring points, probably caused by the inadequate handling of livestock, which increases the amount of solids and Turb in water. The positive correlation and the greater contribution of VSS may be associated with cattle trampling on soil, which intensifies the erosion process on the margins, thus increasing the amount of solids associated with organic matter, possibly animal waste. The impact of cattle trampling on soil erosion processes on the margins of water bodies was also observed in Cerqueira et al. (2017).

The third principal component (C3), in the dry season, was responsible for 12.24% of the total data variance. This component had weak positive correlations with Q, TSS and VSS, moderate positive correlation with Temp, and strong negative correlation with pH. It was observed that there was greater contribution of variables TSS, VSS, Q and Temp, and negative contribution of pH in the points It-1, It-3 and It-4, in April.

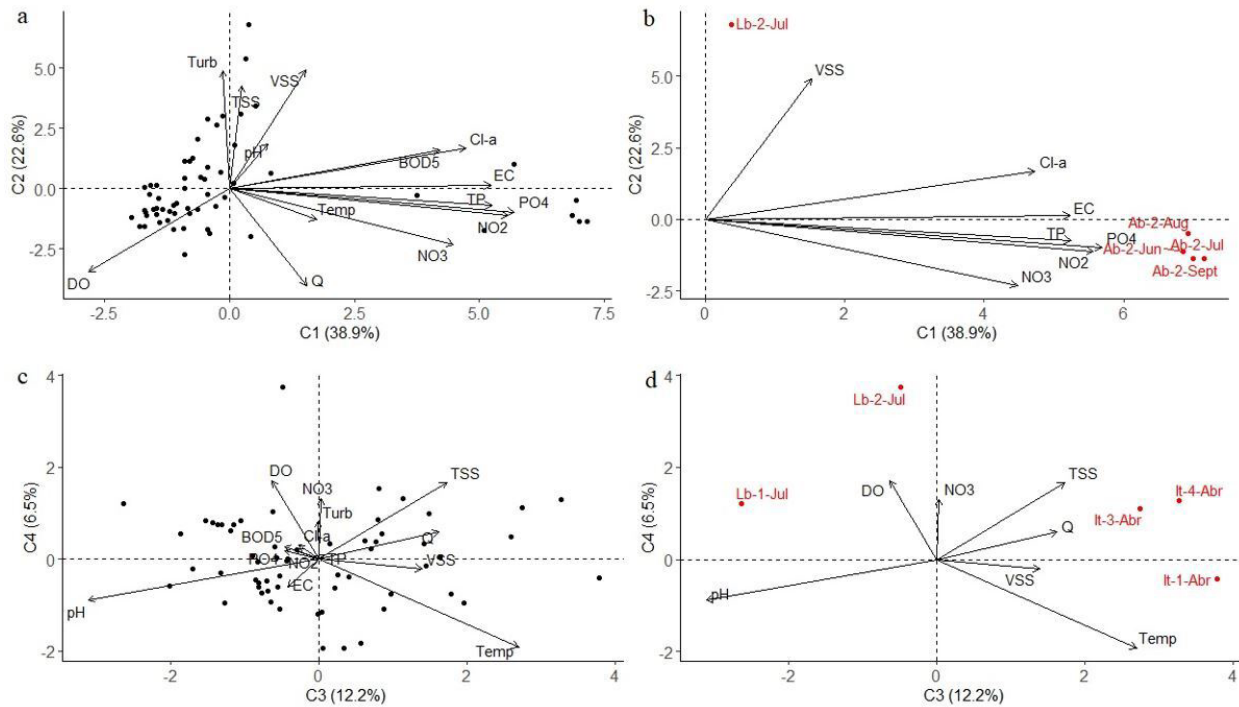This component also represents the pollution generated by the presence of

Figure 6. PCA analysis for dry season data set showing the C1 and C2 (a); variables and sample points of C1 and C2 with higher contribution (b); variables of the C3 and C4 (c); variables and sample points of C3 and C4 with higher contribution (d).

solids in water. However, as it showed greater correlation with Q, it can be said that this result may be associated with the high accumulated precipitation that occurred 15 days before data collection in April, responsible for intensifying the erosive processes and the leaching of acidic soil in the basin, thus increasing the amount of solids associated with organic compounds in the water body and reducing water pH.

Rainy season

The results obtained from the PCA, for the rainy season, are shown in Table V and Figure 7a-d.

In the rainy season, C1 was responsible for 33.72% of the total data variance and showed weak positive correlations with TSS, moderate positive correlations with the variables VSS, TP, PO_4, NO_2, EC and Temp, strong positive correlations with BOD_5 and Cl-a, and strong negative correlation with DO. The observed correlations indicate that C1, in the rainy season,

represents both point pollution caused by wastewater, and diffuse pollution associated with runoff and carrying of fertilizers used in agricultural areas.

However, it can be seen in Figure 7a that the dumping of the STP Itirapina at point Ab-2 had greater influence on data variability and the contribution of variables related to the eutrophication process occurred practically for the entire period analyzed.

As for the dry season, the increase in nutrients in water was responsible for the increase in plant biomass, causing favorable conditions for eutrophication, increased organic matter, reduced DO and, consequently, reduced water quality. Mixed pollution from wastewater and agricultural activities was also represented in the first principal component in (Zhao et al. 2011).

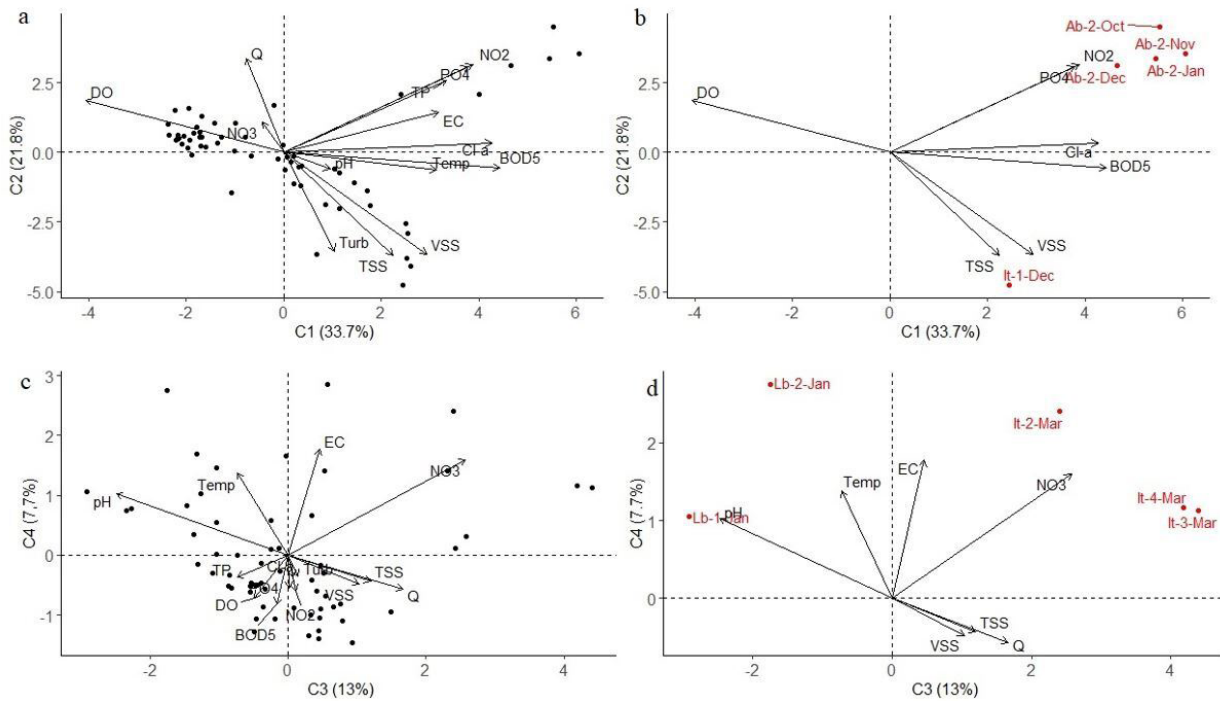Component C2, in the rainy season, was responsible for 21.85% of the total data

Figure 7. PCA analysis for rainy season data set showing the C1 and C2 (a); variables and sample points of C1 and C2 with higher contribution (b); variables of the C3 and C4 (c); variables and sample points of C3 and C4 with higher contribution (d).

variance showing weak positive correlation with DO and TP, moderate positive correlations with variables Q, PO_4 and NO_2, and moderate negative correlations with Turb, TSS and VSS. However, through the results generated, it was not possible to identify which observations (month and monitoring point) had the greatest contribution.

This component may be associated with the effect of the flow in dilution of solids caused by a greater amount of water, which also reduces Turb, in addition to showing that depending on the characteristics of the river channel, higher flow promotes oxygenation due to turbulence of water in these areas, thereby increasing dissolved oxygen levels. The increase in DO caused by the increase in water flow during the rainy season was also observed by Mustapha et al. (2012).

In the rainy season, C3 was responsible for 12.97% of the total data variance, with weak positive correlation with TSS and VSS, moderate positive correlation with Q, strong positive correlation with NO_3 and moderate negative correlation with pH. This component can be attributed to the natural processes of runoff over the soil that have an acidic characteristic, responsible for reducing water pH . Similar results were observed by Barakat et al. (2016).

However, the correlation of NO_3 may be a consequence of diffuse pollution, caused by runoff and soil leaching from agricultural areas that have extensive use of nitrogen fertilizers. There is greater contribution of the variable NO_3 in March in the Itaqueri River monitoring points, which may be a consequence of the nitrogen fertilization possibly carried out after planting eucalyptus and pine around these points. The cover fertilization for silviculture, in the state of São Paulo, is generally carried out from 3 months after planting, that occurs between October and March (Da Silva & Angeli 2006).

Table V. Factor loading of variables and principal components for the rainy season.

Variables	C 1	C 2	C 3	C 4
Q	-0.15	**0.64**	**0.51**	-0.17
pH	0.18	-0.12	**-0.75**	**0.31**
Temp	**0.59**	-0.12	-0.22	**0.42**
EC	**0.6**	0.27	0.14	**0.54**
DO	**-0.77**	**0.36**	-0.15	-0.22
Turb	0.2	**-0.67**	0.04	-0.11
BOD_5	**0.84**	-0.11	-0.05	-0.25
NO_3	-0.08	0.2	**0.78**	**0.49**
NO_2	**0.74**	**0.6**	0.04	-0.18
PO_4	**0.72**	**0.58**	0.01	-0.17
TP	**0.63**	**0.49**	-0.22	-0.11
Cl-a	**0.81**	0.06	0.02	-0.11
TSS	**0.42**	**-0.7**	**0.37**	-0.13
VSS	**0.56**	**-0.69**	**0.32**	-0.15
Eigenvalues	4.72	3.06	1.82	1.07
Explained variance (%)	33.72	21.85	12.97	7.66
Accumulated Variance (%)	33.72	55.57	68.54	76.2

According to Correa et al. (2006), NO_3 is one of the ions that is most likely to be leached, since it is not adsorbed by the components of the soil fractions and it is easily carried to the water body. For this reason, nitrogen in the form of NO_3 is one of the most important sources of water contamination caused by agricultural activities (Jadoski et al. 2010).

Component C4, in the rainy season, was responsible for 7.66% of the total data variance, showing weak positive correlation with pH, Temp and NO_3 and moderate positive correlation with EC. NO_3 was also one of the variables with the greatest contribution to C4, occurring mainly

at the Lb-2 point, being responsible for the increase of EC in the water body.

The last component also corresponds to mixed pollution from point and diffuse sources mainly associated with NO_3. However, it is noted that this component represents milder pollution, where the eutrophication process has not occurred.

Cluster analysis

The CA was applied to classify the monitoring points according to their similarities in water quality. The dendrograms were obtained using the WARD method for the average data of the dry and rainy season, as shown in Figure 8a-b.

It is observed that for the two periods analyzed, there was the formation of two large groups that represent the similarity between the type and intensity of pollution. In the two periods analyzed, the first group was formed only by point Ab-2, indicating that regardless of environmental influences, such as precipitation and air temperature, this point has the worst water quality conditions. While the second group, formed by the other points, represents moderate pollution, in the two periods analyzed.

However, it can be seen that the second group presented a subdivision, where the points It-3, It-4, Lb-3 and Lb-4, located in the areas close to the Lobo reservoir, are separated from the others that are located in areas close to its sources. Still, it can be observed that among the points close to the springs, the points It-2 and Lb-2 showed greater similarity to each other.

Comparing the two periods analyzed, it is observed that the Lb-1 point was classified in different groups, being closer to Ab-2 for the rainy season, which can be explained by the fact that the runoff was responsible for the worsening of water quality conditions at this point.
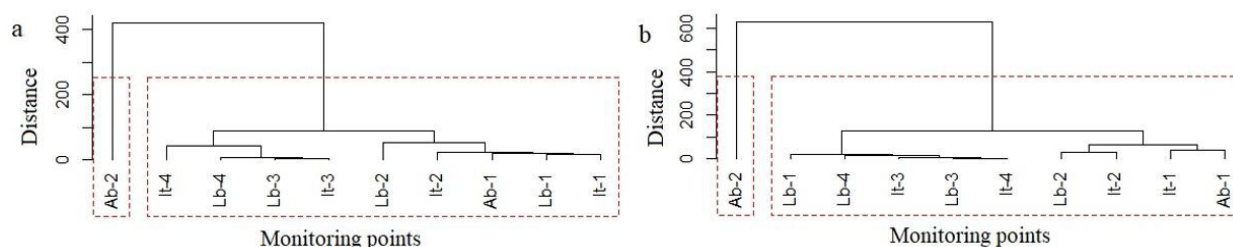
Figure 8. Dendogram for spatial analysis using the WARD method for the dry season (a) and for the rainy season (b).

CONCLUSIONS

The analysis of water quality at LSRB monitoring points allowed to verify the conditions of water bodies over the period of 12 months. Regarding compatibility with the legislation, it was observed that points It-1, It-2 and Ab-2 presented parameters outside the standards established by CONAMA Resolution 357/2005 for water bodies class 2.

In view of the large number of data obtained in the present study and the complexity of the variables involved in water pollution processes, the use of multivariate statistical techniques was essential for the identification of main factors responsible for temporal and spatial variability of LSRB water quality characteristics, as well as pointing out the possible sources of pollution in water bodies.

Through PCA, it was possible to identify the main variables and monitoring points that most contributed to data variability. It was noted that, regardless of environmental factors, one of the main causes is related to point pollution due to the discharge of effluents from the sewage treatment plant of Itirapina city at point Ab-2, responsible for the increase in nutrients, organic contamination, DO reduction, and consequently, deterioration of water quality.

In the dry season, erosion processes on river banks are responsible for increasing the concentration of suspended solids in water, conditions that are intensified by inadequate livestock management in the region. This impact is evident at point Lb-2. However, it is observed that this type of point pollution occurs occasionally and not throughout the year.

For the rainy season, the causes pointed out as responsible for the variation in water quality are associated with point pollution, such as wastewater discharge, and also with diffuse pollution, caused by runoff and leaching of agricultural areas that carry, together with solids, fertilizer residues to water bodies, thereby increasing the amount of nitrogen compounds and reducing water quality.

The application of multivariate statistical analysis showed great advantages over the descriptive approaches traditionally used in most studies, generating more significant and more relevant results to the understanding of data variability, which will complement other studies already carried out at LSRB that used simplified analysis techniques.

In addition, it is encouraged that further studies are carried out and consider the use of statistical methodologies and the incorporation of other factors related to water quality, which were not addressed in this study, such as land use variables.

Finally, considering multivariate statistical techniques as important tools, it is expected that the results generated can contribute to the planning and management of water resources in the LSRB.

Acknowledgments

REFERENCES

ALVARENGA LA, MARTINS MPP, CUARTAS LA, PENTEADO VA & ANDRADE A. 2012. Estudo da qualidade e quantidade da água em microbacia, afluente do rio Paraíba do Sul – São Paulo, após ações de preservação ambiental. Rev Ambient Água 7: 228-240.

ALVARENGA LA, MELLO CR DE, COLOMBO A, CUARTAS LA & CHOU SC. 2016. Hydrological responses to climate changes in a headwater watershed. Cien e Agrotecnologia 40: 647-657.

ALVARES CA, STAPE JL, SENTELHAS PC, MORAES GONÇALVES JL DE & SPAROVEK G. 2013. Köppen's climate classification map for Brazil. Meteorol Z 22: 711-728.

ANJINHO P DA S, BARBOSA MAGA, COSTA CW & MAUAD FF. 2021. Environmental fragility analysis in reservoir drainage basin land use planning: A Brazilian basin case study. Land Use Policy 100: 104946.

ANJINHO PS, NEVES GL, BARBOSA MAGA & MAUAD FF. 2020. Análise da qualidade das águas e do estado trófico de cursos hídricos afluentes ao reservatório do Lobo, Itirapina, São Paulo, Brasil. Rev Bras Geog Fís 13: 364-376.

BAI Y, OCHUODHO TO & YANG J. 2019. Impact of land use and climate change on water-related ecosystem services in Kentucky, USA. Ecol Indic 102: 51-64.

BARAKAT A, EL BAGHDADI M, RAIS J, AGHEZZAF B & SLASSI M. 2016. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. Int Soil Water Conserv Res 4: 284-292.

BEN-ELEDO V, KIGIGHA L, IZAH S & ELEDO B. 2017. Water Quality Assessment of Epie Creek in Yenagoa Metropolis, Bayelsa State, Nigeria. Arch Curr Res Int 8: 1-24.

BRAZIL. 2005. National Environmental Council Resolution (CONAMA) 357. Dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu enquadramento, bem como estabelece as condições e parâmetros de lançamento de efluentes, e dá outras providências (in Portuguese). Union Official Diary, DF. http://www2.mma.gov.br/port/conama/legiabre.cfm?codlegi=459.

CARVALHO LLS DE, LACERDA CF DE, ANDRADE EM DE, LOPES FB, JÚNIOR MV & CARVALHO CM DE. 2017. Variabilidade espacial e temporal da qualidade da água de poços no perímetro irrigado do Baixo Acaraú - CE. Rev Bras Agric Irrig 11: 1348-1357.

CERQUEIRA J DOS S, ALBUQUERQUE HN DE & ARAÚJO SM DE. 2017. Criação de bovinos no Complexo Aluízio Campos e os impactos do pastejo sobre a compactação do solo. Rev Espacios 38: 27-34.

CETESB. 2019. Apêndice E – Significado Ambiental das Variáveis de Qualidade – Águas Interiores, In: Qualidade das águas interiores no estado de São Paulo.

CHEN R, JU M, CHU C, JING W & WANG Y. 2018. Identification and Quantification of Physicochemical Parameters Influencing Chlorophyll-a Concentrations through Combined Principal Component Analysis and Factor Analysis: A Case Study of the Yuqiao Reservoir in China. Sustainability 10: 936.

CORREA RS, WHITE RE & WEATHERLEY AJ. 2006. Risk of nitrate leaching from two soils amended with biosolids. Water Resour 33: 453-462.

CRISPIM DL, FERNANDES LL, FILHO DFF & LIRA BRP. 2020. Comparação de métodos de agrupamentos hierárquicos aglomerativos em indicadores de sustentabilidade em municípios do estado do Pará. Res Soc Dev 9: e60922067.

DA SILVA PHM & ANGELI A. 2006. Implantação e manejo de florestas comerciais. http://www.rsflorestal.com.br/arquivos/artigos/f/df18.pdf. Accessed 7 out. 2020.

MUÑIZ CD, GARCÍA NIETO PJ, ALONSO FERNÁNDEZ JR, MARTÍNEZ TORRES J & TABOADA J. 2012. Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain). Sci Total Environ 439: 54-61.

FINKLER NR, PERESIN D, COCCONI J, BORTOLIN TA, RECH A & SCHNEIDER VE. 2015. Qualidade da água superficial por meio de análise do componente principal. Rev Ambient Agua 10: 782-792.

FRASCARELI D, CARDOSO-SILVA S, MIZAEL JOSS, ROSA AH, POMPÊO MLM, LÓPEZ-DOVAL JC & MOSCHINI-CARLOS V. 2018. Spatial distribution, bioavailability, and toxicity of metals in surface sediments of tropical reservoirs, Brazil. Environ Monit Assess 4: 190:199.

GRIZZETTI B, LANZANOVA D, LIQUETE C, REYNAUD A & CARDOSO AC. 2016. Assessing water ecosystem services for water resource management. Environ Sci Policy 61: 194-203.

GRIZZETTI B, PISTOCCHI A, LIQUETE C, UDIAS A, BOURAOUI F & BUND W VAN DE. 2017. Human pressures and ecological status of European rivers. Sci Rep 7: 205.

GULGUNDI MS & SHETTY A. 2018. Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques. Appl Water Sci 8: 43.

HAIR JF, BLACK WC, BABIN BJ & ANDERSON RE. 2014. Multivariate Data Analysis. 7ª ed. Ed. United States of America: Pearson, 734 p.

HONGYU K, SANDANIELO VLM & OLIVEIRA JUNIOR GJ DE. 2015. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. Eng Sci Technol 1: 83-90.

HOSSAIN MA, ALI NM, ISLAM MS & HOSSAIN HMZ. 2015. Spatial distribution and source apportionment of heavy metals in soils of Gebeng industrial city, Malaysia. Environ Earth Sci 73: 115-126.

JADOSKI S, SAITO LR, PRADO C, LOPES É & SALES LLSR. 2010. Características da lixiviação de nitrato em áreas de agricultura intensiva. Pesqui Apl Agrotec 3: 193-200.

LI T, LI S, LIANG C, BUSH RT, XIONG L & JIANG Y. 2018. A comparative assessment of Australia's Lower Lakes water quality under extreme drought and post-drought conditions using multivariate statistical techniques. J Clean Prod 190: 1-11.

LIM J & CHOI M. 2015. Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea. Environ Monit Assess 187.

LIU C-W, LIN K-H & KUO Y-M. 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ 313: 77-89.

MATTA G, SRIVASTAVA S, PANDEY RR & SAINI KK. 2017. Assessment of physicochemical characteristics of Ganga Canal water quality in Uttarakhand. Environ Dev Sustain 19: 419-431.

MEA. 2005. Millennium Ecosystem Assessment (Program) (Ed), 2005. Ecosystems and human well-being: synthesis. Washington: Island Press, 137 p.

MELLO K DE, TANIWAKI RH, PAULA FR DE, VALENTE RA, RANDHIR TO, MACEDO DR, LEAL CG, RODRIGUES CB & HUGHES RM. 2020. Multiscale land use impacts on water quality: Assessment, planning, and future perspectives in Brazil. J Environ Manage 270: 110879.

MENCIÓ A, BOY M & MAS-PLA J. 2011. Analysis of vulnerability factors that control nitrate occurrence in natural springs (Osona Region, NE Spain). Sci Total Environ 409: 3049-3058.

MENDES L DA S & FERREIRA IM. 2014. Influência da sazonalidade na qualidade da água bruta no município de Ituiutaba - MG. Hygeia 10: 97-105.

MIZAEL JDOSS, CARDOSO-SILVA S, FRASCARELI D, POMPÊO MLM & MOSCHINI-CARLOS V. 2020. Ecosystem history of a tropical reservoir revealed by metals, nutrients and photosynthetic pigments preserved in sediments. Catena 184: 104242.

MORUZZI RB, HONDA FP & NAVARRO GRB. 2012. Avaliação de cargas difusas e simulação de autodepuração no córrego da Água Branca, Itirapina (SP). Geosciences 3: 447-458.

MUSTAPHA A, ARIS AZ, RAMLI MF & JUAHIR H. 2012. Spatial-temporal variation of surface water quality in the downstream region of the Jakara River, north-western Nigeria: A statistical approach. J Environ Sci Health Part A 47: 1551-1560.

NOZAKI CT, MARCONDES MA, LOPES FA, SANTOS KF DOS & LARIZZATTI PS DA C. 2014. Comportamento temporal do oxigênio dissolvido e pH nos rios e córregos urbanos. Atas Saúde Ambient 2: 29-44.

PANTANO G, GROSSELI GM, MOZETO AA & FADINI PS. 2016. Sustainability in phosphorus use: a question of water and food security. Quím Nova 39: 732-740.

PENA MG, MOREIRA GCC, GUIMARÃES LFD, LAURETO CR, ALBUQUERQUE PHM, CARVALHO AXY DE & BASSO GG. 2017. Clusterização Espacial e Não Espacial: Um Estudo Aplicado à Agropecuária Brasileira. TEMA (São Carlos) 18: 69-84.

PHAM HV, TORRESAN S, CRITTO A & MARCOMINI A. 2019. Alteration of freshwater ecosystem services under global change – A review focusing on the Po River basin (Italy) and the Red River basin (Vietnam). Sci Total Environ 652: 1347-1365.

RÉHMAN S, HUSSAIN Z, ZAFAR S, ULLAH H, BADSHAH S, AHMAD S, SALEEM J & JINNAH F. 2018. Assessment of Ground Water Quality of Dera Ismail Khan, Pakistan, Using Multivariate Statistical Approach. Sci Technol Soc 37: 173-183.

RIBEIRO TG, BOAVENTURA GR, CUNHA LS DA & PIMENTA SM. 2017. Estudo Da Qualidade Das Águas Por Meio Da Correlação De Parâmetros Físico-Químicos, Bacia Hidrográfica Do Ribeirão Anicuns. Geochim Bras 30: 84-94.

ROCHA FC, ANDRADE EM & LOPES FB. 2014. Water quality index calculated from biological, physical and chemical attributes. Environ Monit Assess 187: 4163.

ROYSTON JP. 1983. Some Techniques for Assessing Multivarate Normality Based on the Shapiro- Wilk W. App Statist 32: 121.

SANTOS M, MATUCK C, ADAMI F, REIS K & BARRELLA W. 2017. Análise de Agrupamento Hierárquico Aglomerativo aplicada à Ecologia - Teoria e Prática. Unisanta BioScience 6: 68-77.

SAXENA A, PRASAD M, GUPTA A, BHARILL N, PATEL OP, TIWARI A, ER MJ, DING W & LIN C-T. 2017. A review of clustering techniques and developments. Neurocomputing 267: 664-681.

SHRESTHA S & KAZAMA F. 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environ Model Softw 22: 464-475.

TILBURG CE, JORDAN LM, CARLSON AE, ZEEMAN SI & YUND PO. 2015. The effects of precipitation, river discharge, land use and coastal circulation on water quality in coastal Maine. R Soc Open Sci 2: 140429.

TREBUŇA P & HALČINOVÁ J. 2013. Mathematical Tools of Cluster Analysis. Appl Math 4: 814-816.

TUNDISI JG, MATSUMURA-TUNDISI T, CIMINELLI VS & BARBOSA FA. 2015. Water availability, water quality water governance: the future ahead. Proc Int Assoc Hydrol 366: 75-79.

VON SPERLING M. 2005. Introdução à Qualidade Das Águas e Ao Tratamento de Esgotos. Belo Horizonte: Editora UFMG/DESA, 452 p.

ZHANG C, ZHANG W, HUANG Y & GAO X. 2017. Analysing the correlations of long-term seasonal water quality parameters, suspended solids and total dissolved solids in a shallow reservoir with meteorological factors. Environ Sci Pollut Res 24: 6746-6756.

ZHAO H, JIANG Q, MA Y, XIE W, LI X & YIN C. 2018. Influence of urban surface roughness on build-up and wash-off dynamics of road-deposited sediment. Environ Pollut 243: 1226-1234.

ZHAO J, FU G, LEI K & LI Y. 2011. Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. J Environ Sci 23: 1460-1471.

ZHOU Y, XU JF, YIN W, AI L, FANG NF, TAN WF, YAN FL & SHI ZH. 2017. Hydrological and environmental controls of the stream nitrate concentration and flux in a small agricultural watershed. J Hydrol 545: 355-366.

GABRIELA L. NEVES[1]
https://orcid.org/0000-0002-4959-9005

TAINÁ T. GUIMARÃES[2]
https://orcid.org/0000-0002-6362-6591

PHELIPE S. ANJINHO[1]
https://orcid.org/0000-0002-0162-2042

MARIANA A.G.A. BARBOSA[1]
https://orcid.org/0000-0001-7899-4752

ALLITA R. DOS SANTOS[1]
https://orcid.org/0000-0002-5394-4035

JORIM S. VIRGENS FILHO[3]
https://orcid.org/0000-0003-2722-9481

FREDERICO F. MAUAD[1]
https://orcid.org/0000-0002-2477-2019

[1]Universidade de São Paulo (USP), Centro de Recursos Hídricos e Estudos Ambientais, Avenida Domingos Innocentini, Km 13, Itirapina, SP, Brazil

[2]Universidade do Vale do Rio dos Sinos (Unisinos), Centro de Ciências Exatas e Tecnológicas, Avenida Unisinos, 950, 93022-750 São Leopoldo, RS, Brazil

[3]Universidade Estadual de Ponta Grossa, Departamento de Matemática e Estatística, Avenida Carlos Cavalcanti, 4748, 84030-900 Ponta Grossa, PR, Brazil

Correspondence to: **Gabriela L. Neves**
E-mail: gabriela.leiteneves@usp.br

Author Contributions

Gabriela Leite Neves developed the study and worked on all methodological stages, collection and organization of data, statistical analysis, interpretation of results and writing of the manuscript. Tainá Thomassim Guimarães, Phelipe da Silva Anjinho and Mariana Abibi Guimarães Araujo Barbosa assisted in data colletion and worked on interpretation of results, critical review and final writing of the manuscript. Jorim Sousa das Virgens Filho, Allita Rezende dos Santos and Frederico Fabio Mauad worked on interpretation of results, critical review and final writing of the manuscript.