Environ Monit Assess (2023) 195:1240
https://doi.org/10.1007/s10661-023-11889-0

RESEARCH

Linking multivariate statistical methods and water quality indices to evaluate the natural and anthropogenic geochemical processes controlling the water quality of a tropical watershed

Débora Francisca Morais da Silva · Letícia Mariana Lopes da Silva · Jeremie Garnier · Daniel Ferreira Araújo · Daniel Santos Mulholland

Received: 23 February 2023 / Accepted: 16 September 2023

Abstract The improvement of water management requires monitoring techniques that accurately evaluate water quality status and detect the effects of land use changes on water chemistry. This study aimed to evaluate how multivariate statistical methods and water quality indices can be applied together to evaluate the processes controlling water chemical composition and the overall water quality status of a tropical watershed. Thirty-four water samples were collected in the Formoso River basin, located on the border of the Amazon Forest. Water parameters were measured in situ using a multiparameter and in the lab using spectroscopic and volumetric techniques. The water quality dataset was interpreted through principal component analysis, multivariate linear regression, and water quality indices. Statistical methods allowed us to identify the sources and geochemical processes controlling water quality chemistry, which were carbonate dissolution, runoff/erosion, nutrient input due to anthropogenic activities, and redox reactions in flooded zones. They were also used to create linear functions to evaluate the effects of land use changes on the geochemical processes controlling water chemistry. Conversely, the water quality indices provide information about the overall condition of the water. The Weight-Arithmetic Quality Index correctly evaluates water suitability for its multiple uses, according to the Brazilian guidelines. Conversely, the Ontario Water Quality Index is not suitable to evaluate the water quality of tropical rivers, since the usual higher water temperature and the low oxygen contents associated with tropical environments result in biased water quality evaluations by this index.

Keywords Water monitoring · Formoso river · Water geochemistry · Contamination

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10661-023-11889-0.

D. F. M. da Silva · L. M. L. da Silva · D. S. Mulholland (✉)
Laboratório de Águas e Efluentes e Laboratório de Análises Ambientais - Química Ambiental, Universidade Federal de Tocantins, Rua Badejós, Lote 7, Chácaras 69/72, Gurupi, Tocantins 77402-970, Brazil
e-mail: danielsm@uft.edu.br

J. Garnier
Laboratório de Geoquímica - Instituto de Geociências, Universidade de Brasília, Campus Darcy Riberio, , Brasília, Distrito Federal 70910-900, Brazil

D. F. Araújo
Ifremer, CCEM-Contamination Chimique des Ecosystèmes Marins, F-44000 Nantes (Loire-Atlantique), France

Introduction

Water is imperative for life; nevertheless, it remains one of the world's most vulnerable resources to chemical emissions from anthropogenic sources (punctual and diffuse) and drastic land use modifications,

Published online: 22 September 2023

mainly related to urbanization sprawl and agriculture expansion (Varol, 2020a). These activities affect water quality by altering the natural biogeochemical cycles and the dynamics of soil erosion, transport, and deposition. The input of metals and potentially toxic substances into aquatic systems can cause adverse effects on human health and aquatic biota, impairing the multiple uses of water and the adequate fulfillment of the demands of society (Iticescu et al., 2019; Şener et al., 2017). Therefore, regulation and policies on pollutant emissions and the improvement of water management require monitoring techniques to accurately evaluate water quality status and identify natural and anthropogenic processes and sources controlling water chemical composition (Gnanachandrasamy et al., 2020; Nong et al., 2020).

One of the most used water monitoring techniques applies water quality indices since they have been conceived to substantially decrease the data volume and simplify the interpretation of the water quality status by using key parameters. They can be used as an accessible and easy tool for managing and monitoring water resources and assessing overall water quality. An index is considered as a number resulting from a mathematical or statistical operations on a group of indicators (parameters) that expresses a certain quality value in a dimensionless way (Bordalo et al., 2006; Sánchez et al., 2007; Şener et al., 2017). Horton (1965) and Brown et al. (1970) were the first to propose the usage of a Water Quality Index (WQI). Subsequently, several national and international organizations developed water quality indices, such as the Weight Arithmetic Water Quality Index (WAWQI), the National Sanitation Foundation Water Quality Index (NSFWQI), the Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI), and the Oregon Water Quality Index (OWQI) (Tyagi et al., 2020). However, these indices were developed using quality criteria for river waters in temperate climates and do not consider the inherent geological heterogeneity. Therefore, using solely these indices may produce misperceptions while evaluating water quality worldwide.

Multivariate statistical techniques have also been extensively applied to evaluate water quality and the ecological status of aquatic systems by reducing data volume and simplifying the interpretation of complex water quality datasets. The most used statistical techniques applied to water quality evaluation are correlation analysis, principal component analysis, factor analysis, the hierarchical cluster analysis, and discriminant analysis (Shrestha & Kazama, 2007; Singh et al., 2004). These methods have been also applied to identify the pollution sources and the natural or anthropogenic factors or processes that influence aquatic systems conditions (Helena et al., 2000; Mar da Costa et al., 2016; Mulholland et al., 2012; Varol, 2020b).

The Formoso River watershed hosts several municipalities with poor sanitary conditions without sewage collection and treatment systems (SRHMA, 2007). Therefore, the disposal of untreated sewage in the soil and in the waters of the Formoso basin is a potential contamination source of organic matter, nutrients, and metals in surface and groundwaters. The Formoso River basin also hosts several agriculture projects that caused the suppression of vegetation in extensive areas. Floodplain areas usually contain several irrigation projects by flooding and sub-irrigation methods. Networks of water channels with more than 20,000 ha are used for rice crops in the rainy season (SRHMA, 2007). The intensive fertilizer loads applied to these crops can be transported to rivers through surface runoff, causing disruptive changes in the aquatic ecological balance, such as eutrophication.

The Formoso River basin is an emblematic case to evaluate the feasibility of water quality indices and multivariate statistic methods to understand the key processes controlling the water chemical composition and to predict the water quality status of rivers in developing countries of tropical environment where impacts on water quality are rarely stablished satisfactorily. To this end, this study aimed to evaluate how multivariate statistical methods and water quality indices can be applied together to evaluate the processes controlling water chemical composition and overall water quality status of a tropical watershed.

Material and methods

Study area

The Formoso River basin has an area of approximately 20,654 km^2 and is located in the southeast of the State of Tocantins (Brazil), bordering the Amazon Forest (Fig. 1). It hosts 21 small municipalities (<200 thousand inhabitants), of which only 7 have

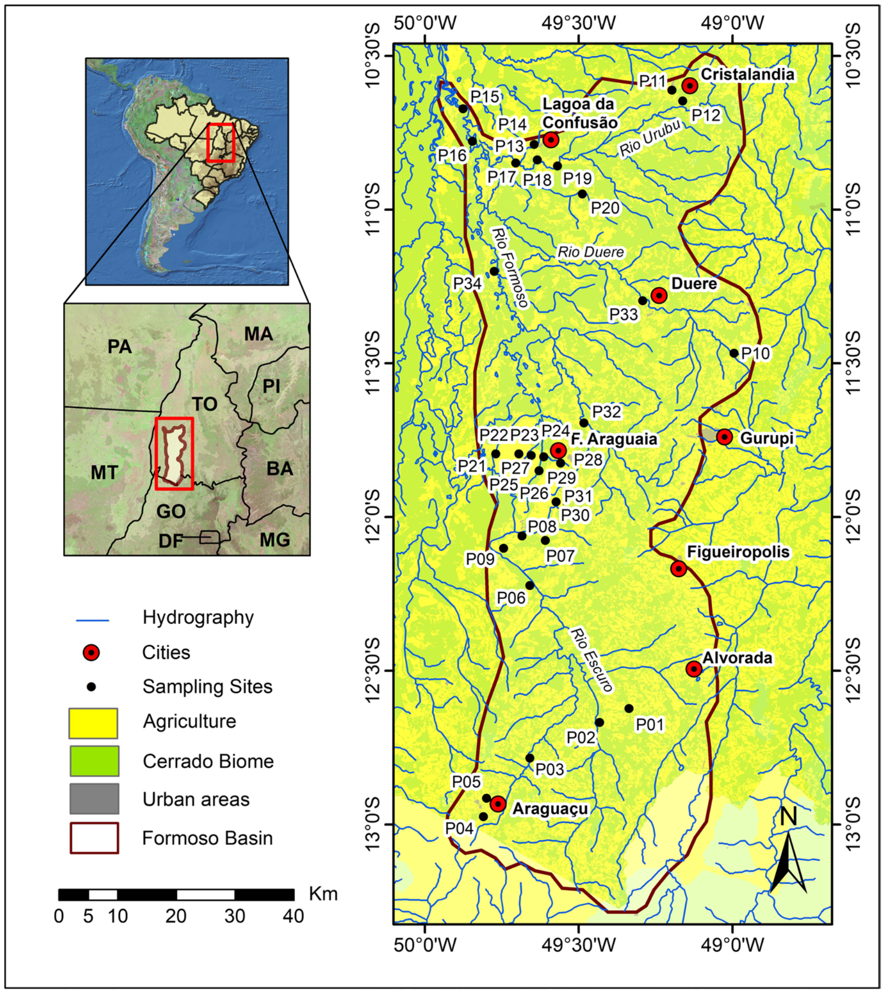

Fig. 1 Map showing the location of the Formoso River basin and the water sampling sites

urban centers inside the basin area. Agriculture and livestock projects represent approximately 90% of the land use practices. The most significant agriculture activities occur along the basin downstream, where the land is used for rice crops in the rainy season and for soybeans, corn, beans, and watermelon crops in the dry season (SRHMA, 2007).

The basin is situated in the northern sector of the Tocantins Province, which is bordered to the south by the Transbrasiliano Lineament, to the southwest and northwest by the Amazon Craton, to the southeast by the São Francisco Province, and to the northeast by the Parnaiba Province. The most important lithostratigraphic units found in the area are the following: (i) alluvial deposits found in Formoso River downstream; (ii) Couto Magalhaes e Xambioá Formations that host phyllites, metargillites, quartzites, and carbonates in the north-central region of the basin; and (iii) Rio dos Mangues Complex that hosts mainly gneisses in the basin upstream area (dos Santos, 2016). Highly weathered soils cover approximately 90% of the basin area, mainly composed of Plinthosols (15%), Ferralsols (33%), and concretionary soils (41%) (Fagundes, 2021; SRHMA, 2007). The basin is situated in a tropical wet and dry climate region, according to the Köppen climate classification. The average annual precipitation ranges from 1400 to 2200 mm, of which 70% occurs between November and March. The Formoso River discharge ranges from 0.611 m^3/s in the dry season to 97.02 m^3/s (Fagundes, 2021).

Water sampling and analysis

Thirty-four water samples were collected along the entire length of the Formoso River, its tributaries, and the irrigation channels located near the city of Formoso do Araguaia (Fig. 1). The samples were stored in 1-L polyethylene bottles previously rinsed with 1 M HCl. In the field, the bottles were previously conditioned with the riverine water from the sampling site. After collection, the samples were refrigerated at approximately 3 °C and protected from sunlight. At the laboratory, approximately 200 mL of the sample was filtered by a frontal vacuum filtration system equipped with HA membranes in cellulose esters, Millipore©, sterile, with 0.45 μm pore size. Aliquots of the filtered samples were separated and acidified (pH < 2) with conc. Merk Suprapur HNO_3 for the subsequent metal determination.

Water analysis was performed according to the methods reported in the "Standard methods of the examination of water and wastewater" (APHA, 2005). The determination of electrical conductivity (EC), total dissolved solids (TDS), dissolved oxygen (DO), and pH were performed in situ using a Hannah HI9828 multiparameter. In the laboratory, PO_4^{3-}, NO_3-N, NH_3-N, and color were measured in filtered samples by visible spectrophotometry using a Kasvi K37-VIS' spectrophotometer. CO_3^{2-}, HCO_3^-, and Cl^-, hardness, and alkalinity were determined in bulk samples by volumetric methods. The acidified filtered samples were analyzed by Inductively Coupled Plasma Optical Emission Spectrometry using an ICP-OES, 5100, Agilent at University of Brasília to measure the concentrations of dissolved Al, Fe, Mn, Zn, Co, Ni, Cr, Cu, Cd, As, and Pb. For the analyses performed herein, only Fe, Si, Mn, and Sr had concentrations above the limit of detection (LOD) of approximately 0.01 mg/L. Exceptionally, Na^+ was determined by flame emission spectrometry (FAES) using a Quimis-Q498M2 photometer. The NRC (SLRS-6) (CRM Environment Canada) was used for quality assurance (QA)/quality control (QC) of the water dissolved fraction analysis. The averaged measured concentration of the certified reference materials deviates within ± 5% of the certified values.

Multivariate statistical methods

Principal component analysis (PCA) is a statistical exploratory method to identify the relationships among the parameters. It reduces the variable dataset into uncorrelated combinations called principal components (PC), making easier to infer their geochemical associations (Tripathi & Singal, 2019). PCA was performed using a correlation matrix with Varimax rotation to maximize the variance weights in which only significant components with eigenvalues greater than 1.5 were selected (Mar da Costa et al., 2016; Mulholland et al., 2012). Parameters with concentrations below the limit of detection (LOD) were replaced by a LOD/$\sqrt{2}$ value as suggested by Verbovšek (2011) for geochemical data. The parameters with factor loadings > 0.7 in a same PC had high direct correlations, whereas the ones with factor loadings < −0.7 had indirect correlations. Direct and indirect correlations among water parameters are usually caused by a specific natural or anthropogenic process, and, therefore, they were used to understand the main geochemical mechanisms controlling water chemistry. The factor loadings and factor scores provided by PCA were calculated using the SPSS software v.18.0 (IBM).

The factor loadings describe the contribution of each water parameter to a particular principal component, in which large (positive or low) factor loadings represent strong relationship between the water parameters. The strong correlations between the water parameters permit the identification of the geochemical process controlling their variability. The factor scores (FS) represent the position of each observation (sample) in a new coordinated system of principal components. They are calculated by a linear combination of the water parameters (x_{ij}) and the factor loadings (a_{ij}) (Elemile et al., 2021). Therefore, the factor scores can indicate the intensity of each geochemical process in a given sample site and may be used as an index. The factor scores were calculated according to the following equation:

$$FS_{ij} = a_{1j}x_{1j} + a_{i2}x_{2j} + \cdots + a_{im}x_{mj} \tag{1}$$

where i, j, and m represent the component, the sample, and the total number of variables, respectively.

A further step applying a multivariate linear regression was used to reduce the number of water parameters needed to calculate the factor scores to create a more feasible index. Multivariate linear regression is a statistical method that predicts the value of a dependent variable (factor score) using the values of the independent variables (water parameters). The multivariate linear regressions were calculated using the forward regression method to maximize the coefficients of determination (R^2) and the adjusted R^2 (adj-R^2) with a p-value<0.01, using the least possible amount of water quality parameters. The new dependent variables generated by the multivariate linear regression were calculated using standardized and non-standardized values of the water quality parameters and are called now on as indices Z_x and I_x, respectively.

Water quality indices

Oregon Water Quality Index

The Oregon Water Quality Index (OWQI) is calculated using theoretical quality functions (curves) that transform variables with different units to a non-dimensional scale value called Sub-Index. The latter is then aggregated with a mathematical function to form a water quality index (Cude, 2001). OWQI developed a score that can integrate up to eight water quality characteristics into a single number to evaluate the general water quality. DO, pH, NH_3-N, and NO_3-N, as well as total solids were used herein to calculate this index. The original OWQI index uses the notion of harmonic averaging instead of arbitration in weighing the parameters and is mathematically expressed as follows:

$$OWQI = \sqrt{\frac{n}{\sum_{i=1}^{n} \frac{1}{Si^2}}} \tag{2}$$

where n is the number of subindices and Si is the subindex of the ith parameter.

Weight arithmetic quality index

This index assesses the water quality status using variables commonly presented in water monitoring programs and guideline established by local legislation with a mathematical approach. This index was calculated using the following parameters: Turbidity, color, Cl^-, NO_3-N, pH, TDS, Fe-dis, and DO. The WAWQI is calculated according to Eq. 3, in which each water quality parameter's unit weight (W_i) is calculated according to Eq. 4, the proportionality constant (K) is calculated using Eq. 5, and each parameter's quality rating scale (Q_i) is calculated using Eq. 6 (Tyagi et al., 2020). The variable V_i expresses the concentration of the ith parameter in the analyzed water, V_0 is the ideal value in pure water ($V_0=0$, except for pH and DO, which ideal values are 7.0 and 14.6 mg/L, respectively) and Si is the threshold value of the ith parameter stablished by local water quality guidelines.

$$WAWQI = \frac{\sum QiWi}{\sum Wi} \tag{3}$$

$$Wi = K/Si \tag{4}$$

$$K = \frac{1}{\sum \frac{1}{Si}} \tag{5}$$

$$Qi = 100[\left(\frac{Vi - Vo}{Si - Vo}\right)] \tag{6}$$

Irrigation water quality indices

The water quality for irrigation purposes usually associates its major ion concentrations with their effects on soils and plants. For instance, high salt concentration in irrigation water can be harmful to crops by changing soil structure, and plant metabolic processes, decreasing plant growth rates, and promoting salt accumulation in soil profiles (Singh et al., 2020). This study evaluated water quality for irrigation by four indices, i.e., sodium adsorption ratio (SAR), permeability index (PI), magnesium hazard (MH), and residual sodium carbonate (RSC).

The SAR evaluates the suitability of the water to be used for agricultural irrigation. Adsorbed Na ions promote soil clay particle dispersion, change soil structure, affect water infiltration rate, and lead to problems with crop production (Nagaraju et al., 2016). It evaluates Na concentration with respect to calcium and magnesium, expressed in milliequivalents per liter (Chebet et al., 2020), as shown below:

$$SAR = \frac{Na^+}{\sqrt{\frac{Ca^{2+}+Mg^{2+}}{2}}} \tag{7}$$

Similar to the SAR, the PI evaluates the effects of long-term use of mineral rich water in soil permeability and, consequently, in crop production. The PI is calculated using the criteria proposed by Doneen according to Eq. 8 (Nagaraju et al., 2016), using Ca^{2+},

Mg^{2+}, HCO_3^-, and Na^+ concentrations expressed in milliequivalents per liter.

$$PI = \frac{(Na^+ + \sqrt{HCO_3^-}) \times 100}{(Na^+ + Mg^{2+} + Ca^{2+})} \tag{8}$$

The MH index is applied to evaluate possible adverse effects of high concentration of Mg to plants (Ali & Ali, 2018; Wakeel, 2013). In most natural environments, Ca and Mg ions are present in the state of equilibrium. However, plant growth can decrease when Mg exceeds Ca in irrigation water due to a Mg-induced Ca deficiency, affecting crop production. The MH index is calculated using Eq. 9 with ion concentrations expressed in milliequivalents per liter (Chebet et al., 2020).

$$MH = \mathrm{Mg}^{2+} + \frac{100}{\mathrm{Ca}^{2+}\mathrm{Mg}^{2+}} \tag{9}$$

The RSC index expresses the excessive HCO_3^- and CO_3^{2-} concentrations when compared to Ca^{2+} and Mg^{2+} concentrations. Irrigation waters with excessive carbonate species can be balanced by Na^+ ions and cause Na_2CO_3 precipitation in soils, affecting crop production (Murtaza et al., 2021). Together with the SAR index, the RSC index can be used to evaluate sodicity hazard of irrigation waters. The RSC is calculated according to Eq. 10.

$$RSC = [(\mathrm{HCO}_3^- + \mathrm{CO}_3^{2-}) - (\mathrm{Ca}^{2+} + \mathrm{Mg}^{2+})] \tag{10}$$

Results and discussion

Water chemistry

The results from Table 1 and Online Resources show that the waters of the Formoso River have low nutrient concentrations. For instance, the mean and 95% confidence interval (CI) for the following parameters were as follows: NO_3-N (0.10 mg/L, 95% CI = 0.03–0.17), NH_3-N (0.03 mg/L, 95% CI = 0.01–0.05), and PO_4^{3-} (< 0.03 mg/L, 95%

Table 1 Statistical data obtained through the analysis of the Formoso River basin water and quality guidelines stablished by CONAMA for class 2 waters

	Mean	SD	Min	Max	CI (95%)	CV	CONAMA 357/05
pH	6.96	0.6	5.8	8.0	6.8–7.1	0.1	6.0—9.0
EC	96.3	72.2	16.0	244.0	72.1–120.6	0.7	–
TDS	48.4	36.5	8.0	122.0	36.2–60.7	0.8	< 500
DO	5.0	1.0	2.7	7.1	4.7–5.3	0.2	> 5
Turbidity	24.3	12.5	0.9	63.0	20.1–28.5	0.5	< 100
SS	7.8	7.9	0.4	31.5	5.2–10.4	1.0	–
Fe-total	1.0	1.2	< 0.02	6.3	0.6–1.4	1.1	–
Color	26.1	23.2	0.0	92.0	18.3–33.9	0.9	< 75
Alka	42.8	38.5	6.6	147.4	29.9–55.8	1.3	–
HCO_3^-	52.2	47.0	8.0	179.8	36.8–68.0	0.9	–
CO_3^{2-}	0.0	0.0	0.0	0.0	0.0–0.0	0.0	–
TH	40.8	37.4	6.0	138.0	28.3–53.4	0.9	–
Cl^-	7.2	6.4	3.0	37.0	5.1–9.3	0.9	< 250
NO_3-N	0.10	0.21	< 0.03	1.26	0.03–0.17	2.1	< 10
NH_3-N	0.03	0.05	< 0.03	0.20	0.01–0.05	2.0	–
PO_4^{3-}	< 0.03	0.03	< 0.03	0.10	0.01–0.03	1.5	–
Ca	11.2	9.5	2.4	40.0	8.0–14.4	1.2	–
Mg	3.2	3.7	< 0.01	14.6	2.0–4.5	1.1	–
Na	3.0	3.7	0.6	14.4	1.8–4.3	1.2	
Fe-Dis	0.15	0.15	< 0.01	0.68	0.10–0.21	1.0	< 0.3
Mn	0.05	0.11	< 0.01	0.61	0.01–0.09	2.23	–
Si	4.80	2.17	0.54	10.78	4.03–5.49	0.46	–
Sr	0.04	0.02	< 0.01	0.10	0.03–0.05	0.63	–

All data expressed in milligrams per liter, except color (mg/L Pt–Co), turbidity (FTU), EC (µS/cm), and pH

EC electrical conductivity, *TDS* total dissolved solids, *DO* dissolved oxygen, *SS* suspended solids, *Alka.* alkalinity, *TH* total hardness, *SD* standard deviation, *CI* 95% confidence interval, *CV* coefficient of variation

CI = 0.01–0.03). The waters also had a circumneutral pH (6.96, 95% CI = 6.8–7.1), low alkalinity (42.8 mg/L $CaCO_3$, 95% CI = 29.9–55.8), and low TDS (48.4 mg/L, 95% CI = 36.2–60.7). These findings shows that the waters are typical of oligotrophic aquatic environments and porewater of soils commonly found in the Brazilian Central-West Region (Lilienfein et al., 2000; Mar da Costa et al., 2016; Mulholland et al., 2012). The high coefficient of variation (> 1.0) found for some paraments measured herein (Table 1) shows that water chemical composition varies spatially throughout the basin.

Geochemical processes and sources

The principal component analysis (PCA) was applied to the water quality dataset summarized in Table 1. The dataset variability can be explained by three main components (PCs), representing 82% of its total variance (Fig. 2). PC1 had large factor loading (> 0.7) for HCO_3^-, Ca^{2+}, Mg^{2+}, Sr^{2+}, total hardness, alkalinity, TDS, and pH, representing 33.7% of the variance (Fig. 2a). PC2 obtained large factor loadings (> 0.7) for turbidity, SS, total and dissolved Fe, Si, and color, representing 23.9% of the variance (Fig. 2a). PC3 had large factor loadings (> 0.7) for NO_3-N, Cl^-, and Na^+ (Fig. 2b), representing 14.9% of the total variance. PC4 had factor loadings of 0.86 for Mn and − 0.84 for DO, representing 9.5% of the variance and showing a strong inverse correlation between them (Fig. 2b).

The PCA analysis allowed to infer the key geochemical processes that regulate the chemical composition of Formoso River waters. PC1 shows the dissolution of carbonates by the solubilization of HCO_3^-, Ca^{2+}, Mg^{2+}, and Sr^{2+} that direct influence on pH values and total hardness, alkalinity, and TDS concentration (Mar da Costa et al., 2016; Morse et al., 2007). PC2 denotes the runoff and erosion processes that increase water turbidity and color by transporting particulate and colloidal materials rich in Fe-oxyhydroxides, commonly found in highly weathered soils of tropical environments. PC2 can also be influenced by oxidation and precipitation of Fe-oxyhydroxides in punctual sampled sites near riverheads, where the reduction conditions of the groundwater input high concentrations of aqueous Fe^{2+} into river waters through their springs. PC3 shows the anthropogenic influence in rivers from both urban and rural areas, which influences NO_3-N, Na^+, and Cl^- concentrations in water. High concentrations of these parameters are not linked to local geology and are commonly associated with urban and rural wastewater releases and runoff.

The geochemical indices (Table 2) were applied to evaluate the intensity of each process throughout

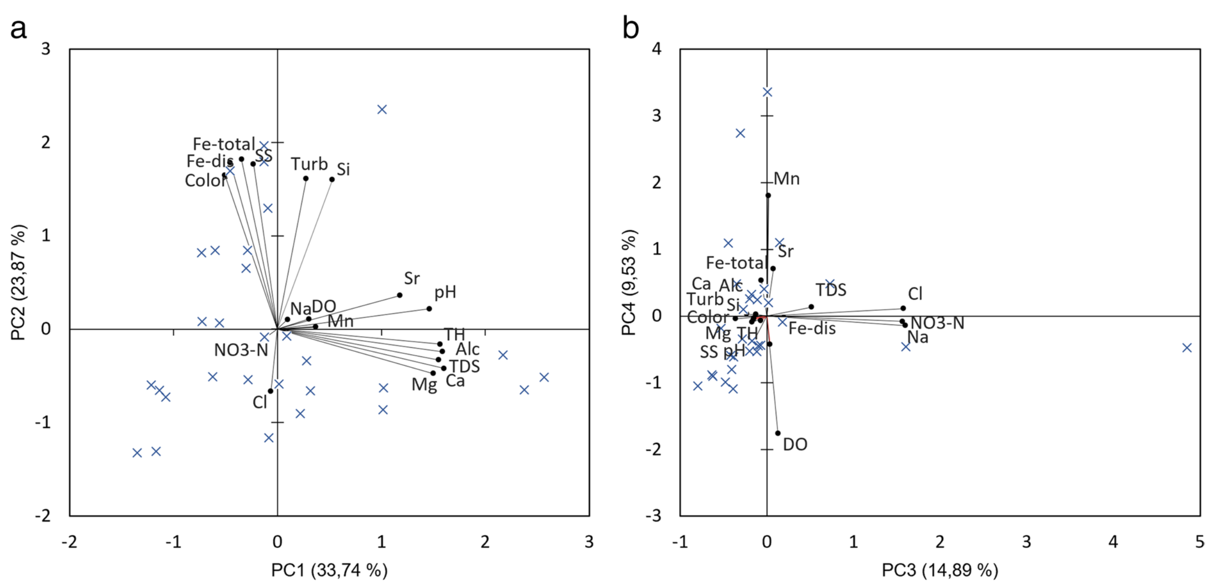

Fig. 2 Biplots calculated by the principal component analysis (PCA), showing the correlations between water quality parameters linked to different natural and anthropogenic sources and processes, i.e., carbonate dissolution (PC1), runoff and erosion (PC2), nutrient input (PC3), and redox processes (PC4), and their effects on spatial of data variability

Table 2 Indices calculated by the multivariate linear regression as function of the standardized and non-standardized values of water quality parameter

Process	Index	R^2	adj-R^2	*p-value*
Carbonate dissolution	$I_1 = [TH] \times 0.015 + [TDS] \times 0.008 + [pH] \times 0.490 - 4.329$ $Z_1 = [TH] \times 0.512 + [TDS] \times 0.275 + [pH] \times 0.256$	0.973	0.970	< 0.01
Runoff/erosion or Fe precipitation	$I_2 = [Turb.] \times 0.039 + [Color] \times 0.010 + [Fe\text{-}T.] \times 0.787 - 1.798$ $Z_2 = [Turb.] \times 0.394 + [Color] \times 0.204 + [Fe\text{-}T.] \times 0.515$	0.913	0.903	< 0.01
Nutrient input	$I_3 = [Cl^-] \times 0.039 + [NO_3\text{-}N] \times 1.845 + [Na^+] \times 0.134 - 0.855$ $Z_3 = [Cl^-] \times 0.248 + [NO_3\text{-}N] \times 0.404 + [Na^+] \times 0.434$	0.986	0.985	< 0.01
Mn reduction due to DO depletion	$I_4 = [Mn] \times 10.202 - [DO] \times 0.618 + 2.748$ $Z_4 = [Mn] \times 0.597 - [DO] \times 0.555$	0.978	0.976	< 0.01

The values between the brackets corresponds to the standardized or non-standardized values found for each water quality parameter

I_x indices calculated used non-standardized values, Z_x indices calculated using standardized values, *TDS* total of dissolved solids, *TH* total hardness, *DO* dissolved oxygen, *Turb.* turbidity

the basin (Fig. 3). The new linear combinations generated by the multivariate linear regression were calculated using standardized and non-standardized values of the water quality parameters. The new linear combinations generated by the regressions significantly predicted the factor scores as showed by p-value < 0.01, the coefficient of determination (R^2), and the adjusted-R^2 (adj-R^2; Table 2). The indices Z_x were calculated using the standardized values of the variables and show the relative contribution of each water quality parameter to their total variance, i.e., the best indicators of each geochemical processes controlling water chemistry.

The indices were assessed based on their deviation from the mean values ($Z_x = 0$), since they are standardized. The intensity of each process was assessed as negligible ($Z_x \leq 1$), moderate ($1 < Z_x \leq 2$), high ($2 < Z_x \leq 3$), and extreme ($Z_x > 3$). The index values calculated for each geochemical process are showed in Fig. 3 in a cumulative distribution plot. The index Z_1 expresses the intensity of the dissolution of carbonate rocks, demonstrating that total hardness,

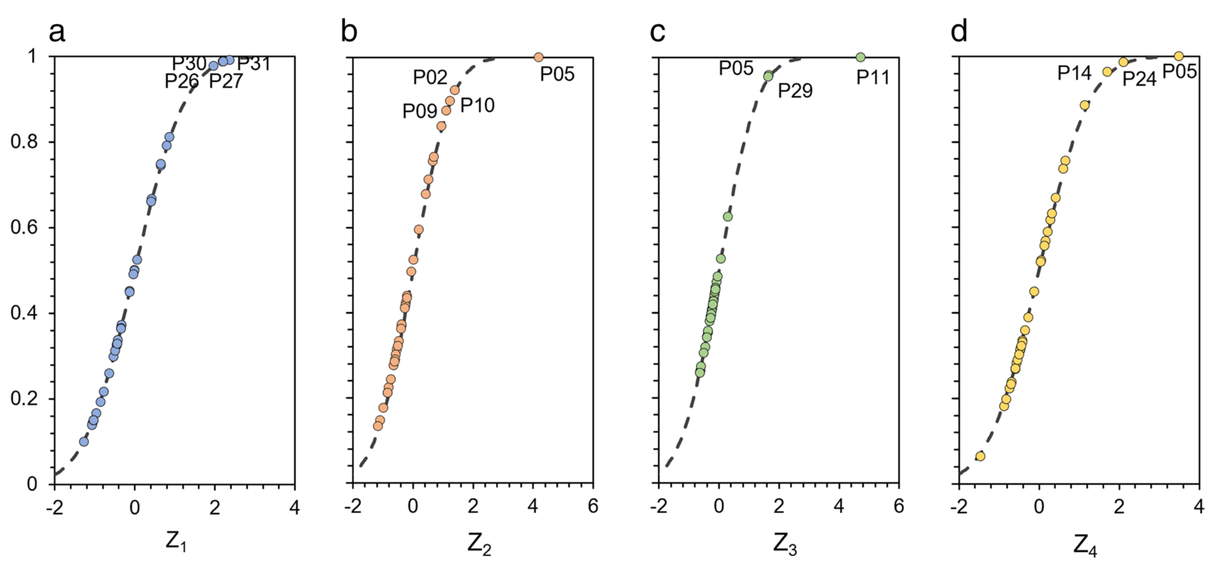

Fig. 3 Cumulative distribution plots of the indices calculated for each sampled site, showing the intensity of the different natural and anthropogenic processes, i.e., carbonate dissolution (Z_1), runoff and erosion or Fe precipitation (Z_2), nutrient input (Z_3), and redox processes (Z_4)

TDS, and pH are the best indicators of this process. The sites that showed the highest values in Z_1 were P26 and P27, located in the irrigation channels, and P30 and 31 (Fig. 3a), located over to the carbonate lithologies of the Magalhaes Formation that outcrop in the fluvial plains of the Formoso do Araguaia region. The P26 had moderate carbonate dissolution ($1 < Z_1 \leq 2$), whereas P27, P30, and P31 had high carbonate dissolution ($2 < Z_1 \leq 3$).

The index Z_2 expresses the intensity of runoff and erosion or iron oxidation and precipitation and is best represented by turbidity, color, and total Fe. The P2, P9, and P10 sites had moderate runoff ($1 < Z_2 \leq 2$; Fig. 3b) and are located mainly in the upper course of the Formoso River and its tributaries, close to the Middle Araguaia Depression, where the intensity of erosion and runoff is higher than on the Formoso-Javaés Plains and Intermediate Fluvial Plains located in the lower course of the watershed. The P5 had Fe precipitation with extreme intensity ($Z_2 > 3$; Fig. 3b) and is located close to a groundwater spring where the high input of aqueous Fe^{2+} into the river water followed by its subsequent oxidation cause the precipitation of Fe-oxyhydroxides species, which increases water turbidity and color.

The index Z_3 expresses the intensity of nutrient input due to anthropogenic activities. The water quality parameters NO_3-N, Cl^-, and Na^+ were the main indicators of this process. The sites that presented the highest values in Z_3 were P5, P11, and P29 (Fig. 3b), located in the cities of Araguaçu, Cristalândia, and Formoso do Araguaia, respectively. The sites P05 and P29 had nutrient input with moderate intensity ($1 < Z_3 \leq 2$; Fig. 3c), whereas P11 had nutrient input with extreme intensity ($Z_3 > 3$; Fig. 3c).

The index Z_4 demonstrates the influence of redox processes on the mobility of the elements, with Mn and DO as the best indicators. The P24 and P25 sites had moderate ($1 < Z_4 \leq 2$) and high intensity ($2 < Z_4 \leq 3$) of Mn reduction due to DO depletion, respectively (Fig. 3d), and are in the irrigation channels of the rice crops. When the soil is flooded for the rice crop, the microbial decomposition of organic matter can exhaust oxygen supply and start to use Mn oxides as electron acceptors. Thus, in environments with DO depletion, Mn reduction promotes their remobilization to the water column, increasing its concentrations. The P5 site had extreme intensity ($Z_4 > 3$; Fig. 3d) of Mn reduction due to DO depletion in the groundwater spring, which inputs aqueous Mn^{2+} and Fe^{2+} into river water (Online Resources).

Domestic and ecological quality evaluation

The water quality of the Formoso Basin was evaluated using two indices (OWQI and WAQI; Fig. 4). The OWQI showed that 9% of the samples were assessed as having very bad quality, 59% had bad quality, and 32% of the samples had terrible conditions (Fig. 4a). Conversely, the WAQI index showed that 9% of the samples were assessed as having terrible conditions, 18% were assessed with very bad quality, 26% had bad quality, 9% were of good quality, and 38% of the samples were assessed with excellent conditions (Fig. 4b). The contrasting water quality status found by these indices are due to the average DO concentration (5.0 ± 1.0 mg/L, 95% CI = 4.7–5.3) found by the present study and the way that the OWQI uses this parameter to evaluate the water quality status. The OWQI index considers that the ideal oxygen concentration (i.e., Sub-index = 100) should be approximately 11 mg/L. Conversely, the WAQI index recommends to use a quality guideline, which

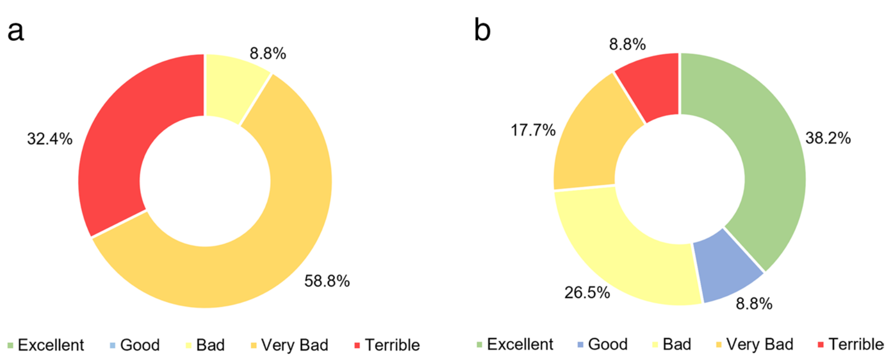

Fig. 4 Results for the Ontario Water Quality Index (OWQI) (**a**) and the Weighted Arithmetic Quality Index (WAQI) (**b**)

was 6.0 mg/L based on the Brazilian Environment Council (CONAMA) Class 2 guidelines (CONAMA, 2005). As DO concentration tend to decrease in waters with higher temperatures, tropical systems are more propense to oxygen depletion, leading to major bias in the water quality evaluation when using indices conceived in temperate environments. While our findings shows that the WAQI index can be applied to correctly define the water quality status of a tropical environment, the OWQI require further adaptation for tropical watersheds.

The water dataset and the results from the WAQI were compared to the Brazilian water quality guidelines to evaluate whether this index can correctly evaluate if the water is suitable for its multiple uses. The Brazilian water quality threshold values applied for surface waters are established CONAMA, according to water most restrictive uses (CONAMA, 2005). Among the many activities developed along the Formoso River basin, the most water restrictive uses found therein are water consumption after conventional treatment, irrigation, and recreational activities. Therefore, the waters of the Formoso Basin were accessed as class 2, in which the threshold values are reported in the Online Resources. Among the parameters analyzed, only color and dissolved Fe had noncompliance values. Color had noncompliance values in P5 and P10 sampling sites, whereas dissolved Fe had noncompliance values in P5, P1, and P10 sampling sites. All these sites were assessed as having terrible water quality conditions by the WAQI, showing that this index is suitable for predicting water quality, according to the Brazilian guidelines. Moreover, P1 and P10 sample sites had high values in I_2, whereas P5 sampled site had high value in I_3. These findings shows that soil erosion and the nutrient input from urban areas are the geochemical processes impairing the fulfillment of water multiple uses.

Irrigation water quality evaluation

The water quality for irrigation purpose was evaluated using the SAR, PI, RSC, and MH indices (Table 3). According to sodium hazard risk, SAR values greater than 9 meq/L are considered unsatisfactory for irrigation use (Nagaraju et al., 2016). The SAR values calculated for the samples collected in the Formoso River basin ranged from 0.009 to 0.532 meq/L, which are significantly lower than the threshold value and, therefore, considered suitable for irrigation. The PI is assessed in three categories, i.e., class I (PI > 75%) considered good for irrigation, class II (25% < PI < 75%) considered suitable, and class III (PI < 25%) considered unsuitable for irrigation (Nagaraju et al., 2016). The PI values of Formoso River basin water ranged from 14.6 to 69.6%. Approximately, 23% of the samples analyzed were considered unsuitable for irrigation. Irrigation water with MH values greater than 50 meq/L is deemed hazardous, and therefore, they are unsuitable for this purpose (Ali & Ali, 2018; Wakeel, 2013). The MH values found in the waters of the Formoso River basin ranged from 0 to 59.1 meq/L. Among the 34 samples, only 2 had MH higher than 50 meq/L and therefore are considered inappropriate for crop production. To be considered suitable for irrigation, the water should also have RSC values lower than 1.25 meq/L or preferentially less than 0.5 meq/L. The RSC values of Formoso River basin waters ranged from − 2.56 to − 0.12 meq/L and, therefore, are considered suitable for irrigation by this index.

Table 3 Water quality indices for irrigation purposes

Index	Mean	S.D.	C.V.	Min.	Max.
SAR	0.08	0.13	1.58	0.01	0.53
RSC	− 0.53	0.68	− 1.27	− 2.56	− 0.12
PI	37.75	15.95	0.42	14.65	69.56
MH	30.62	16.97	0.55	24.03	59.06

SAR sodium adsorption ratio, *PI* permeability index, *MH* magnesium hazard, *RSC* residual sodium carbonate

The combined evaluation of the irrigation water quality indices shows that the Formoso River basin waters are suitable for crop irrigation, although some sites were considered inappropriate. Some specific sites, mainly located at the carbonate lithologies of the Magalhaes Formation that outcrop in the fluvial plains of the Formoso do Araguaia region (P9, P14, P16, P30, and P31) and at the irrigation channels used mainly for rice crop irrigation (P25, P26, P27), were assessed as unsuitable for irrigation according to the PI. These regions can, therefore, deal with carbonate precipitation within soil profiles affecting water infiltration rates and crop production. Two sites (P23 and P26) located at irrigation channels used mainly for rice crop irrigation had MH indices higher than the threshold values, suggesting that these regions may deal with a Mg-induced Ca deficiency in plant, which can decrease crop production.

Conclusion

The present study showed that multivariate statistical methods combined with water quality indices constitute a holistic approach to evaluate the water quality condition and the effect of land uses changes in the geochemical processes controlling water chemistry. The quality indices provide information about the overall water quality condition of the basin based on quality criteria (e.g., logarithmic quality models or quality guidelines). In contrast, statistical approaches can extract information about the sources and geochemical processes intrinsic to the watershed. Additionally, the irrigation quality indices allow to identify regions that might deal with carbonate precipitation within soil profiles and Mg-induced Ca deficiency in plant, which together can affect crop production. The study also showed that the OWQI is not suitable to evaluate the water quality of tropical rivers, since the usual water higher temperature and the associated lower oxygens contents usually found in tropical environments result in biased water quality evaluation by this index.

Acknowledgements The authors thanks Timothy Mulholland for grammar check of the manuscript and Sebastião Tavares dos Reis for supporting field activities.

Author contributions Débora Francisca Moraes da Silva and Daniel Santos Mulholland contributed to the study's conception and design. Material preparation, sample collection, and field measurements were performed by Daniel Santos Mulholland., Débora Francisca Moraes da Silva and Letícia Mariana Lopes da Silva. Sample analyses were performed by Jeremie Garnie, Daniel Santos Mulholland, Débora Francisca Moraes da Silva, Daniel Ferreira Araújo and Letícia Mariana Lopes da Silva. All authors worked on data interpretation and discussion. The first draft of the manuscript was written by Débora Francisca Moraes da Silva and Daniel Santos Mulholland. All authors commented on and reviewed the previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by Universidade Federal do Tocantins (Edital Universal 30/2020 PROPESQ/UFT).

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests The authors declare no competing interests.

References

Ali, S. A., & Ali, U. (2018). Hydrochemical characteristics and spatial analysis of groundwater quality in parts of Bundelkhand Massif, India. *Applied Water Science 2018, 8*, 1–15. https://doi.org/10.1007/S13201-018-0678-X

APHA. (2005). Standard methods for the examination of water and wastewater, 23rd edn. American Public Health Association.

Bordalo, A. A., Teixeira, R., & Wiebe, W. J. (2006). A water quality index applied to an international shared river basin: The case of the Douro River. *Environmental Management, 38*, 910–920. https://doi.org/10.1007/S00267-004-0037-6/FIGURES/7

Brown, R. M., McClelland, N. I., Deininger, R. A., & Tozer, R. G. (1970). A water quality index - Do we dare. *Water Sew Works, 117*.

Chebet, E. B., Kibet, J. K., & Mbui, D. (2020). The assessment of water quality in river Molo water basin, Kenya. *Applied Water Science, 10*. https://doi.org/10.1007/s13201-020-1173-8

CONAMA. (2005). Conselho Nacional do Meio Ambiente (CONAMA). Resolução no. 357, de 17 de março de 2005. Dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu enquadramento, bem como estabelece as condições e padrões de lançamento de efluentes.

Cude, C. G. (2001). Oregon water quality index a tool for evaluating water quality management effectiveness. *Journal of the American Water Resources Association, 37*.

dos Santos D. (2016). A rede de drenagem e seu significado geomorfológico: anomalias de drenagens e tectônica recente na bacia do rio formoso, tocantins. Universidade Federal do Tocantins.

Elemile, O. O., Ibitogbe, E. M., Folorunso, O. P., et al. (2021). Principal component analysis of groundwater sources pollution in Omu-Aran Community, Nigeria. *Environmental Earth Sciences, 80*. https://doi.org/10.1007/s12665-021-09975-y

Fagundes, F. (2021). Análise do uso da água para irrigação na bacia hidrográfica do rio formoso . Universidade Federal do Tocantins.

Gnanachandrasamy, G., Dushiyanthan, C., Jeyavel Rajakumar, T., & Zhou, Y. (2020). Assessment of hydrogeochemical characteristics of groundwater in the lower Vellar river basin: Using Geographical Information System (GIS) and Water Quality Index (WQI). *Environment, Development and Sustainability, 22*, 759–789. https://doi.org/10.1007/s10668-018-0219-7

Helena, B., Pardo, R., Vega, M., et al. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research, 34*, 807–816. https://doi.org/10.1016/S0043-1354(99)00225-0

Horton, R. K. (1965). An index number system for rating water quality. *Journal (Water Pollution Control Federation), 37*, 300–306.

Iticescu, C., Georgescu, L. P., Murariu, G., et al. (2019). Lower Danube water quality quantified through WQI and multivariate analysis. *Water (Switzerland), 11*. https://doi.org/10.3390/w11061305

Lilienfein, J., Wilcke, W., do Carmo Lima S, et al. (2000). Nutrient concentrations in soil solution of some Brazilian oxisols under conventional and no-tillage systems in the early part of the rainy season. *Australian Journal of Soil Research, 38*, 851–866.

Mar da Costa, N. Y., Boaventura, G. R, Mulholland, D. S., et al. (2016). Biogeochemical mechanisms controlling trophic state and micropollutant concentrations in a tropical artificial lake. *Environmental Earth Sciences, 75*. https://doi.org/10.1007/s12665-016-5629-y

Morse, J. W., Arvidson, R. S., & Lüttge, A. (2007). Calcium carbonate formation and dissolution. *Chemical Reviews, 107*, 342–381.

Mulholland, D. S., Boaventura, G. R., & Araújo, D. F. (2012). Geological and anthropogenic influences on sediment metal composition in the upper Paracatu River Basin, Brazil. *Environment and Earth Science, 67*, 1307–1317. https://doi.org/10.1007/s12665-012-1574-6

Murtaza, G., Rehman, M. Z., Qadir, M., et al. (2021). High residual sodium carbonate water in the Indian subcontinent: Concerns, challenges and remediation. *International Journal of Environmental Science and Technology, 18*, 3257–3272.

Nagaraju, A., Sharifi, Z., & Balaji, E. (2016). Statistical and analytical evaluation of groundwater quality of Tirupati area, Chittoor district, Andhra Pradesh, South India. *Journal of the Geological Society of India, 88*(2 88), 222–234. https://doi.org/10.1007/S12594-016-0481-Y

Nong, X., Shao, D., Zhong, H., & Liang, J. (2020). Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Research, 178*. https://doi.org/10.1016/j.watres.2020.115781

Sánchez, E., Colmenarejo, M. F., Vicente, J., et al. (2007). Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators, 7*, 315–328. https://doi.org/10.1016/J.ECOLIND.2006.02.005

Şener, Ş, Şener, E., & Davraz, A. (2017). Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). *Science of the Total Environment, 584–585*, 131–144. https://doi.org/10.1016/j.scitotenv.2017.01.102

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling and Software, 22*, 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - A case study. *Water Research, 38*, 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Singh, K. R., Goswami, A. P., Kalamdhad, A. S., & Kumar, B. (2020). Development of irrigation water quality index incorporating information entropy. *Environment, Development and Sustainability, 22*, 3119–3132. https://doi.org/10.1007/S10668-019-00338-Z/TABLES/5

SRHMA. (2007). Plano de Bacia Hidrográfica do Rio Formoso-PBH Rio Formoso, No Estado do Tocantins.

Tripathi, M., & Singal, S. K. (2019). Use of principal component analysis for parameter selection for development of a novel water quality index: A case study of river Ganga India. *Ecological Indicators, 96*, 430–436. https://doi.org/10.1016/j.ecolind.2018.09.025

Tyagi, S., Sharma, B., Singh, P., & Dobhal, R. (2020). Water quality assessment in terms of water quality index. *American Journal of Water Resources, 1*, 34–38. https://doi.org/10.12691/ajwr-1-3-3

Varol, M. (2020a). Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution, 266*. https://doi.org/10.1016/j.envpol.2020.115417

Varol, M. (2020b). Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution, 266*, 115417. https://doi.org/10.1016/J.ENVPOL.2020.115417

Verbovšek, T. (2011). A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. *RMZ–Materials and Geoenvironment, 58*, 393–404

Wakeel, A. (2013). Potassium–sodium interactions in soil and plant under saline-sodic conditions. *Journal of Plant Nutrition and Soil Science, 176*, 344–354. https://doi.org/10.1002/JPLN.201200417