



# Use of multivariate statistical methods to analyze the monitoring of surface water quality in the Doce River basin, Minas Gerais, Brazil

Micael de Souza Fraga<sup>1</sup> · Guilherme Barbosa Reis<sup>1</sup> · Demetrius David da Silva<sup>1</sup> · Hugo Alexandre Soares Guedes<sup>2</sup> · Abrahão Alexandre Alden Elesbon<sup>3</sup>

Received: 10 April 2020 / Accepted: 17 June 2020 / Published online: 26 June 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

The objective of the present study was to evaluate the water quality data in the Minas Gerais portion of the Doce River basin in order to analyze the current monitoring network by identifying the main variables to be maintained in the network, their possible sources of pollution, and the best sampling frequency. Multivariate statistical techniques (factor analysis/principal components analysis, FA/PCA and cluster analysis, CA) complemented by the analysis of violation of the framing classes were used for this purpose. Water quality variables common to 64 monitoring sites were analyzed for the base period from 2010 to 2017. The water quality variables were analyzed considering the different monitoring campaigns: (a) partial campaigns; (b) total campaigns; and (c) monthly campaigns. It was identified from the FA/PCA results, that, when the partial campaign data were analyzed, the variables selected represent the high susceptibility that the basin presents to erosion and the release of domestic effluents in its water bodies. When the data of total campaigns were evaluated, representative variables of the contamination by heavy metals from industrial and mining activities were included. Therefore, the analysis of violation of the framing classes made possible to identify five critical variables: thermotolerant coliforms, dissolved iron, total phosphorus, and total manganese, which reinforced the results obtained in FA/PCA. Based on the results of the analyses, it was recommended to include variables associated with heavy metal contamination in the partial campaigns, prioritizing the dissolved iron and total manganese, as well as total chloride sampling only for the total campaigns. The evaluated data from the monthly campaigns, the CA showed that although the quarterly monitoring frequency is satisfactory, the monthly monitoring is more appropriate for the monitoring of water quality in the Minas Gerais portion of the Doce River basin.

**Keywords** Brazilian watershed · Doce River · Monitoring network · Sewage · Water quality standards · Water pollution

---

Responsible Editor: Xianliang Yi

---

✉ Micael de Souza Fraga  
micelfraga@gmail.com

Guilherme Barbosa Reis  
guilherme.eaa.reis@gmail.com

Demetrius David da Silva  
demetrius@ufv.br

Hugo Alexandre Soares Guedes  
hugo.guedes@ufpel.edu.br

Abrahão Alexandre Alden Elesbon  
abrahao@ifes.edu.br

<sup>1</sup> Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

<sup>2</sup> Universidade Federal de Pelotas, Pelotas, RS 96010-020, Brazil

<sup>3</sup> Instituto Federal do Espírito Santo, Colatina, ES 29700-558, Brazil

## Introduction

The water quality of a water body can be influenced by several factors, reason why it presents great variability (Fritzsons et al. 2009; Singh et al. 2009; Soares et al. 2020). In natural environments, water quality can be influenced by climatic factors, weathering of rocks, and soil erosion. In anthropogenic basins, agricultural expansion and accelerated population and industrial growth are evident (Ayorlo et al. 2013; Dupas et al. 2015; Muangthong and Shrestha 2015). According to the National Water Agency (ANA 2013a), such factors cause changes in the nutrients, sediments, toxins, heavy metals, among others, causing serious damage to human health and the aquatic ecosystem.

Knowing that water quality reflects the environmental conditions of the river basin, it is becoming increasingly necessary to diagnose and predict future impacts resulting from

certain actions. A qualitative monitoring program is the first step to establish a reliable and representative water quality database (Simeonov et al. 2003; Shrestha and Kazama 2007), allowing the detection of spatial and temporal variations in the variables in addition to supporting the management of water resources in the implementation of management instruments, such as the granting of permits for the use of water and charging and framing water bodies in classes of use (ANA 2013a).

In the Minas Gerais portion of the Doce River basin, qualitative monitoring began in 1997 through the Waters of Minas Project, under the responsibility of the Minas Gerais Water Management Institute (IGAM). Water quality is one of the main vulnerability aspects of the basin, since several determinants in the occurrence of specific and diffuse contaminations are observed, such as discharge of domestic wastewater without proper treatment, inadequate disposal of solid waste, high effluent generation, and inadequate land use (ECOPLAN-LUME 2010a).

In the current monitoring network, the Minas Gerais portion of the Doce River basin presents 65 stations in operation, performing four annual campaigns with quarterly frequency for most of the monitoring stations, being two complete and two partial ones. In the complete campaigns, carried out every 6 months, 56 water quality variables are analyzed, of which 51 are common to the group of stations. In the partial campaigns, carried out between the complete campaigns, 19 variables are analyzed in common to the set of stations and to the four monitoring campaigns. For the stations located in the main channel of the Doce River, the campaigns have monthly frequency (IGAM 2016).

In general, the monitoring campaigns analyze variables that make possible to characterize the water quality and the degree of contamination of the water bodies. However, the temporal variability and lack of proper understanding of water quality parameters due to the extensive database generated make it difficult to control water pollution (Chowdhury and Al-Zahrani 2014; Soares et al. 2020). Therefore, given the large amount of information that the monitoring campaigns have been generating and the lack of specific studies related to this subject for the Doce River basin, it is necessary to use statistical tools to analyze this database and identify the main variables that explain the variability of water quality, the main sources of pollution, and the best sampling frequency.

Among the methodologies available to interpret qualitative data sets, multivariate statistical techniques such as principal component analysis followed by factor analysis (FA/PCA) and cluster analysis (CA) have been widely used in recent years to support the management of water resources (Tanriverdi et al. 2010; Zhang et al. 2011; Ajorlo et al. 2013; Yu et al. 2013; Sabino et al. 2014; Lopes et al. 2014; Chowdhury and Al-Zahrani 2014; Rocha and Costa 2015; Mohamed et al. 2015; Finkler et al. 2015; Ji et al. 2016;

Rocha and Pereira 2016; Varekar et al. 2016; Zeinalzadeh and Rezaei 2017; Herojeet et al. 2017; Le et al. 2017; Calazans et al. 2018a, 2018b; Zhong et al. 2018). In water quality studies, FA/PCA make use of the correlation structure among multiple variables analyzed to produce a small number of new independent variables that contain most of the information in the original dataset (Olsen et al. 2012), allowing to correlate water quality variables with their possible sources of pollution and to select the most important ones for their characterization. The CA allows to identify the best sampling frequency, based on the similarity of the analyzed water quality data.

In addition to the multivariate techniques, other analysis can be used to complement such studies, for instance, the calculation of the percentage of monitored samples that are in disagreement with the standards established by law (Sabino et al. 2014; Martins et al. 2017; Oliveira et al. 2017). It is possible to understand that variables with a high degree of framing class violation may be considered indicative of deterioration of the water quality of the river basin.

In this context, from a data set containing physical, chemical, and biological characteristics of water, statistical techniques were employed with the objective of analyzing the qualitative monitoring network in the Minas Gerais portion of the Doce River basin, identifying the main variables to be maintained in the monitoring network, the possible sources of pollution, and the better sampling frequency, thus providing a guidance to the management bodies for actions of planning and management of water resources aiming to improve water quality.

## Materials and methods

### Characterization of the study area

The study was executed in the Minas Gerais state portion of the Doce River basin, which corresponds to 87% of the total area of approximately 82,427 km<sup>2</sup> (ANA 2013b). The Doce River originates in the state of Minas Gerais, in the Mantiqueira and Espinhaço Mountains, and its waters flows approximately 850 km until it reaches the Atlantic Ocean, in the city of Linhares, Espírito Santo state (ECOPLAN-LUME 2010a). The basin is part of the Southeast Atlantic River Basin and it is located between the parallels 17° 30' 00" and 21° 30' 00" S and the meridians 39° 30' 00" and 44° 00' 00" W.

In its entirety, the Doce River basin comprises 228 cities, which their territories are totally or partially inserted in the basin, 200 cities in Minas Gerais, and 28 in the state of Espírito Santo (CBH-Doce 2016a). There are 209 city offices located in the territory of the basin, with a resident population of approximately 3.6 million (IBGE 2010). In the context of water quality, these values bring consequences from the

precarious treatment of domestic sewage, one of the main problems in the basin. The negative impact on water quality is observed in some river parts of the basin, especially in the Doce River tributaries, because in its main channel, this impact is minimized by the increase in the river flow (ANA 2016).

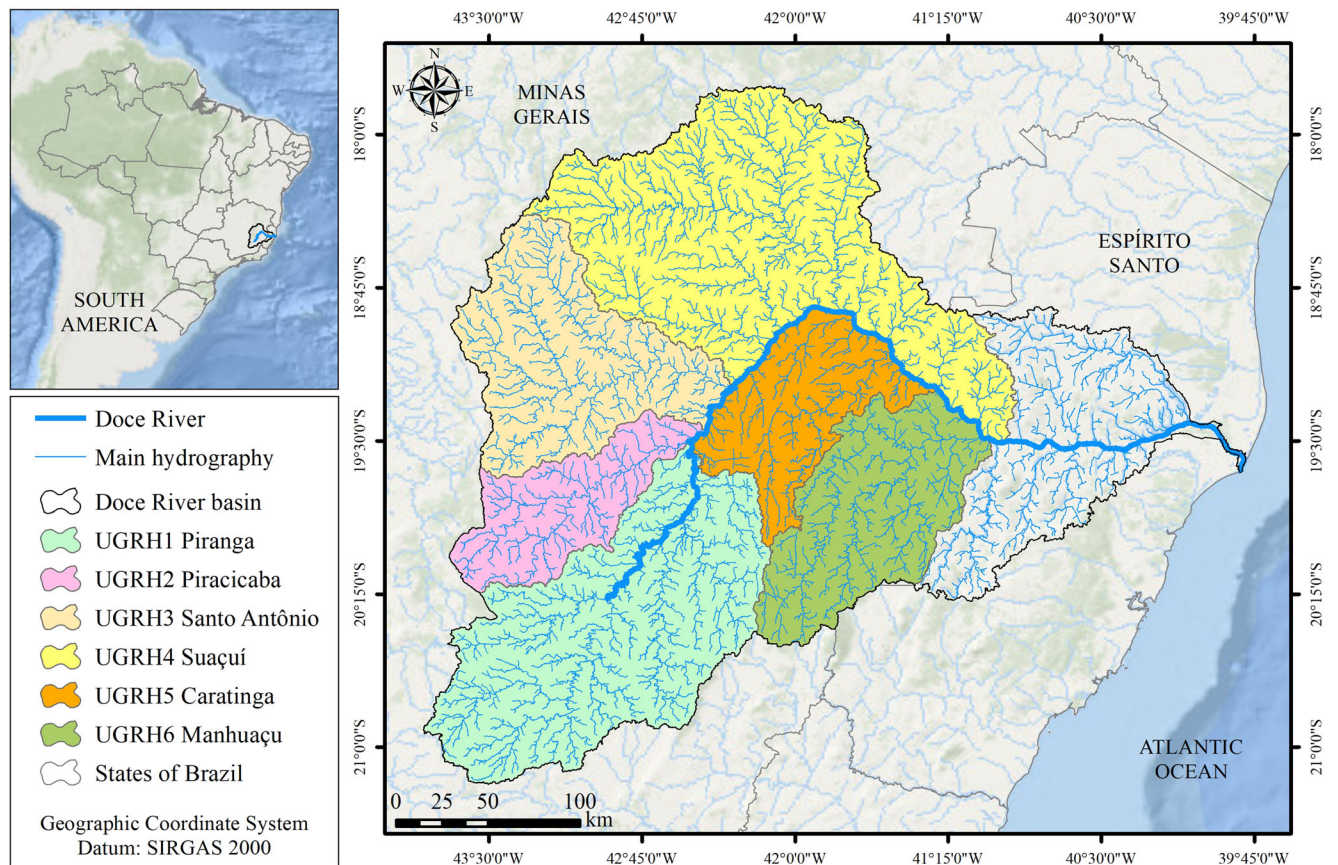
The economic activity of the basin is very diversified, specially: agriculture (reforestation, traditional crops, coffee beans, sugar cane, farming); the agribusiness (sugar and ethanol); mining (iron, gold, bauxite, precious stones, etc.); industry (pulp, steel and dairy); trading and support services of industrial complexes; and electricity generation (ECOPLAN-LUME 2010a). The region has the largest steel mill complex in Latin America, which is associated with mining and reforestation companies (CBH-Doce 2016a).

In the state of Minas Gerais, the Doce River basin is subdivided into six Water Resources Management Units (UGRHs), which correspond to UGRH1 Piranga, UGRH2 Piracicaba, UGRH3 Santo Antônio, UGRH4 Suaçuí, UGRH5 Caratinga, and UGRH6 Manhuaçu (CBH-Doce 2016b). The UGRHs are characterized by physical, socio-cultural, economic, and political aspects (IGAM 2016). In Fig. 1, it is possible to observe the separation of the Doce River basin into UGRHs in the state of Minas Gerais.

Within the economic and environmental context, the basin was the target of a major environmental crime in Brazil. On November 05, 2015, the Fundão tailings dam, operated by Samarco Mineração SA collapsed. It was located in the district of Bento Rodrigues, municipality of Mariana, state of Minas Gerais. The dam, classified as Class III, with high environmental damage potential, was destined to receive and store the waste generated by the iron ore beneficiation activity (IGAM 2017a). The dam contained 56.4 million m<sup>3</sup> of tailings, of which 43 million m<sup>3</sup> (80% of the total volume) were released into the environment. This amount reached 668 km of rivers and streams of the Doce River basin, in the states of Minas Gerais and Espírito Santo (Carmo et al. 2017), resulting in several impacts on water resources and their uses, such as public supply, irrigation, industrial use, power generation electrical, leisure and fishing, destruction of permanent preservation areas, and silting and morphological alterations of water bodies (ANA 2016).

**Database used**

The water quality data used in the study came from the water quality monitoring campaigns carried out by the “Waters of Minas Project”, where the water quality analyses are carried



**Fig. 1** Geographic location of the Doce River basin and the separation into each UGRH

out by a laboratory accredited by the National Institute of Metrology, Quality, and Technology (INMETRO), which regularly participates in analytical quality control assessments and follows standardized methods for water and sewage analysis (APHA, AWWA, WEF 2012). The monitoring results are available in the IGAM website (IGAM 2018).

For the analysis, it was chosen to use only the variables common to the set of stations, removing initially only the variable “air temperature.” Total of 50 variables (Table 1).

Although the campaigns have been carried out since 1997 and currently the network has 65 monitoring stations, the data used in the study are those from the data collections conducted in the period from 2010 to 2017 in 64 stations. Such action was taken due to the following factors: (i) there is no complete set of data that includes the majority of monitoring stations available in the pre-2010 period; (ii) multivariate analysis does not allow missing values in the dataset; and (iii) the station code RD011, located at UGRH1 Piranga, was recently implemented in 2016, thus providing a small database that made it impossible to use it in the analysis.

In order to carry out the analysis, the whole database with the 50 water quality variables was divided into three data sets: (i) partial campaigns, where the data of the 64 stations and of the 18 quality variables were analyzed, monitored, since the “air temperature” was removed from the analyses; (ii) total campaigns, where the data of the 64 stations and of the 50 water quality variables monitored were analyzed in common to the set of stations; and (iii) monthly campaigns, where the data of the 12 stations located in the riverbed of the Doce River and the 18 variables of water quality that are monitored monthly were analyzed. Those variables are monitored in the partial campaigns.

It is noteworthy that, due to the data base used, the results found were partially affected by the collapse of the Fundão tailings dam in Mariana in 2015, since the IGAM historical series include variables sensitive to the impacts resulting from the accident, such as turbidity, solids, total manganese, and dissolved iron. It is also worth noting that of the 64 IGAM monitoring stations evaluated, only 13 were affected by the collapse, among which the RD011 station, corresponded to approximately 20% of the stations evaluated.

**Table 1** Total of 50 water quality variables used in the study

Variables	
Total alkalinity (mg L <sup>-1</sup> )	Dissolved iron (mg L <sup>-1</sup> )
Dissolved aluminum (mg L <sup>-1</sup> )	Total phosphorus (mg L <sup>-1</sup> )
Total arsenic (mg L <sup>-1</sup> )	Total magnesium (mg L <sup>-1</sup> )
Total barium (mg L <sup>-1</sup> )	Total manganese (mg L <sup>-1</sup> )
Total boron (mg L <sup>-1</sup> )	Total mercury (µg L <sup>-1</sup> )
Total cadmium (mg L <sup>-1</sup> )	Total nickel (mg L <sup>-1</sup> )
Total calcium (mg L <sup>-1</sup> )	Nitrate (mg L <sup>-1</sup> )
Total lead (mg L <sup>-1</sup> )	Nitrite (mg L <sup>-1</sup> )
Free cyanide (mg L <sup>-1</sup> )	Total ammoniacal nitrogen (mg L <sup>-1</sup> )
Total chloride (mg L <sup>-1</sup> )	Organic nitrogen (mg L <sup>-1</sup> )
Chlorophyll a (µg L <sup>-1</sup> )	Oils and greases (mg L <sup>-1</sup> )
Dissolved copper (mg L <sup>-1</sup> )	Dissolved oxygen (DO) (mg L <sup>-1</sup> )
Total coliforms (MPN 100 mL <sup>-1</sup> )	pH
Electrical conductivity (µmho cm <sup>-1</sup> )	Dissolved potassium (mg L <sup>-1</sup> )
True color (mg L <sup>-1</sup> )	Total selenium (mg L <sup>-1</sup> )
Total chromium (mg L <sup>-1</sup> )	Dissolved sodium (mg L <sup>-1</sup> )
Biochemical oxygen demand (BOD) (mg L <sup>-1</sup> )	Total dissolved solids (TDS) (mg L <sup>-1</sup> )
Chemical oxygen demand (COD) (mg L <sup>-1</sup> )	Total suspended solids (TSS) (mg L <sup>-1</sup> )
Calcium hardness (mg L <sup>-1</sup> )	Total solids (mg L <sup>-1</sup> )
Magnesium hardness (mg L <sup>-1</sup> )	Surfactants (mg L <sup>-1</sup> )
Total hardness (mg L <sup>-1</sup> )	Total sulfate (mg L <sup>-1</sup> )
Thermotolerant coliforms (MPN 100 mL <sup>-1</sup> )	Sulfide (mg L <sup>-1</sup> )
Fecal streptococci (MPN 100 mL <sup>-1</sup> )	Water temperature (°C)
Total phenols (mg L <sup>-1</sup> )	Turbidity (NTU)
Pheophytin a (µg L <sup>-1</sup> )	Total zinc (mg L <sup>-1</sup> )

## Analysis methods used

The identification and selection of the determining variables of the water quality variability of the Doce River was based on the application of two analyses: principal component analysis followed by factor analysis (FA/PCA) and the analysis of violation of the limits established by the current class of framing. The analysis of the best sampling frequency was performed using cluster analysis (CA). The statistical software XLSTAT® was used to perform all the analyses.

### Factor analysis/principal component analysis

Factor analysis/principal component analysis (FA/PCA) was used to select the most significant water quality variables in the interpretation of the data set analyzed. Since FA/PCA does not allow missing values in the data set, the percentage of missing data for each water quality variable was calculated, parameters with more than 10% of missing data were disregarded (Calazans et al. 2018a). For the other missing data, it was considered the mean value of the variable (Olsen et al. 2012).

The FA/PCA was held in two rounds. The first one used only the water quality variables monitored in the partial campaigns, since they have a quarterly frequency and, consequently, a better data representativeness. The second one was performed using the water quality variables monitored in the complete campaigns, which, although they are biannual, they include a greater number of variables.

The Kaiser-Meyer-Olkin (KMO) and Bartlett’s sphericity tests were first performed to confirm the adequacy of FA/PCA to the water quality data. The KMO test verifies the correlation measure between the independent variables. The value of the test varies from 0 to 1, whereas values below 0.5 indicate that the application of FA/PCA is inappropriate. The Bartlett’s test of sphericity evaluates whether the correlation matrix is an identity matrix, which would indicate that there is no correlation between the data and that the factorial model is inappropriate (Muangthong and Shrestha 2015; Jung et al. 2016). FA/PCA was performed by decomposing the correlation matrix into its eigenvalues and eigenvectors. The Spearman’s correlation (Spearman *R* coefficient) was used due to considering the non-normal distribution of data from the water quality variables (Sabino et al. 2014; Winter et al. 2016), checked by application of the normal Shapiro-Wilk test at a significance level of 5%.

In the FA/PCA, PCA provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information (Helena et al. 2000; Shrestha and Kazama 2007). The principal component (PC) can be expressed as:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj} \quad (1)$$

where “*z*” is the component score, “*a*” is the component loading, “*x*” the measured value of variable, “*i*” is the component number, “*j*” the sample number, and “*m*” the total number of variables.

FA follows PCA. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules, and constructing new variables, also called factors (*F*). PCs were subjected to *varimax* rotation generating *F*s, which is often used in water quality studies (Zhang et al. 2011; Guedes et al. 2012; Ajorlo et al. 2013; Rocha et al. 2014; Mohamed et al. 2015; Barakat et al. 2016; Villas-Boas et al. 2017). The final effect of rotating the factorial matrix is to redistribute the variance of the first factors to the latter, with the objective of achieving a simpler and theoretically more significant factorial pattern (Hair Jr. et al. 2009). The FA can be expressed as:

$$z_{ij} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \dots + a_{fm}f_{mi} + e_{fi} \quad (2)$$

where “*z*” is the measured variable, “*a*” is the factor loading, “*f*” is the factor score, “*e*” the residual term accounting for errors or other source of variation, “*i*” the sample number, and “*m*” the total number of factors.

In FA/PCA, the factors are extracted in the order of the most explanatory to the least explanatory, and the number of factors is always equal to the number of variables. However, only those with an eigenvalue greater than 1 were considered, so any factor explains a higher variance in comparison with that one presented by a simple variable (Hair Jr. et al. 2009). On the selection of variables to characterize the factors, it was adopted the classification of values of the factor loading proposed by Liu et al. (2003): strong (> 0.75), moderate (< 0.75 and > 0.50), and weak (< 0.50 and > 0.30). It was decided to select the factors that presented factorial load ≥ 0.7, a value widely used by other authors (Chowdhury and Al-Zahrani 2014; Rocha and Pereira 2016; Shrestha and Kazama 2007).

### Framing class violation analysis

In addition to the FA/PCA, the percentage of framing class violation was also calculated for the water quality variables that have concentration limits established by the COPAM/CERH-MG Normative Resolution No. 01/2008 (Minas Gerais 2008), considering the water body framing in the location of the monitoring stations. In the Minas Gerais portion of the Doce River basin, only the Piracicaba River Basin has a framing approved by the State Council for Water Resources (CERH-MG), so the Class 2 framing was adopted for the other water bodies as approved by CNRH Resolution n° 91/2008.

In the selection of variables to be prioritized, it was used a percentage of violation of the framing class equal to or greater

than 20%. According to the Integrated Water Resources Plan of the Doce River basin (PIRH-Doce), the variables that are above this percentage are indicative of deterioration of water quality, being essential to maintain them in the monitoring program (ECOPLAN-LUME 2010a).

In the analysis, it was chosen to use all the variables monitored, and only one previous analysis was performed to filter those that have their limits established in the legislation, since this is the only necessary requirement for the application of the method.

### Cluster analysis

Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess (Shrestha and Kazama 2007). In this study, cluster analysis (CA) was used to evaluate data from the monthly campaigns conducted only in the Doce River riverbed, aiming to gather the 12 months of the year into groups (clusters) according to the similarities of the water quality variables, so that the months within a group are similar to each other but different from other groups. In the analysis, the hierarchical grouping was applied through the Ward method in the normalized data set, using the Euclidean distance as a measure of dissimilarity (bond length), as also used in several other studies (Zhang et al. 2011; Ajourlo et al. 2013; Muangthong and Shrestha 2015). With the result of the CA, it was possible to verify in which months the water quality presents a greater similarity, and therefore, to evaluate the monthly frequency sampling adopted by the IGAM and to compare it with the quarterly frequency adopted for the partial campaigns.

## Results and discussion

### Analysis of water quality through FA/PCA

In the evaluation of data suitability to the FA/PCA, the existence of significant correlations between the variables for both datasets ( $p$  value  $< 0.05$ ) was verified with the Bartlett's sphericity test. Regarding to the KMO test, the value found was 0.74 when the data of the partial campaigns were evaluated and 0.85 when the data of the complete campaigns were evaluated, demonstrating correlation of the variables. Due to the results obtained in both tests, it was verified the adequacy of the FA/PCA application to the data set.

When extracting the factors of water quality variables from the partial campaigns, 18 factors were found, six of them with an eigenvalue greater than one, explaining 71% of the total variability of the data. Fig. 2 shows the eigenvalues in descending order and the cumulative variance among the obtained factors.

Table 2 shows the non-rotational factorial weight matrix for the water quality variables of the partial campaigns. The modulus values of factor loading  $\geq 0.7$  suggest which are the most significant variables in each factor.

Based on the factor weight matrix (Table 2), it can be observed that only factors F1 and F6 presented factor loading greater than or equal to 0.7, while the others presented loadings close to this value, as well as similar values in more than one factor, such as the total chloride and electrical conductivity variables, thus hindering the analysis of the results. In this way, it was reasonable to rotate the factors, since the process maximizes their variance without affecting the proportion of the total variance explained by the set (Hair Jr. et al. 2009).

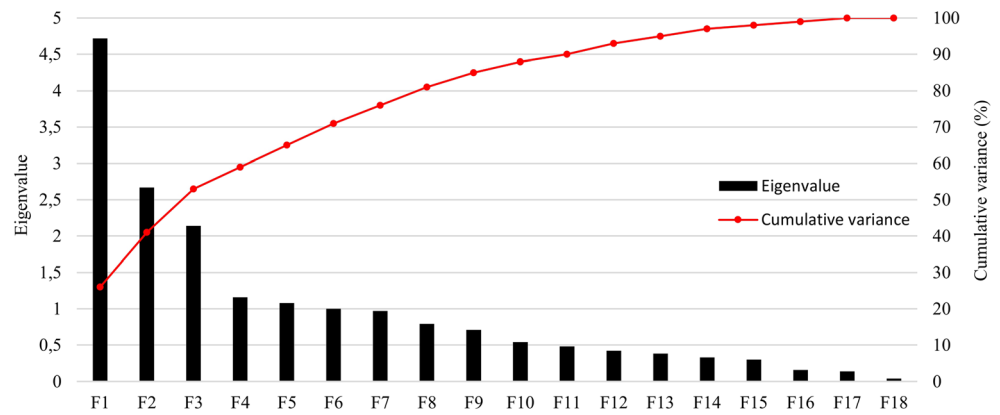
Table 3 shows the contribution of each component after the redistribution of the total variance among the factors by applying the *varimax* algorithm, without changing the total variance explained. Considering weights greater than or equal to 0.7 as indicative factors of strong loading among water quality parameters, 12 variables were selected.

Table 3 shows that the rotation of the factors provided significant improvements in the results, since parameters that did not present a high factor loading in some of the factors in the non-rotated matrix (Table 2), started to show some after the *varimax* rotation. Another positive aspect of the rotation was the better distribution of the factor loading between the factors, and therefore, each variable had a greater numerical value in only one factor, facilitating the interpretation of the result and the identification of possible sources of pollution in each of them. Improvements from the rotation process have also been observed by other studies on water quality (Guedes et al. 2012; Lopes et al. 2014; Rocha and Pereira 2016; Villas-Boas et al. 2017).

The first factor (F1) was responsible for 20% of the total variance of the data. When analyzing the factor loadings of the variables, it can be interpreted that the high values of total suspended solids (TSS), total solids, and turbidity represent the high susceptibility that the basin presents to erosion. According to the PIRH-Doce, the characteristics of soils and relief lead the Doce River basin to a condition of fragility in terms of susceptibility to erosion, which is divided into four levels: very strong; strong; moderate; and low or zero, of which 58% of the total area is classified as strong and 30% as moderate (ECOPLAN-LUME 2010a). Also, according to the PIRH-Doce, the most problematic areas in the Doce River basin are the high stream of the Piracicaba River and the Suaçuí Grande River basin.

As for the Piracicaba River basin, it is observed that the elevated portions of the unit produce the largest amount of sediment with values varying between 100,000 and 200,000 kg km<sup>-2</sup> year<sup>-1</sup>. From the confluence of the Piracicaba River with the Doce River, the production decreases to 50,000 kg km<sup>-2</sup> year<sup>-1</sup>. Among the aggravating factors of high sediment generation rates are the torrential

**Fig. 2** Eigenvalues and percentage of the cumulative variance of the factors when analyzing the water quality variables of the partial campaigns



rains, susceptible soils, and the land use in the basin, which has about 60% of anthropogenic areas (ECOPLAN-LUME 2010b). Regarding to the Suaçuí River basin, the values also vary between 100,000 and 200,000 kg km<sup>-2</sup> year<sup>-1</sup>. The extensive areas of the basin occupied by animal husbandry and mining (ECOPLAN-LUME 2010c) collaborate with the erosive process.

The F2 was responsible for 16% of the total variability of the data, inferring about the inorganic material dissolved in the water through the variables total chloride and electrical conductivity. According to Barakat et al. (2016), these variables may reflect the natural conditions of the basin through the weathering of rocks and consequent surface runoff. In

addition to their natural origin, total chloride levels may also be related to releases of industrial and domestic effluents (Ramesh kumar and Anbazhagan 2018; Rocha and Pereira 2016). F2 also showed that the pH and TDS, although not selected, presented moderate factor loading, since the electrical conductivity is positively correlated with the dissolved solids (R = 0.73), a result that corroborates with several other studies presented in the literature (Zhang et al. 2011; Frančišković-Bilinski et al. 2013; Muangthong and Shrestha 2015; Barakat et al. 2016; Pavlidis et al. 2018).

The F3, accounting for 14% of the total variability of the data, is represented by total ammoniacal nitrogen and BOD, indicating that the water bodies of the basin suffer variation

**Table 2** Matrix of non-rotational factor weight of water quality variables analyzed in partial campaigns

Variables	F1	F2	F3	F4	F5	F6
Total chloride	0.50	0.53	0.39	0.07	0.07	0.04
Chlorophyll a	0.10	0.04	0.17	0.69	-0.28	0.27
Total coliforms	0.39	0.05	-0.45	0.47	0.30	-0.14
Electrical conductivity	0.56	0.62	0.33	-0.04	0.17	0.04
BOD	0.53	0.36	-0.44	-0.09	-0.08	0.10
COD	0.76	-0.14	-0.04	-0.06	-0.08	0.05
Thermotolerant coliforms	0.46	0.13	-0.54	0.41	0.22	-0.10
Pheophytin a	0.07	-0.06	0.00	0.12	-0.04	-0.86
Total phosphorus	0.72	-0.01	-0.21	-0.13	-0.01	0.06
Nitrate	0.18	0.19	0.24	-0.29	0.61	-0.13
Total ammoniacal nitrogen	0.46	0.34	-0.48	-0.20	-0.12	0.15
Dissolved oxygen	-0.44	-0.30	0.39	0.31	0.39	0.16
pH	0.17	0.38	0.56	0.14	-0.01	-0.01
TDS	0.76	0.14	0.38	0.03	0.07	0.03
TSS	0.66	-0.69	0.15	0.00	-0.01	0.00
Total solids	0.71	-0.64	0.18	0.00	0.00	0.01
Water temperature	0.21	0.31	0.36	0.04	-0.53	-0.26
Turbidity	0.61	-0.67	0.16	-0.09	-0.04	0.01
Eigenvalue	4.72	2.67	2.14	1.16	1.08	1.00
Variance explained	0.26	0.15	0.12	0.06	0.06	0.06
Cumulative variance	0.26	0.41	0.53	0.59	0.65	0.71

Highlighted values in italics refer to factor loadings equal or greater, in modulus, than 0.7

**Table 3** Rotated factor weight matrix of water quality variables analyzed in partial campaigns

Variables	F1	F2	F3	F4	F5	F6
Total chloride	0.04	<i>0.81</i>	0.18	0.07	-0.02	-0.02
Chlorophyll a	0.04	0.25	-0.19	0.22	<i>0.71</i>	-0.13
Total coliforms	0.10	0.01	0.15	<i>0.81</i>	0.03	0.09
Electrical conductivity	0.02	<i>0.85</i>	0.28	0.10	-0.16	-0.05
BOD	0.05	0.15	<i>0.71</i>	0.30	-0.01	-0.10
COD	0.62	0.21	0.39	0.15	0.02	-0.01
Thermotolerant coliforms	0.09	0.02	0.32	<i>0.79</i>	0.06	0.05
Pheophytin a	0.05	-0.02	-0.06	0.16	-0.09	<i>0.86</i>
Total phosphorus	0.48	0.16	0.52	0.22	-0.08	-0.05
Nitrate	0.03	0.37	-0.10	0.08	-0.66	-0.03
Total ammoniacal nitrogen	0.01	0.07	<i>0.75</i>	0.20	-0.04	-0.14
Dissolved oxygen	-0.04	-0.04	- <i>0.81</i>	0.03	-0.02	-0.22
pH	-0.05	0.68	-0.14	-0.12	0.09	0.06
TDS	0.50	0.68	0.17	0.10	-0.03	0.00
TSS	<i>0.96</i>	-0.02	-0.02	0.07	0.02	0.03
Total solids	<i>0.97</i>	0.05	0.00	0.07	0.01	0.03
Water temperature	0.00	0.45	0.19	-0.31	0.34	0.41
Turbidity	<i>0.92</i>	-0.04	0.00	-0.03	-0.01	0.02
Eigenvalue	3.61	2.82	2.52	1.70	1.11	1.02
Variance explained	0.20	0.16	0.14	0.09	0.06	0.06
Cumulative variance	0.20	0.36	0.50	0.59	0.65	0.71

Highlighted values in italics refer to factor loadings equal or greater, in modulus, than 0.7

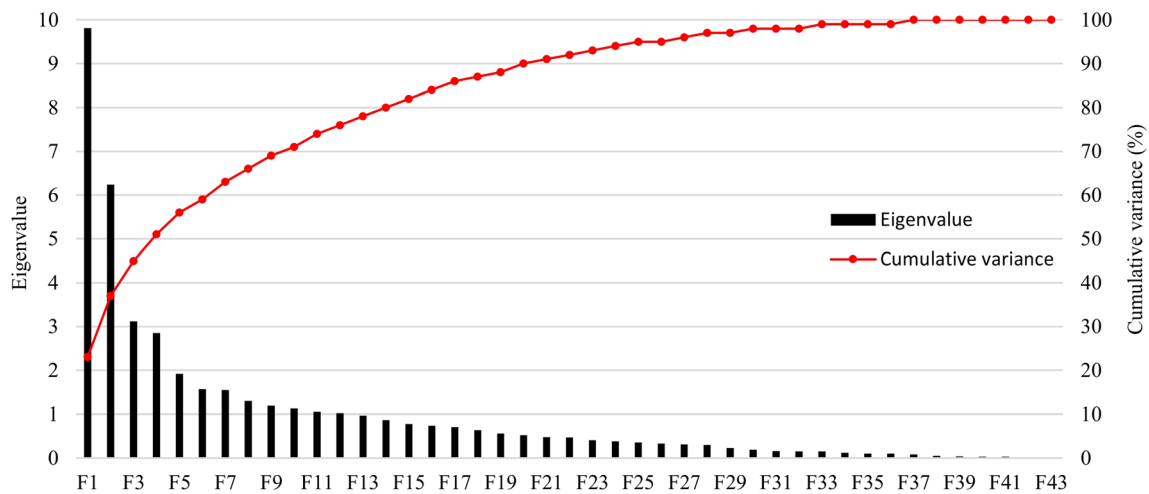
due to contamination by organic fertilizers from agricultural areas and by the discharge of untreated or partially treated domestic effluents. It can be observed in Table 3 that the variables total ammoniacal nitrogen and BOD presented a positive factor loading, while dissolved oxygen presented negative factor loading. In other words, F3 also shows the inverse relationship between dissolved oxygen and other variables, since BOD is related to the amount of oxygen required to degrade organic matter (Obade and Moore 2018). F4 is basically explained by total coliforms and thermotolerant coliforms, again indicating the precariousness of domestic sewage treatment and its release in the basin's water bodies.

According to the Water Resources Action Plan of the Piracicaba River basin (PARH-Piracicaba), coliform contamination in UGHR2 Piracicaba is above the standards in almost all the stations in the basin, demonstrating that the discharge of domestic sewage is a constant problem (ECOPLAN-LUME 2010b). As stated by the Water Resources Action Plan of the Caratinga River basin (PARH-Caratinga), in UGHR5 Caratinga, it is clear the condition of domestic effluents overload in surface waters reproduced in the non-conforming results in relation to the class 2 limit for thermotolerant coliforms (61%), total phosphorus (32%), BOD (13%), and dissolved oxygen (11%), as well as the detection of small violations of total ammoniacal nitrogen in isolated monitoring stations (ECOPLAN-LUME 2010d).

The fifth (F5) and the sixth (F6) factors are represented by chlorophyll a and pheophytin a, respectively. These variables refer to the primary productivity in the water bodies, being indicative of the physiological state of the phytoplankton and the degree of eutrophication of the aquatic environment (Giovanardi et al. 2018; Sun et al. 2018), again results from the overload of sanitary sewage without treatment and diffuse pollution from agricultural areas.

For the second round of FA/PCA, where the water quality variables monitored in the complete campaigns were considered, it was necessary to remove the total boron, dissolved copper, oils and grease, and total selenium from the calculation of the missing data percentage. The total cadmium variable was also removed because all values were equal to the minimum detection limit of the test used in the laboratory analysis ( $0.0005 \text{ mg L}^{-1}$ ), resulting in a standard deviation of zero. However, the variables calcium hardness and magnesium hardness were removed from the analysis because they essentially had the same attributes as the total calcium and total magnesium variables, respectively, and this was observed due to the high correlation between the variables ( $\sim 1.0$ ). Therefore, 43 out of 50 water quality variables monitored in the complete campaigns were analyzed. Figure 3 shows the eigenvalues in descending order and the cumulative variance among the 43 factors obtained for the 43 variables analyzed in the total campaigns.





**Fig. 3** Eigenvalues and percentage of the cumulative variance of the factors when analyzed the water quality variables monitored in the total campaigns

Considering the factors with an eigenvalue greater than one, the 43 variables analyzed were reduced to 12 uncorrelated factors, which together explain 76% of the total variance of the data (Fig. 3). In Table 4, the matrix of factor weights after the *varimax* rotation is presented. Considering weights equal to or greater than 0.7 as indicative factors of a strong factor loading, 29 variables of water quality were selected.

As can be seen in Table 4, the FA/PCA results for the total campaigns reinforce those obtained in the partial campaigns. However, because more variables were analyzed in the total campaigns, the selection of new variables considered as representative of the water quality variability in the Minas Gerais portion of the Doce River basin was conducted and, consequently, the identification of additional sources of pollution in the basin that were not considered in the analysis using data from the partial campaigns.

F1, which previously represented the high susceptibility of the basin to erosion, now also has representative variables of heavy metal pollution: total barium, total lead, total chromium, total manganese, and total nickel. Several studies have shown the relationship between heavy metals and solids, demonstrating that only a small number of them remain in the liquid mass and most of them are deposited in the sediments (Thuong et al. 2013; Malvandi 2017; Zhuang et al. 2018). Thus, sediments in the aquatic environment may play an important role in the deposition and transmission of heavy metals, justifying the fact that both have high factor loadings in F1. In the Doce River basin, heavy metals are associated with regional geology; however, their concentration in surface waters is enhanced by the releasing of domestic effluents, by the use of agrochemicals and by mining and metallurgy. They are all dominant economic activities in the Doce River basin (ECOPLAN-LUME 2010a).

A study on the water quality of the Doce River after the collapse of the iron ore tailings dam in the municipality of Mariana reinforces the strong correlation found among the

variables in F1. According to ANA (2016), when analyzing only the monitoring stations affected by the collapse of the dam, a strong correlation between the turbidity and the concentration of total suspended solids, total solids, and total manganese was verified, since these variables presented increase in the same order of magnitude. The same study showed that variables such as total lead, total chromium, total arsenic, and total mercury also had the highest maximums above acceptable limits, according to the current legislation, after the occurrence of the dam collapse. However, although they were also linked to mining, total arsenic and total mercury only showed high factor loadings in F9 and F11, respectively.

Although the disaster discharged 34 million m<sup>3</sup> of iron ore tailings in the waters of the Doce River basin, the dissolved iron variable presented a low correlation with the other variables and, consequently, a low factor loading in all 12 selected factors. In their study, ANA (2016) also found that, despite increasing the concentration after the dam collapse, the dissolved iron showed a different dynamic than the other analyzed variables. Thus, this fact may justify the low factor loading and the non-selection of the iron dissolved by FA/PCA in the present work.

The F2 continued to represent the natural conditions of the basin through the weathering of rocks and the consequent surface runoff; however, in this second analysis, we can note the addition of other variables associated with the same causes: total alkalinity, total calcium, total hardness, total magnesium, dissolved potassium, and dissolved sodium. Although the dissolved aluminum only presents a high factor loading in F4, it is also influenced by the natural conditions of the basin, since the soil of the region has in its chemical composition large concentrations of aluminum (ECOPLAN-LUME 2010a).

For the other factors, it can be stated that they represent basically the contamination by organic fertilizers originating

**Table 4** Rotated factor weight matrix of water quality variables monitored in total campaigns

Variables	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Total alkalinity	0.01	<i>0.94</i>	0.08	0.03	0.02	-0.04	0.02	0.04	0.02	-0.02	-0.03	0.04
Dissolved aluminum	0.20	0.07	-0.06	<i>0.81</i>	0.09	0.06	-0.03	-0.01	-0.08	0.00	-0.05	-0.04
Total arsenic	0.11	0.09	-0.02	-0.07	0.03	0.17	0.10	0.08	<i>0.75</i>	-0.02	0.05	-0.04
Total barium	<i>0.83</i>	0.25	-0.01	0.22	0.11	-0.11	0.06	0.03	0.04	0.04	-0.22	0.05
Total calcium	0.07	<i>0.88</i>	0.08	-0.09	0.12	-0.02	-0.03	0.01	0.16	0.02	-0.03	0.00
Total lead	<i>0.96</i>	-0.01	0.04	0.05	0.00	0.02	-0.01	0.01	0.01	-0.01	0.02	0.02
Free cyanide	-0.03	0.00	0.04	0.02	-0.04	-0.21	-0.80	0.01	0.05	0.40	-0.01	0.02
Total chloride	-0.01	<i>0.70</i>	0.13	0.26	0.03	0.20	0.01	0.12	-0.22	0.16	0.10	0.03
Chlorophyll a	0.02	0.07	-0.07	0.00	0.07	0.01	-0.02	0.10	-0.09	<i>0.83</i>	0.00	-0.01
Total coliforms	0.04	0.06	0.03	0.03	<i>0.82</i>	-0.03	0.07	0.07	0.08	-0.03	0.07	0.01
Electrical conductivity	-0.01	<i>0.91</i>	0.22	0.01	0.04	0.27	0.00	0.03	0.04	0.03	0.02	0.01
True color	0.44	0.04	0.04	0.60	0.00	-0.14	-0.39	0.20	0.02	0.09	0.08	0.11
Total chromium	<i>0.91</i>	0.02	0.02	-0.06	-0.05	0.01	-0.03	-0.01	-0.02	-0.04	-0.20	-0.01
BOD	0.00	0.16	<i>0.87</i>	-0.02	0.17	0.06	-0.06	0.00	-0.07	0.03	0.00	-0.08
COD	0.58	0.18	0.30	0.46	0.13	0.04	0.08	0.08	-0.06	0.00	0.15	0.01
Total hardness	0.11	<i>0.94</i>	0.06	-0.03	0.07	-0.07	0.01	0.00	0.16	-0.04	-0.03	0.00
Thermotolerant coliforms	0.00	0.06	0.24	-0.04	<i>0.81</i>	0.02	0.02	0.00	0.04	0.10	-0.03	-0.02
Fecal streptococci	0.15	0.02	0.19	0.15	<i>0.74</i>	0.01	-0.01	-0.05	-0.05	-0.01	-0.04	0.01
Total phenols	0.01	0.04	0.08	0.03	0.03	-0.09	<i>0.81</i>	-0.01	0.12	0.09	-0.04	0.07
Pheophytin a	0.07	0.02	0.00	-0.10	-0.05	-0.01	-0.03	0.04	-0.08	-0.02	-0.05	<i>0.84</i>
Dissolved iron	0.09	0.05	0.28	0.60	-0.02	-0.23	0.16	0.12	0.07	-0.10	-0.16	-0.11
Total phosphorus	0.39	0.11	0.59	0.20	0.10	0.11	0.10	0.04	0.20	0.09	0.28	0.02
Total magnesium	0.13	<i>0.84</i>	0.02	0.03	0.00	-0.11	0.05	-0.01	0.14	-0.09	-0.03	0.00
Total manganese	<i>0.83</i>	0.05	0.06	-0.02	0.05	0.04	0.02	0.03	0.32	-0.01	0.27	0.00
Total mercury	0.28	0.02	-0.01	-0.08	0.01	-0.05	-0.01	0.03	0.01	-0.05	<i>0.74</i>	-0.05
Total nickel	<i>0.93</i>	0.03	-0.01	0.11	0.03	0.00	-0.02	-0.01	-0.01	-0.02	-0.19	0.01
Nitrate	-0.02	0.16	0.00	-0.08	-0.06	<i>0.80</i>	0.08	0.03	0.04	-0.12	-0.04	0.05
Nitrite	0.08	0.16	0.08	0.06	0.05	0.65	-0.13	-0.10	0.47	0.03	-0.08	0.05
Total ammoniacal nitrogen	0.01	0.13	0.51	0.04	0.20	-0.03	-0.04	-0.06	0.09	-0.06	-0.09	0.18
Organic nitrogen	0.11	0.05	0.12	0.35	0.19	0.19	0.25	0.01	0.13	0.06	0.08	0.46
Dissolved oxygen	-0.05	-0.12	-0.52	-0.15	-0.20	0.07	0.23	-0.56	-0.20	0.22	0.04	-0.11
pH in loco	-0.03	0.44	-0.07	-0.25	-0.11	0.30	0.10	0.35	-0.23	-0.03	0.06	-0.14
Dissolved potassium	0.09	<i>0.72</i>	0.10	0.41	0.01	0.15	0.06	0.21	-0.24	0.03	0.12	0.03
Dissolved sodium	-0.06	<i>0.80</i>	0.09	0.01	-0.04	0.44	-0.01	0.03	-0.01	0.12	0.02	-0.01
TDS	0.45	0.61	0.08	0.45	0.11	0.11	-0.01	0.07	-0.04	0.05	0.03	0.05
TSS	<i>0.93</i>	0.02	0.01	0.11	0.07	0.00	0.01	0.01	0.01	0.00	0.22	0.03
Total solids	<i>0.92</i>	0.07	0.01	0.14	0.08	0.01	0.01	0.02	0.00	0.01	0.21	0.03
Surfactants	0.02	0.14	<i>0.89</i>	0.06	0.07	0.04	0.18	-0.03	-0.10	-0.08	0.02	-0.04
Total sulfate	0.04	0.16	0.08	0.04	0.12	0.47	0.38	-0.04	0.38	-0.09	0.03	-0.11
Sulfide	-0.06	-0.02	-0.04	-0.01	-0.05	-0.39	-0.36	-0.15	0.18	0.61	-0.10	0.01
Water temperature	-0.02	0.19	-0.08	0.15	-0.01	-0.03	-0.04	<i>0.87</i>	0.05	0.10	0.08	0.08
Turbidity	<i>0.84</i>	-0.01	0.06	0.09	0.00	0.03	0.01	0.02	0.04	0.01	0.35	0.06
Total zinc	0.34	-0.04	0.01	0.00	0.01	0.06	0.15	0.40	-0.08	0.03	-0.32	-0.15
Eigenvalue	7.65	6.64	2.88	2.48	2.17	2.13	2.01	1.54	1.49	1.41	1.27	1.08
Variance explained	0.18	0.15	0.07	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03
Cumulative variance	0.18	0.33	0.40	0.46	0.51	0.56	0.60	0.64	0.67	0.71	0.74	0.76

Highlighted values in italics refer to equal or greater factorial loads, in modulus, than 0.7

from agricultural areas and the discharge of untreated or partially treated domestic effluents in the water bodies of the basin. The following variables were selected: BOD, surface active substances, total coliforms, thermotolerant coliforms, fecal streptococci, nitrate, total phenols, water temperature, chlorophyll a, and pheophytin a.

Comparing the first and second rounds of the FA/PCA, it is observed that the difference is basically in the variability of the water quality explained by the heavy metals: total arsenic, barium, chromium, lead, manganese, mercury, and nickel. For these variables, several studies emphasize the importance of monitoring due to their bioaccumulative capacity, since they cause disturbances in the metabolic processes and damage to the biological system of living beings (Lozano et al. 2010; Riguetti et al. 2015; Zapata et al. 2017). Therefore, the importance of the second round of FA/PCA with the data from the total campaigns was demonstrated, as well as the need to include some of these variables in the partial campaigns.

### **Analysis of violations of the limits established by the COPAM/CERH-MG Normative Resolution No. 01/2008**

With the result of the analysis of violation of the framing class, it was possible to identify the variables that most represented deterioration of water quality (20% of violations or more) in the Minas Gerais portion of the Doce River basin in each UGRHs (Table 5).

As can be seen in Table 5, only the variables thermotolerant coliforms, dissolved iron, total phosphorus, and total manganese presented values of violation of the framing class higher than 20% among the UGRHs, being the maintenance of the UGRHs in the monitoring program a priority, as well as the variables indicated by the FA/PCA. Thermotolerant coliforms and total manganese were also pointed out in the second round of the FA/PCA, meaning that, in addition to presenting a high rate of violation of the framing class, they are also part of the main variables responsible for the variability of water quality in the Minas Gerais portion of the Doce River basin.

The high rates of violation of thermotolerant coliforms and total phosphorus variables characterize the release of untreated domestic effluent as the main source of pollution that affects the quality of the water of the Doce River basin, a result that has also been found in several other studies in Brazilian basins (Souza and Gastaldini 2014; Oliveira et al. 2017, 2018; Costa et al. 2017; Vargas et al. 2018; Fraga et al. 2019; Soares et al. 2020).

In a study on water quality in the Xopotó River basin, sub-basin of the Doce River, it was found that the microbiological quality of the water is deteriorating (Drumond et al. 2018). In addition to presenting high concentrations of thermotolerant coliforms, a variety of bacterial genotypes were found that represent a potential risk of diarrheagenic diseases, emphasizing the poor condition of the microbiological quality of the

water bodies of the basin, mainly due to the absence of sewage treatment plants (Drumond et al. 2018). According to ANA (2017), only 31 of the 200 municipalities in the Minas Gerais portion of the Doce River basin have some percentage of sewage treatment, and many of the effluent treatment plants are unable to remove microorganisms since they do not have a tertiary treatment processes. Despite the inadequacy of the effluent collection and treatment system, the surface water presented low levels of DO and BOD violation in all UGRHs, which can be explained by the auto depuration process, which re-establishes DO levels, but does not reduce the coliforms levels (Andrade et al. 2018).

The high percentage of class violation of dissolved iron and total manganese variables reflects the impacts of mining and releasing of steel removals. The largest steel complex in Latin America is located in the Doce River basin. The extraction of iron ore comprises the main mineral exploration activity, with approximately 20% of mining concessions in Minas Gerais. This entire industrial complex is responsible for most of Brazil's iron ore and steel exports (ECOPLAN-LUME 2010a). Total manganese is also related to mining; this metal is widely utilized in siderurgy (iron production in the manufacturing of metal alloys and batteries), textile industries (fabric paints), and other chemical industries (varnishes, fireworks, and fertilizers) (CETESB 2016).

Previous studies have shown that these variables already presented problems related to violation of the framing class. When evaluating the IGAM data from 1997 to 2008 using fewer stations, it was found that the variables thermotolerant coliforms and total manganese had the highest violation rates in all the UGRHs. For the dissolved iron variable, it was verified that the variable exceeded the established limit of 20% in the UGRHs 1, 2, 4, and 5. For the total phosphorus, this limit was exceeded only in the UGRHs 1 and 5 (ECOPLAN-LUME 2010a).

It is worth mentioning that the high violation of the total manganese and dissolved iron variables is also associated with their concentration peaks caused by the dam collapse in Mariana in 2015. For these variables, these peaks exceeded significantly the values of the historical series of data prior to the event (IGAM 2017b). It is worth pointing out that, despite the trend of return of the analyzed variables to the previous conditions, the disturbances imposed on the affected ecosystems left a significant damage in the Doce River. Much of the leaked material after the dam collapse is still deposited in the water bodies, which still potentially compromises various water uses. In addition, the large volume of tailings accumulated in the water bodies affects the balance of aquatic ecosystems, compromising fauna, flora, and ecological processes, such as auto depuration (ANA 2016). In addition to the total manganese and dissolved iron, the variables turbidity and total solids also presented peaks that exceeded the

**Table 5** Percentage of violation of the framing class in the Doce River basin considering the variables with limits established by the COPAM/CERH-MG Normative Resolution No. 01/2008

Variables	Percentage of violation (%)						
	UGRH1	UGRH2	UGRH3	UGRH4	UGRH5	UGRH6	Doce River basin
Thermotolerant coliforms	<i>65.33</i>	<i>79.72</i>	<i>42.52</i>	<i>56.74</i>	<i>77.20</i>	<i>46.77</i>	<i>63.12</i>
Total manganese	<i>39.25</i>	<i>51.36</i>	5.61	<i>23.91</i>	<i>49.77</i>	<i>13.59</i>	<i>32.90</i>
Total phosphorus	<i>35.55</i>	<i>39.31</i>	16.36	<i>24.10</i>	<i>37.01</i>	<i>21.61</i>	<i>30.38</i>
Dissolved iron	<i>26.40</i>	<i>22.99</i>	<i>33.65</i>	<i>24.75</i>	<i>36.03</i>	<i>25.57</i>	<i>27.34</i>
Dissolved aluminum	14.07	12.23	17.09	16.25	14.71	16.99	15.05
Free cyanide	14.12	14.65	14.05	14.08	16.84	14.76	14.70
Turbidity	16.23	13.33	3.27	10.44	12.99	8.39	11.86
Total suspended solids	15.57	10.57	3.27	7.78	13.64	4.52	10.19
True color	5.51	3.59	8.26	14.49	11.51	5.33	7.75
Total lead	3.84	1.93	1.81	3.44	3.74	1.75	2.85
Total arsenic	9.41	0.00	0.00	0.00	0.00	0.00	2.58
Total phenols	2.04	2.86	3.23	2.43	3.25	0.00	2.29
pH	1.64	1.84	7.94	2.09	0.32	0.65	2.04
DO	0.33	0.46	0.00	0.19	9.74	0.00	1.46
BOD	0.00	1.15	0.47	0.19	6.49	0.65	1.21
Chlorophyll a	0.50	0.00	0.00	1.54	1.63	1.95	0.92
Total chromium	0.26	0.30	0.00	2.03	0.68	1.37	0.73
Zinc	0.27	0.30	2.88	1.41	0.50	0.00	0.72
Total mercury	0.51	0.00	0.00	1.22	1.36	0.68	0.61
Dissolved copper	0.50	0.46	0.93	0.38	0.32	1.29	0.58
Surfactants	0.00	0.00	0.00	0.48	3.91	0.00	0.51
Total nickel	0.00	0.28	0.00	1.03	0.49	0.49	0.38
Total ammoniacal nitrogen	0.00	0.00	0.00	0.00	2.60	0.00	0.33
TDS	0.00	0.00	0.00	0.19	0.00	0.00	0.04
Nitrate	0.00	0.00	0.00	0.00	0.32	0.00	0.04
Total boron	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total cadmium	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total chloride	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nitrite	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total selenium	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total sulfate	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sulfide	0.00	0.00	0.00	0.00	0.00	0.00	0.00

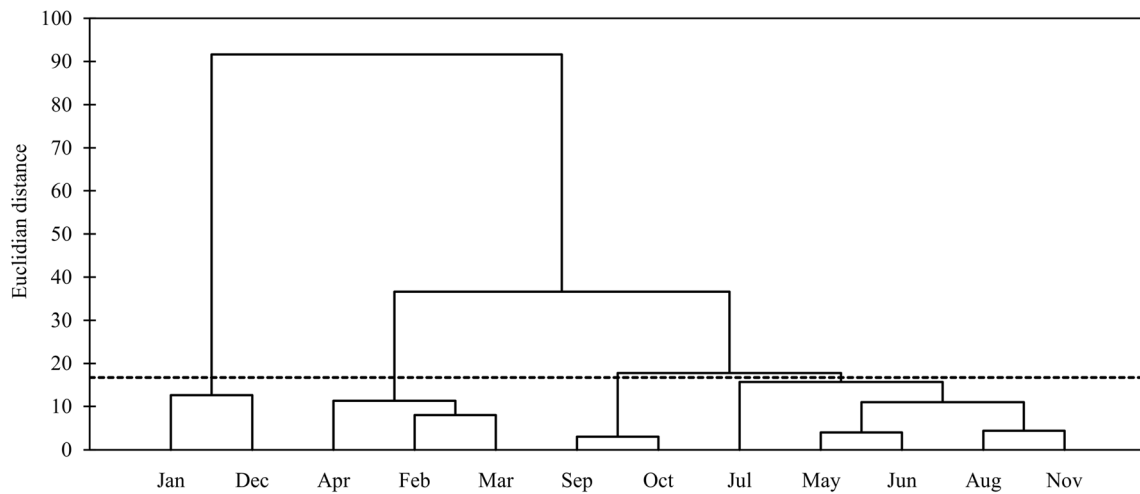
Values highlighted in italics correspond to percentages greater than 20%

maximum values of the historical series of data before the dam collapse.

Even though it presented violation values of less than 20%, the dissolved aluminum proved to be an important variable and, although it is associated to the regional geology, its transport to the surface waters is potentialized by the dominant economic activities in the basin. Violations of other heavy metals such as total arsenic, total lead, dissolved copper, total chromium, total mercury, and total nickel were also observed. On the other hand, variables that did not present violations

were observed: total boron, total cadmium, total chloride, nitrite, total selenium, total sulfate, and sulfide.

In order to prioritize the most impacting variables in the basin and reduce the costs associated with monitoring, the variables that did not present a percentage of violation may have a biannual sampling frequency, which would result in the analysis of the total chloride only in the complete campaigns. Although the total chloride had a high factor loading in F2 in the first round of the FA/PCA, this change would not pose major problems, since the electrical conductivity was



**Fig. 4** Dendrogram resulting from CA, showing the grouping of the 12 months of the year

also selected in the same factor and both represented the same pollution group.

**Monitoring frequency analysis through CA**

In the analysis of the frequency of the monthly monitoring performed only using data from the stations installed in the Doce River riverbed, the CA gathered the 12 months of the year into four groups, as it can be seen on the dendrogram in Fig. 4.

When analyzing the dendrogram (Fig. 4), you can see the influence of seasonality in the formed clusters. Clusters 1 (January and December) and 2 (April, February, and March) correspond to the rainy season, and clusters 3 (September and October) and 4 (July, May, June, August, and November) correspond to the dry season. Because they are performed quarterly, this result has a similarity with the months in which the partial campaigns are carried out, demonstrating that for the Doce River basin, the quarterly frequency can be satisfactory. On the other hand, greater Euclidean distances are observed between the months of the rainy season (groups 1 and 2), showing that for this period, the water quality does not present as much similarity, which emphasizes the importance of adopting a monthly sampling frequency. Similar results were also found by Calazans et al. (2018b) when evaluating the water quality monitoring network in the Velhas River basin, also located in the state of Minas Gerais.

Due to the high cost of monitoring campaigns, the monthly frequency can only be maintained in the riverbed of the Doce River. However, it is also recommended for the monthly frequency to perform the suggested changes for the variables monitored in the partial campaigns. Thus, the monthly and partial campaigns would

monitor the most representative variables of the water quality in the Minas Gerais portion of the Doce River basin.

**Conclusions**

A total of 14 out of 50 variables were identified as priority variables in the monitoring network: chlorophyll a, total coliforms, electrical conductivity, BOD, thermotolerant coliforms, pheophytin a, dissolved iron, total phosphorus, total manganese, total ammoniacal nitrogen, DO, total suspended solids, total solids, and turbidity. Contamination in the Doce River basin is due to a series of factors, including natural processes and anthropic activities, such as the high susceptibility of the basin to erosion; contamination by heavy metals, which are associated with the economic activities and the soils of the region; and the release of untreated or partially treated domestic effluents in the water bodies of the basin.

The high values of framing class violation for thermotolerant coliforms and total phosphorus indicate inappropriate sanitary conditions in the Doce River basin. The percentages of violation of total manganese and dissolved iron were also significant, potentialized by the economic activities of the basin.

Based on the analyses, it is recommended to include the dissolved iron and total manganese variables in the partial campaigns and the total chloride sampling only in the complete campaigns. This change would make the partial campaigns represent all sources of pollution in the Doce River basin.

The cluster analysis showed that the water quality variation of the Doce River is determined in part by the seasonality, reiterating the importance of monthly frequency monitoring in the stations of the Doce River basin.

**Funding information** This study is funded by the Brazilian Agencies CNPq (National Council for Scientific and Technological Development), CAPES (Coordination of Improvement of Personal Higher Education) and Fapemig (Foundation for Research Support of the State of Minas Gerais).

## References

- Ajorlo M, Abdullah RB, Yusoff MK, Halim RA, Hanif AHM, Willms WD, Ebrahimian M (2013) Multivariate statistical techniques for the assessment of seasonal variations in surface water quality of pasture ecosystems. *Environ Monit Assess* 185:8649–8658. <https://doi.org/10.1007/s10661-013-3201-8>
- ANA (2013a) Cuidando das Águas - Soluções para melhorar a qualidade dos recursos hídricos, 2nd edn Brasília
- ANA (2013b) Base hidrográfica Ottocodificada da bacia do rio Doce 1: 50.000/1.100.000. In: Agência Nac. Águas. <http://metadados.ana.gov.br/geonetwork/srv/pt/main.home>. Accessed 7 Jul 2018
- ANA (2016) Encarte Especial sobre a Bacia do Rio Doce - Rompimento da Barragem em Mariana/MG
- ANA (2017) Atlas Esgotos: Despoluição de Bacias Hidrográficas. Brasília
- Andrade LC de, Andrade RDR, Camargo FA d O (2018) The historical influence of tributaries on the water and sediment of Jacuí's Delta, Southern Brazil. *Ambient e Agua - An Interdiscip J Appl Sci* 13:1. <https://doi.org/10.4136/ambi-agua.2150>
- APHA, AWWA, WEF (2012) Standard Methods for the Examination of Water and Wastewater, 22nd edn
- Barakat A, El Baghdadi M, Rais J et al (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. <https://doi.org/10.1016/J.ISWCR.2016.11.002>
- Calazans GM, Pinto CC, da Costa EP, Perini AF, Oliveira SC (2018a) The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. *Environ Monit Assess* 190:491. <https://doi.org/10.1007/s10661-018-6873-2>
- Calazans GM, Pinto CC, da Costa EP, Perini AF, Oliveira SC (2018b) Using multivariate techniques as a strategy to guide optimization projects for the surface water quality network monitoring in the Velhas river basin, Brazil. *Environ Monit Assess* 190:726. <https://doi.org/10.1007/s10661-018-7099-z>
- Carmo FF do, LHY K, Junior RT et al (2017) Fundão tailings dam failures: the environment tragedy of the largest technological disaster of Brazilian mining in global context. *Perspect Ecol Conserv* 15: 145–151
- CBH-Doce (2016a) A bacia do rio Doce. <http://www.cbhdoce.org.br/institucional/a-bacia>
- CBH-Doce (2016b) Deliberação Normativa CBH-Doce nº:51/2016
- CETESB (2016) Significado Ambiental e Sanitário das Variáveis de Qualidade das Águas e dos Sedimentos e Metodologias Analíticas e de Amostragem. <https://www.google.com/search?q=cetesb&oq=cetesb&aqs=chrome..69i57j0l5j69i60l2.2421j0j7&sourceid=chrome&ie=UTF-8>.
- Chowdhury S, Al-Zahrani M (2014) Water quality change in dam reservoir and shallow aquifer: analysis on trend, seasonal variability and data reduction. *Environ Monit Assess* 186:6127–6143. <https://doi.org/10.1007/s10661-014-3844-0>
- Costa EP, Pinto CC, Soares ALC et al (2017) Evaluation of violations in water quality standards in the monitoring network of São Francisco River basin, the third largest in Brazil. *Environ Monit Assess* 189: 590. <https://doi.org/10.1007/s10661-017-6266-y>
- Drumond SN, Santiago A da F, Moreira M et al (2018) Identificação molecular de *Escherichia coli* diarrreio gênica na Bacia Hidrográfica do Rio Xopotó na região do Alto Rio Doce. *Eng Sanit e Ambient* 23: 579–590. <https://doi.org/10.1590/s1413-41522018165696>
- Dupas R, Delmas M, Dorioz J-M, Garnier J, Moatar F, Gascuel-Oudou C (2015) Assessing the impact of agricultural pressures on N and P loads and eutrophication risk. *Ecol Indic* 48:396–407. <https://doi.org/10.1016/J.ECOLIND.2014.08.007>
- ECOPLAN-LUME (2010a) Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Doce - Volume I
- ECOPLAN-LUME (2010b) Plano de Ação de Recursos Hídricos da Unidade de Planejamento e Gestão dos Recursos Hídricos - Piracicaba
- ECOPLAN-LUME (2010c) Plano de Ação de Recursos Hídricos da Unidade de Planejamento e Gestão dos Recursos Hídricos - Suaçuí
- ECOPLAN-LUME (2010d) Plano de Ação de Recursos Hídricos da Unidade de Planejamento e Gestão dos Recursos Hídricos - Caratinga
- Finkler NR, Peresin D, Cocconi J, Bortolin TA, Rech A, Schneider VE (2015) Qualidade da água superficial por meio de análise do componente principal. *Ambient e Agua - An Interdiscip J Appl Sci* 10:782–792. <https://doi.org/10.4136/ambi-agua.1468>
- Fraga M de S, da Silva DD, Elesbon AAA, Guedes HAS (2019) Methodological proposal for the allocation of water quality monitoring stations using strategic decision analysis. *Environ Monit Assess* 191:776. <https://doi.org/10.1007/s10661-019-7974-2>
- Frančičković-Bilinski S, Cuculić V, Bilinski H, Häusler H, Stadler P (2013) Geochemical and stable isotopic variability within two rivers rising under the same mountain, but belonging to two distant watersheds. *Chemie der Erde - Geochemistry* 73:293–308. <https://doi.org/10.1016/J.CHEMER.2013.02.004>
- Fritzsons E, Mantovani LE, Chaves Neto A, Hindi EC (2009) A influência das atividades mineradoras na alteração do pH e da alcalinidade em águas fluviais: o exemplo do rio Capivari, região do carste paranaense. *Eng Sanit e Ambient* 14:381–390. <https://doi.org/10.1590/S1413-41522009000300012>
- Gerais M (2008) Deliberação Normativa Conjunta COPAM/CERH-MG nº 01. 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, e dá outras providências
- Giovanardi F, Francé J, Mozetič P, Precali R (2018) Development of ecological classification criteria for the Biological Quality Element phytoplankton for Adriatic and Tyrrhenian coastal waters by means of chlorophyll a (2000/60/EC WFD). *Ecol Indic* 93:316–332. <https://doi.org/10.1016/J.ECOLIND.2018.05.015>
- Guedes HAS, da Silva DD, Elesbon AAA et al (2012) Aplicação da análise estatística multivariada no estudo da qualidade da água do Rio Pomba, MG. *Rev Bras Eng Agrícola e Ambient* 16:558–563. <https://doi.org/10.1590/S1415-43662012000500012>
- Hair JF Jr, Black WC, Babin BJ et al (2009) Análise Multivariada de Dados, 6th edn. Bookman, Porto Alegre
- 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 Res* 34:807–816. [https://doi.org/10.1016/S0043-1354\(99\)00225-0](https://doi.org/10.1016/S0043-1354(99)00225-0)
- Herojeet R, Rishi MS, Lata R, Dolma K (2017) Quality characterization and pollution source identification of surface water using multivariate statistical techniques, Nalagarh Valley, Himachal Pradesh, India. *Appl Water Sci* 7:2137–2156. <https://doi.org/10.1007/s13201-017-0600-y>
- IBGE (2010) Censo Demográfico 2010. Inst. Bras. Geogr. e Estatística, In <https://censo2010.ibge.gov.br/resultados.html>
- IGAM (2016) Qualidade das águas superficiais de Minas Gerais em 2016. Belo Horizonte
- IGAM (2017a) Encarte especial sobre a qualidade das águas do rio Doce após 2 anos do rompimento de barragem de Fundão - 2015/2017. Belo Horizonte

- IGAM (2017b) Relatório de Monitoramento das Águas Superficiais nas Bacias Hidrográficas de Minas Gerais em 2016: Projeto: Sistema de Monitoramento da Qualidade das Águas Superficiais do Estado de Minas Gerais - Águas de Minas. Belo Horizonte
- IGAM (2018) Monitoramento de Qualidade das Águas. Inst. Min. Gestão das Águas, In <http://portalinfohidro.igam.mg.gov.br/monitoramento-de-qualidade-das-aguas>.
- Ji X, Dahlgren RA, Zhang M (2016) Comparison of seven water quality assessment methods for the characterization and management of highly impaired river systems. *Environ Monit Assess* 188:1–16. <https://doi.org/10.1007/s10661-015-5016-2>
- Jung KY, Lee K-L, Im TH, Lee IJ, Kim S, Han KY, Ahn JM (2016) Evaluation of water quality for the Nakdong River watershed using multivariate analysis. *Environ Technol Innov* 5:67–82. <https://doi.org/10.1016/J.ETI.2015.12.001>
- Le TTH, Zeunert S, Lorenz M, Meon G (2017) Multivariate statistical assessment of a polluted river under nitrification inhibition in the tropics. *Environ Sci Pollut Res* 24:13845–13862. <https://doi.org/10.1007/s11356-017-8989-2>
- 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. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- Lopes FB, de Andrade EM, Meireles ACM et al (2014) Assessment of the water quality in a large reservoir in semiarid region of Brazil. *Rev Bras Eng Agrícola e Ambient* 18:437–445. <https://doi.org/10.1590/S1415-43662014000400012>
- Lozano G, Herraiz E, Hardisson A, Gutiérrez AJ, González-Weller D, Rubio C (2010) Heavy and trace metal concentrations in three rockpool shrimp species (*Palaemon elegans*, *Palaemon adspersus* and *Palaemon serratus*) from Tenerife (Canary Islands). *Environ Monit Assess* 168:451–460. <https://doi.org/10.1007/s10661-009-1126-z>
- Malvandi H (2017) Preliminary evaluation of heavy metal contamination in the Zarrin-Gol River sediments, Iran. *Mar Pollut Bull* 117:547–553. <https://doi.org/10.1016/J.MARPOLBUL.2017.02.035>
- Martins GLA, Costa ASV, Barros AR, Ramalho FMG (2017) Qualidade da água do rio Setúbal em Jenipapo de Minas-MG após construção de barragem Water quality of Setúbal river in Jenipapo de Minas-MG after dam construction. *Ambient e Agua - An Interdiscip J Appl Sci* 12. <https://doi.org/10.4136/1980-993X>
- Mohamed I, Othman F, Ibrahim AIN, Alaa-Eldin ME, Yunus RM (2015) Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia. *Environ Monit Assess* 187:4182. <https://doi.org/10.1007/s10661-014-4182-y>
- Muangthong S, Shrestha S (2015) Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. *Environ Monit Assess* 187:548. <https://doi.org/10.1007/s10661-015-4774-1>
- Obade V de P, Moore R (2018) Synthesizing water quality indicators from standardized geospatial information to remedy water security challenges: a review. *Environ Int* 119:220–231. <https://doi.org/10.1016/J.ENVINT.2018.06.026>
- Oliveira SC, Amaral RC, Almeida KC de B, Pinto CC (2017) Qualidade das águas superficiais do Médio São Francisco após a implantação dos perímetros irrigados de Gorutuba/Lagoa Grande e Jaíba. *Eng Sanit e Ambient* 22:711–721. <https://doi.org/10.1590/s1413-41522017136784>
- Oliveira DG de, Vargas RR, Saad AR et al (2018) Land use and its impacts on the water quality of the Cachoeirinha Invernada Watershed, Guarulhos (SP). *Ambient e Agua - An Interdiscip J Appl Sci* 13:1. <https://doi.org/10.4136/ambi-agua.2131>
- Olsen RL, Chappell RW, Loftis JC (2012) Water quality sample collection, data treatment and results presentation for principal components analysis – literature review and Illinois River watershed case study. *Water Res* 46:3110–3122. <https://doi.org/10.1016/J.WATRES.2012.03.028>
- Pavlidis G, Ploumistou E, Karasali H, Liapis K, Anagnostopoulos C, Charalampous A, Alexakis D, Gamvroula D, Tsihrintzis VA (2018) Evaluation of the water quality status of two surface water reservoirs in a Mediterranean island. *Environ Monit Assess* 190:570. <https://doi.org/10.1007/s10661-018-6908-8>
- Ramesh Kumar K, Anbazhagan V (2018) Analysis and assessment of heavy metals in soils around the industrial areas in Mettur, Tamilnadu, India. *Environ Monit Assess* 190:519. <https://doi.org/10.1007/s10661-018-6899-5>
- Riguetti PF, Cardoso CAL, Cavalheiro AA, Lenzi E, Fiorucci AR, da Silva MS (2015) Manganês, zinco, cádmio, chumbo, mercúrio e crômio no chorume de aterro sanitário em Dourados, MS, Brasil. *Ambient e Agua - An Interdiscip J Appl Sci* 10:153–163. <https://doi.org/10.4136/ambi-agua.1538>
- Rocha CH de B, Costa HF (2015) Variação temporal de parâmetros limnológicos em manancial de abastecimento em Juiz de Fora , MG. *Rev Bras Recur Hídricos* 20:543–550
- Rocha CHB, Pereira AM (2016) Análise multivariada para seleção de parâmetros de monitoramento em manancial de Juiz de Fora, Minas Gerais. *Ambient e Agua - An Interdiscip J Appl Sci* 11:176–187. <https://doi.org/10.4136/ambi-agua.1590>
- Rocha CHB, Freitas FA, da Silva TM (2014) Alterações em variáveis limnológicas de manancial de Juiz de Fora devido ao uso da terra. *Rev Bras Eng Agrícola e Ambient* 18:431–436. <https://doi.org/10.1590/S1415-43662014000400011>
- Sabino CVS, Lage LV, Almeida KC de B (2014) Uso de métodos estatísticos robustos na análise ambiental. *Eng Sanit e Ambient* 19:87–94. <https://doi.org/10.1590/S1413-41522014019010000588>
- 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. <https://doi.org/10.1016/j.envsoft.2006.02.001>
- Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsas D, Anthemidis A, Sofoniou M, Kouimtzis T (2003) Assessment of the surface water quality in Northern Greece. *Water Res* 37:4119–4124. [https://doi.org/10.1016/S0043-1354\(03\)00398-1](https://doi.org/10.1016/S0043-1354(03)00398-1)
- Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality—a case study. *Ecol Model* 220:888–895. <https://doi.org/10.1016/j.ecolmodel.2009.01.004>
- Soares ALC, Pinto CC, Oliveira SC (2020) Impacts of anthropogenic activities and calculation of the relative risk of violating surface water quality standards established by environmental legislation: a case study from the Piracicaba and Paraopeba river basins, Brazil. *Environ Sci Pollut Res* 27:14085–14099. <https://doi.org/10.1007/s11356-020-07647-1>
- Souza MM de, Gastaldini M do CC (2014) Avaliação da qualidade da água em bacias hidrográficas com diferentes impactos antrópicos. *Eng Sanit e Ambient* 19:263–274. <https://doi.org/10.1590/S1413-41522014019000001097>
- Sun X, Wu M, Xing Q, Song X, Zhao D, Han Q, Zhang G (2018) Spatio-temporal patterns of *Ulva* prolifera blooms and the corresponding influence on chlorophyll-a concentration in the Southern Yellow Sea, China. *Sci Total Environ* 640–641:807–820. <https://doi.org/10.1016/J.SCITOTENV.2018.05.378>
- Tanrıverdi Ç, Alp A, Demirkıran AR, Üçkardeş F (2010) Assessment of surface water quality of the Ceyhan River basin, Turkey. *Environ Monit Assess* 167:175–184. <https://doi.org/10.1007/s10661-009-1040-4>
- Thuong NT, Yoneda M, Ikegami M, Takakura M (2013) Source discrimination of heavy metals in sediment and water of To Lich River in Hanoi City using multivariate statistical approaches. *Environ Monit Assess* 185:8065–8075. <https://doi.org/10.1007/s10661-013-3155-x>

- Varekar V, Karmakar S, Jha R (2016) Seasonal rationalization of river water quality sampling locations: a comparative study of the modified Sanders and multivariate statistical approaches. *Environ Sci Pollut Res* 23:2308–2328. <https://doi.org/10.1007/s11356-015-5349-y>
- Vargas RR, Barros MDS, Saad AR, Arruda RDOM, Azevedo FD'A (2018) Assessment of the water quality and trophic state of the Ribeirão Guaraçau Watershed, Guarulhos (SP): a comparative analysis between rural and urban areas. *Ambient e Agua - An Interdiscip J Appl Sci* 13:1. <https://doi.org/10.4136/ambi-agua.2170>
- Villas-Boas MD, Olivera F, de Azevedo JPS (2017) Assessment of the water quality monitoring network of the Piabanha River experimental watersheds in Rio de Janeiro, Brazil, using autoassociative neural networks. *Environ Monit Assess* 189:439. <https://doi.org/10.1007/s10661-017-6134-9>
- Winter JCF, Gosling SD, Potter J (2016) Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol Methods* 21:273–290. <https://doi.org/10.1037/met0000079.supp>
- Yu D, Shi P, Liu Y, Xun B (2013) Detecting land use-water quality relationships from the viewpoint of ecological restoration in an urban area. *Ecol Eng* 53:205–216. <https://doi.org/10.1016/j.ecoleng.2012.12.045>
- Zapata FCC, Villanueva MC, Esquivel RAP, Payano IGU (2017) Bioaccumulation of heavy metals in *Oncorhynchus mykiss* for export at production centers in the Peruvian Central Highlands. *Ambient e Agua - An Interdiscip J Appl Sci* 12:527. <https://doi.org/10.4136/ambi-agua.2100>
- Zeinalzadeh K, Rezaei E (2017) Determining spatial and temporal changes of surface water quality using principal component analysis. *J Hydrol Reg Stud* 13:1–10. <https://doi.org/10.1016/j.ejrh.2017.07.002>
- Zhang X, Wang Q, Liu Y, Wu J, Yu M (2011) Application of multivariate statistical techniques in the assessment of water quality in the southwest new territories and Kowloon, Hong Kong. *Environ Monit Assess* 173:17–27. <https://doi.org/10.1007/s10661-010-1366-y>
- Zhong M, Zhang H, Sun X, Wang Z, Tian W, Huang H (2018) Analyzing the significant environmental factors on the spatial and temporal distribution of water quality utilizing multivariate statistical techniques: a case study in the Balihe Lake, China. *Environ Sci Pollut Res* 25:29418–29432. <https://doi.org/10.1007/s11356-018-2943-9>
- Zhuang Q, Li G, Liu Z (2018) Distribution, source and pollution level of heavy metals in river sediments from South China. *CATENA* 170:386–396. <https://doi.org/10.1016/J.CATENA.2018.06.037>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.