CrossMark

Assessment of spatial variations in the surface water quality of the Velhas River Basin, Brazil, using multivariate statistical analysis and nonparametric statistics

Carolina Cristiane Pinto ·
Giovanna Moura Calazans · **Sílvia Corrêa Oliveira**

Abstract The Velhas River sub-basin, which is located in the third-largest river basin in Brazil (São Francisco), is in an advanced state of degradation. In this work, the surface water quality of the Velhas River Basin was studied at 65 monitoring sites; 16 water quality parameters were sampled quarterly for 11 years (2008 to 2013). Cluster analysis (CA) and a nonparametric Kruskal–Wallis test were associated with the analysis of violations to water quality standards to interpret the water quality data set from the Velhas River Basin and assess its spatial variations. The CA grouped the 65 monitoring sites into four groups. The Kruskal–Wallis test identified significant differences ($p < 0.05$) between the groups formed by CA. The results show that watercourses located in the upper region of the Velhas River Basin are more affected by the release of industrial effluent and domestic sewage, and the lower region is more affected by diffuse pollution and erosion. This association between multivariate statistical techniques and nonparametric tests was effective for the classification and processing of large water quality datasets and the identification of major differences between water pollution sources in the basin. Therefore, these results provide an understanding of the factors affecting water quality in the Velhas River Basin. The results can aid in decision-making by water managers and these methods can be applied to other river basins.

Introduction

The sub-basin of the Velhas River is one of the most important areas in the São Francisco River Basin, which is the third-largest basin in Brazil in terms of territorial extension, and it is completely within Brazilian territory. However, the sub-basin of the Velhas River is the area most affected by anthropogenic pollution (domestic sewage, industrial effluent, agriculture, and livestock, among other types) in the São Francisco River Basin (Trindade et al. 2017; Costa et al. 2017).

The European Union and the USA have already established treatment facilities for point source pollution, and non-point source pollution has recently received attention (Zhang et al. 2011). However, in Brazil, the release of domestic sewage *in natura* to the waterways still occurs, even in the study region. Data from the National Sanitation Information System (SNIS), the largest and most important information system in the Brazilian sanitation sector, show that only 50.3% of the population have access to sewage collection, and only 42.67% of the collected sewage is treated (SNIS 2017).

C. C. Pinto · G. M. Calazans · S. C. Oliveira (✉)
Escola de Engenharia, Universidade Federal de Minas Gerais, Campus Pampulha, Av. Antônio Carlos, 6627 - Bloco 1 - sala 4525, Belo Horizonte, MG 31.270-901, Brazil
e-mail: silvia@desa.ufmg.br

It is also emphasized that, most of the time, only secondary sewage treatment is conducted, without the removal of pathogens and nutrients.

This sub-basin has been monitored by the state environmental agency, the Minas Gerais Water Management Institute (IGAM), since 1997. Surface water quality monitoring is vital for the sustainable management of water resources. Their results provide an understanding of the current water quality situation and the main changes over time through detecting spatial and temporal trends. In addition, such data are essential for the application of management tools, such as the planning, granting, charging, and framing of watercourses, and to subsidize the inspection, environmental licensing, and formulation of environmental policies. Therefore, their results allow the efficient management of water resources and sustainable socioeconomic development of the sub-basin (Park et al. 2006; Strobl and Robillard 2008; Finotti et al. 2009; ANA 2013).

The results of such monitoring programs are consistent across a large data matrix, and can be interpreted (Wunderlin et al. 2001; Shrestha and Kazama 2007). Therefore, evaluation methods are required to combine large data sets with many variables to generate a single and simple interpretation result, without the loss of useful information (Achieng' et al. 2017).

Multivariate statistical methods are efficient and have been widely applied to matrices of water quality monitoring data for different objectives, such as characterizing and evaluating the water quality of rivers; elucidating temporal and spatial variations and tendencies caused by natural and/or anthropogenic processes, and to determine the parameters responsible for these variations; extracting information regarding similarities and differences between seasons, parameters, and monitoring sites; identifying possible sources of pollution at different points in a body of water; and identifying and analyzing the relationships between water chemistry, land use, and geology (Vega et al. 1998; Wunderlin et al. 2001; Reghunath et al. 2002; Simeonova et al. 2003; Singh et al. 2004; Mendiguchía et al. 2004; Shrestha and Kazama 2007; Alkarkhi et al. 2008; Zhang et al. 2011; Varol et al. 2012; Bhat et al. 2014; Achieng' et al. 2017).

The purpose of this study is to evaluate a large data matrix obtained from a surface water quality monitoring program for the Velhas River Basin between 2008 and 2013. The objective is to use cluster analysis associated with the nonparametric Kruskal–Wallis test and analyze the percentage of violations to the state legislation standards to (1) identify similarities or dissimilarities between sampling sites; (2) identify water quality parameters that explain the structure of the dataset; and (3) determine the influence of natural and anthropogenic factors in the Velhas River Basin.

Materials and methods

Study area

The São Francisco River Basin is one of the most important in Brazil as it contributes a great amount of water to six states and the Federal District. The São Francisco River is the largest in Brazil, and its water basin has faced biodiversity losses and changes to its aquatic ecosystems due to the lack of sanitation services, the construction of large dams, and industrial and agricultural activities (ANA 2015). Environmental education and recovery are urgently required. The São Francisco River is the largest river (2914 km) located wholly within Brazil, and its basin contains areas from six states (Minas Gerais, Goiás, Bahia, Pernambuco, Alagoas, and Sergipe) and the Federal District, where Brasília (Brazil's capital) is located (Lee et al. 2014).

The sub-basin of the Velhas River is one of the main sub-basins in the São Francisco River Basin in terms of its water contribution and territorial extension (807 km). Its drainage area is 27,850 km², it has a population of approximately 4.8 million inhabitants, and it covers 51 municipalities (CBH Velhas 2014). Among these municipalities, 15 are located in the Metropolitan Region of Belo Horizonte (MRBH), an area of intense urbanization that, despite occupying only 10% of the basin's territorial area, contains 70% of its population, which contributes to the deterioration of its surface waters. Water uses in this sub-basin range from the dilution of domestic and industrial effluents to more important usages, such as water supply, fish farming, and primary contact recreation.

Land use in Velhas River basin

The Velhas River Basin is segmented according to the land use characteristics of each region, which provide different sources of pollution to surface water (Table 1).

The Upper Velhas is the most critical and concerning regions of the basin in terms of its surface water quality.

Table 1 Territorial regions of the Velhas River Basin, location, and main characteristics

Regions	Location	Main characteristics
Upper	Upper Velhas covers an area of 2740 km^2 (9.8% of the basin) and the entire called Quadrilátero Ferrífero region (largest national iron ore producer).	It has a larger population and significant economic activity, which are mainly concentrated in the MRBH. This region contains the integrated water supply system of the Velhas River (supplying 74% of Belo Horizonte City, as well as other cities). The main pollutants are untreated industrial and domestic sewage and the effluent of mining activities.
Middle–Upper	The drainage area corresponds to 15.4% of the basin and comprises 20 cities.	These areas have a lower population concentration than the Upper region, and agricultural and livestock activities are predominant.
Middle–Lower	This represents the largest portion of the Velhas River Basin (43.8%, 12.204 km^2), and contains the entire or partial area of 23 cities	
Lower	Composed of 8 cities and represents the second-largest group in terms of territorial area (31%).	The unique aspect of this region is that none of the cities are completely located within the basin. The population concentration is low and agricultural and livestock activities are predominant.

Source: Adapted from CBH Velhas (2014)

As it covers an extensive mining region, its surface waters receive a large amount of domestic and industrial effluent from the MRBH; however, it also includes the main sources of urban water supply.

Pasture areas occupy approximately half of the basin's area (45.6%); as such, areas and agriculture are more extensive in the central and lower regions of the Velhas River Basin. As reported by IGAM (2013b), these areas are mainly responsible for erosion in the region owing to the large percentage of mechanized activity and the use of agricultural inputs, such as fertilizers and pesticides.

The annual reports published by the Minas Gerais Water Management Institute (IGAM), who are responsible for water monitoring in the region, have suggested that the dumping of domestic sewage, mining, industrial, and agricultural activities are the main sources of pollution in the Velhas River Basin (IGAM 2013b). Due to these reasons, the aquatic environment of the basin is increasingly deteriorating.

Water quality monitoring program

In the Velhas River Basin, the surface water quality monitoring network has been managed by the Minas Gerais Water Management Institute (IGAM) since 1997. The dataset used in this study (available on the IGAM website) includes 16 water quality parameters (Table 2) that were sampled quarterly from 2008 to 2013. The locations of the 65 monitoring sites are shown in Fig. 1.

The water quality parameters sampling, transport, and analyses were carried out by a laboratory accredited to the Brazilian National Institute of Metrology, Quality and Technology (INMETRO), which participates regularly on analytical quality control (AQC) evaluations and follows the standards of Standard Methods for the Examination of Water and Wastewater (APHA, AWWA, WEF 2012).

Descriptive statistics

Univariate descriptive statistics were used prior to multivariate analysis, which involved the calculation of the minimum and maximum values, mean, standard deviation, geometric mean, and the percentiles (25, 50—median, and 75) of water quality parameters for each monitoring site. The distribution of the data was verified by applying the Shapiro–Wilk normality test (Shapiro and Wilk 1965) at a significance level of 5%. In addition, the percentage of outliers was calculated following the exploratory method of interquartile amplitude (AIQ) (Naghettini and Pinto 2007; Schwertmana et al. 2004). As suggested by Hair Jr. et al. (2009), the occurrences of outliers were analyzed individually to decide whether they should be included or removed from the analysis.

Statistical analysis

The spatial variability of surface water quality was analyzed using multivariate statistical techniques,

Table 2 Physicochemical and microbiological parameters used in this study

Parameter	Abbreviation	Units	Analytical technique
Biochemical oxygen demand	BOD	mg L^{-1} O$_2$	Titrimetric
Chemical oxygen demand	COD	mg L^{-1} O$_2$	Titrimetric
Chlorophyll-α	Cl-α	µg L^{-1}	Spectrophotometric
Dissolved oxygen	DO	mg L^{-1} O$_2$	Titrimetric
Electric conductivity	EC	µS cm^{-1}	Conductimetric
Nitrate nitrogen	N-NO$_3^-$	mg L^{-1} N	Potentiometric
pH	pH		Potentiometric
Pheophytin-a	Ph-a	µg L^{-1}	Colorimetric
Thermotolerant coliforms	T. *Coli.*	MPN[a].100 mL^{-1}	Multiple-tube fermentation
Total Chloride	Cl$^-$	mg L^{-1} Cl	Colorimetric
Total Phosphorus	P	mg L^{-1} P	Colorimetric
Total Dissolved Solids	TDS	mg L^{-1}	Gravimetric
Total Suspended Solids	TSS	mg L^{-1}	Gravimetric
Total Solids	TS	mg L^{-1}	Gravimetric
Turbidity	Turb.	NTU	Nephelometric
Water temperature	T	°C	Thermometer

[a] Most probable number

nonparametric Kruskal–Wallis test, and analyses of violations to the standards set for water quality parameters recommended by legislation. All mathematical and statistical computations were conducted using Microsoft Office Excel 2010, *Statistica* 8.0, and *XLSTAT* 2014.5.

To eliminate the influence of different measurement units and make the data dimensionless, the data were standardized through *z*-scale transformation (subtraction of the mean and division by the standard deviation of each variable), which generates variables with a mean of 0 and standard deviation of 1 and provides them with the same scale (Liu et al. 2003; Zhang et al. 2011).

Cluster analysis (CA) group observations into clusters according to their similarities or dissimilarities so that the result shows high homogeneity within a group and high heterogeneity between groups (Vega et al. 1998; Singh et al. 2005; Lattin et al. 2011). The hierarchical agglomerative method, which uses Euclidean distance as a measure of similarity, is one of the most-used CA methods and begins with individual objects; other objects are grouped stepwise considering similarity with previous objects, until all form a large group (Johnson and Wichern 2007). The results of CA are typically illustrated using a dendrogram, which represents clusters and allows the similarity between them to be visualized (Panda et al. 2006; Shrestha and Kazama 2007; Sojka et al. 2008; Phung et al. 2015). In this study, hierarchical agglomerative CA was performed on

the normalized data set following the complete linkage method, using Euclidean distance as a measure of similarity.

To identify significant differences between the medians of the water quality parameters for the groups formed by CA, a Kruskal–Wallis nonparametric test was applied, followed by a multiple comparison test (when applicable) at a significance level (α) of 5%. Therefore, the parameters responsible for differentiating the groups formed by the CA could be identified. The Kruskal–Wallis test was applied individually for each parameter.

Analysis of violation of the water quality parameters to the legal standards set

The percentage of violations to the legal standards for water quality parameters was calculated for each monitoring site, considering the previously selected period. The Minas Gerais State legislation Normative Deliberation (ND) 01/2008 (COPAM/CERH-MG 2008)-establishes more restrictive values than national legislation; therefore, for this study, the water quality data series was compared to local values to more precisely evaluate violations to water quality standards. The water quality parameters violated at a rate of 50% or more during the analyzed period were highlighted as "critical parameters."
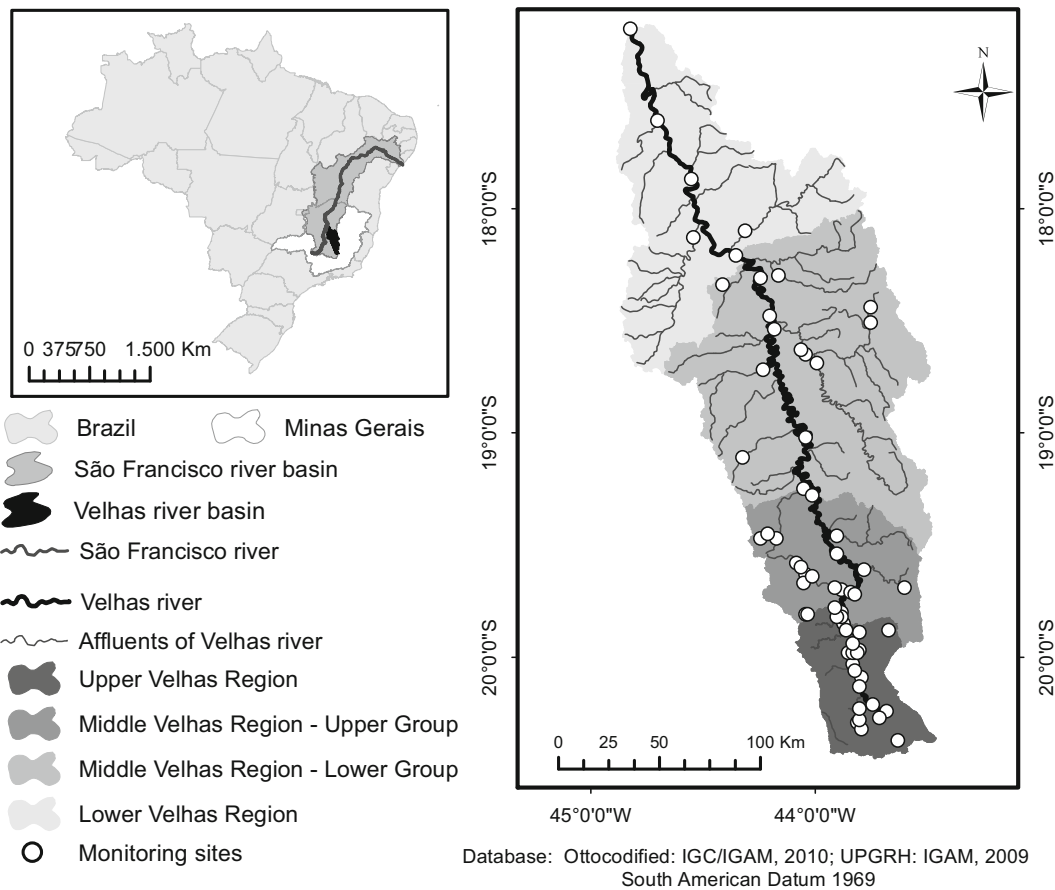
Fig. 1 Location of the study area and surface water quality monitoring sites

Results and discussion

Descriptive statistics

Table 3 presents the results of univariate descriptive statistics considering all monitoring sites. A total of 1430 observations were obtained. These results show variations in the concentrations of parameters in the whole studied hydrographic basin. The median values may suggest hypotheses about water quality in the Velhas River Basin.

From the individual descriptive statistics for each monitoring site, monitoring sites SC25, SC14, SC10, BV154, BV155, and SC26 presented the highest medians for most water quality parameters, while stations SC27 and AV250 presented the smallest values. This may suggest that the water quality at these monitoring sites is more or less degraded than average in the Velhas River Basin; therefore, they are more or less relevant in the monitoring program. For example, AV320 may be

located in an area where there are several pollution sources as its median total dissolved solids value is 3.5 times higher than the second-highest median for the parameter. These assumptions can be confirmed by subsequent analyses.

The results of the Shapiro–Wilk normality test at a significance level of 5% show that the distribution of most parameters was asymmetric, which is expected for environmental data. The probability of the occurrence of the identified outliers was evaluated individually. We included all outliers in the analyses as they are likely to occur, and cannot be considered as measurement errors. Instead, they may often be the result of hydrological variations.

Statistical analysis

CA was used to detect similar groups between the monitoring sites. A dendrogram was generated, grouping 65 of the monitoring sites in the Velhas River Basin into four statistically significant clusters at (Dlink/Dmax) × 100 <

Table 3 Summary statistics of water quality parameters in the Velhas River Basin from 2008 to 2013, considering all monitoring sites

Variables/Units	Min	Max	Mean	Standard deviation	Variance	Median
BOD (mg L^{-1} O$_2$)	2	145	6.4	11.7	136.89	2
COD (mg L^{-1} O$_2$)	5	466	24.7	29.9	894.01	15
Cl-α (μg L^{-1})	0	1134.8	22.3	75.2	5655.04	4.2
DO (mg L^{-1} O$_2$)	0.5	14.3	6.4	1.9	3.61	6.8
EC (μS cm^{-1})	1	2465	194.7	255.5	65,280.25	130.5
N-NO$_3^-$ (mg L^{-1} N)	0	13.3	0.8	1.3	1.69	0.4
pH	5.4	9.9	7	0.6	0.36	7
Ph-a (μg L^{-1})	0	95.8	4	7.1	50.41	1.8
T. *Coli.* (MPN 100 ml^{-1})	2	160.000	46,892	64,772	4,195,411,984	8000
Cl$^-$ (mg L^{-1} Cl)	0.3	168	9.3	14.9	222.01	3.8
P (mg L^{-1} P)	0	3.6	0.2	0.3	0.09	0.1
TDS (mg L^{-1})	7	2564	129.6	207.6	43,097.76	92
TSS (mg L^{-1})	0	6846	121.2	335.2	112,359.04	33
TS (mg L^{-1})	19	7044	250.8	392.1	153,742.41	152
Turb. (NTU)	0.9	4656	99.9	297.2	88,327.84	22.3
T (°C)	12.8	33.2	24	3.2	10.24	24.2

60 (equal to 40), as shown in Fig. 2. Previous studies conducted in the Velhas River Basin also presented similar results for the monitoring sites analyzed using the CA (Almeida 2013; Knupp 2007; Trindade 2013).
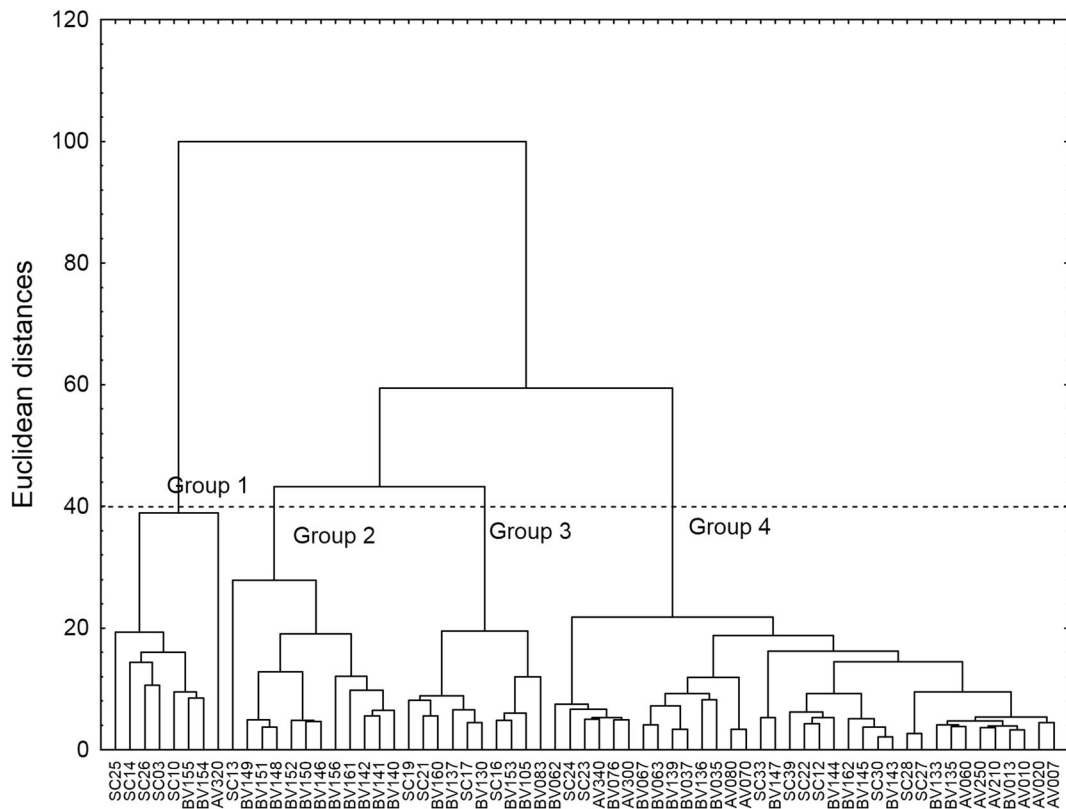

Fig. 2 Dendrogram showing the clustering of monitoring sites

Among the four formed groups, the first stands out, which was formed from previously indicated monitoring sites that have higher medians of water quality parameters. This indicates that monitoring sites SC25, SC14, SC26, SC03, SC10, BV155, BV154, and AV320 should be prioritized in the monitoring network and greater attention should be given in their location, as they represent areas of greater water quality deterioration in the water basin.

The groups formed by CA were identified in Fig. 3 to analyze their geographic locations using ArcGIS 9.1. Similar water quality is observed between the region upstream of the MRBH and in some tributaries of the middle course. There are some differences in surface water quality between the regions of Velhas River, and it should be noted that group 1, which contains the most impacted monitoring sites, are located in the Metropolitan

Region of Belo Horizonte (MRBH). This result indicates that the water quality of these monitoring sites is highly influenced by the urbanization of this area.

The nonparametric Kruskal–Wallis test exhibited differences between all parameters at a significance level of 5% ($p < 0.05$), in comparison to the groups formed by CA. The multiple comparison test was then used to identify clusters that are different. Some of these results are presented in Fig. 4.

Cluster I (monitoring sites SC25, SC14, SC26, SC03, SC10, BV155, BV154, and AV320)

In accordance with previous studies (IGAM 2013b; Knupp 2007), these monitoring sites are extremely polluted. AV320 is located in the Mina Stream, in the city of Raposos, near an iron ore beneficiation company, which
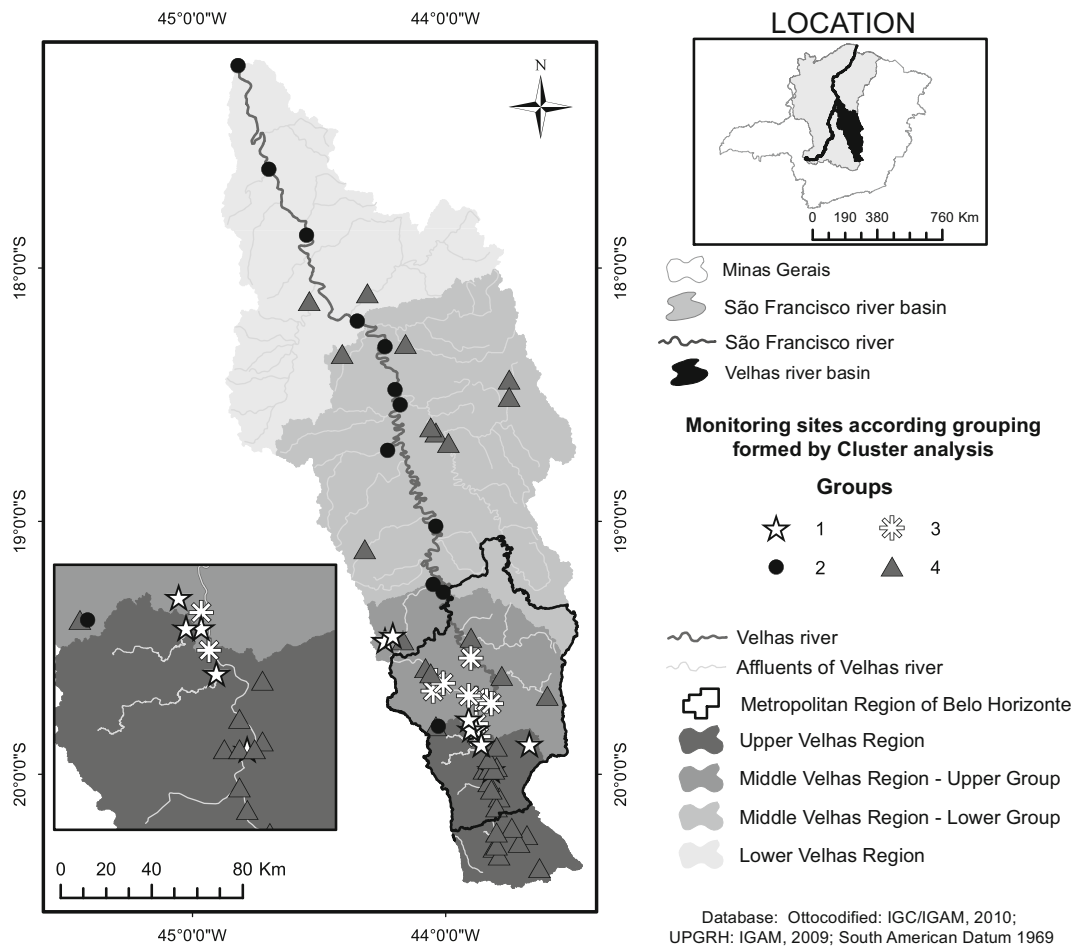

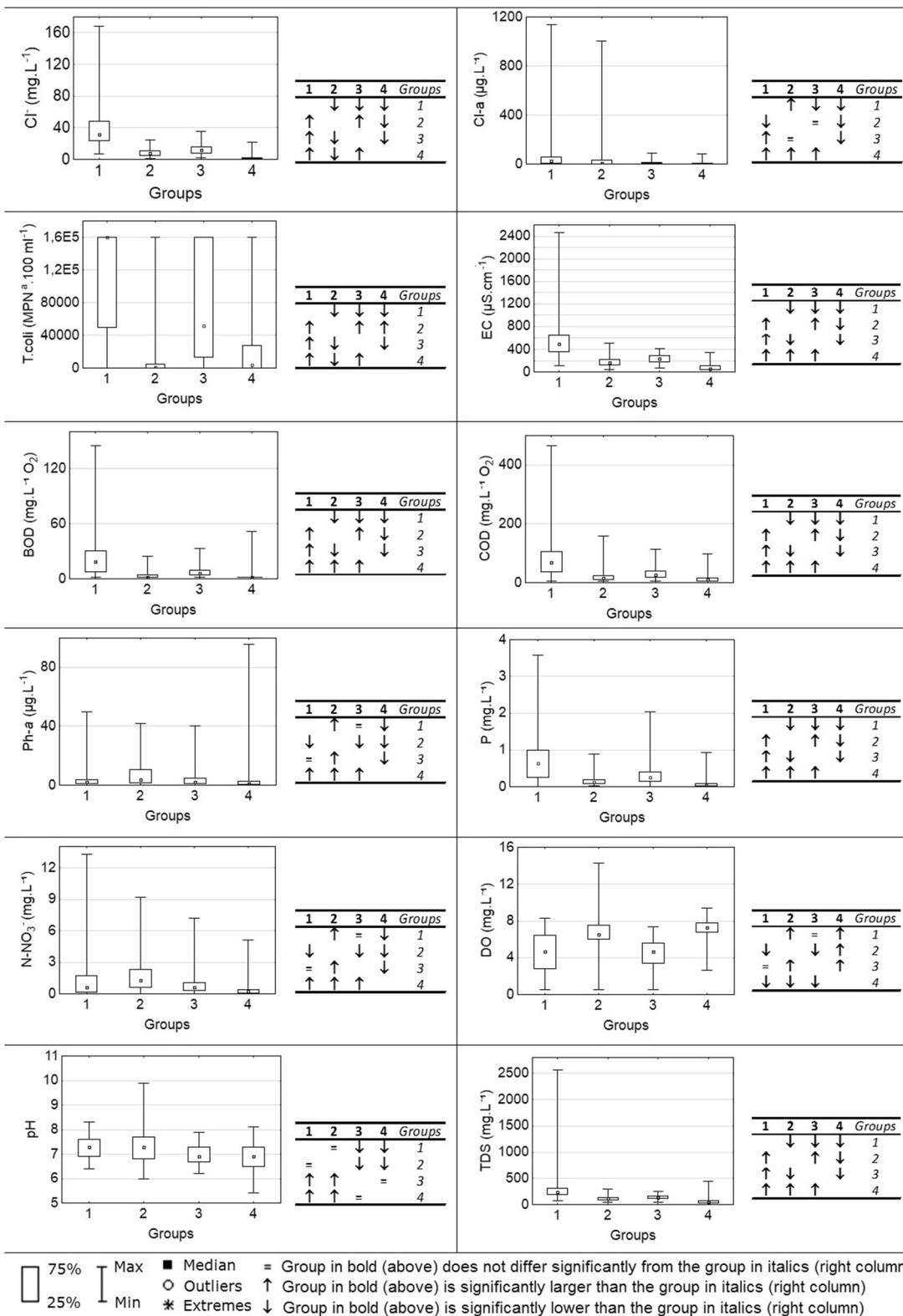**Fig. 3** Identification of the monitoring site clusters formed by the CA

Fig. 4 Box plots of water quality parameter results, and concentration results of the Kruskal–Wallis and multiple comparisons tests ($p < 0.05$), considering the groupings formed by CA

receives industrial and domestic effluents that cause high concentrations of dissolved substances (SEMAD 2006). The nonparametric Kruskal–Wallis test results confirmed the significant high concentrations between electrical conductivity and total dissolved solids in this group.

These monitoring sites have higher BOD, COD, and highly significant P concentrations than those of other groups due to the input of organic matter from domestic sewage. The concentrations of T. *Coli.* in these groups significantly differ from those in other groups, and their medians were at the limit of the analytical method (160,000 NMP 100 mL^{-1}), indicating that these sites are severely contaminated by domestic sewage. SC10, BV155, and BV154 are located in the Onça and Arrudas Streams, which receive domestic and industrial sewage from the cities of Belo Horizonte and Sabará (BV155 and BV154) and Belo Horizonte and Contagem (SC10). The water quality condition of these cities is the most critical in the Velhas River Basin (IGAM 2013b; IGAM 2013a), and these monitoring sites have the highest and most significant solid concentrations.

The surface water quality at some monitoring sites that presented significant amounts of T. *Coli.* is mainly affected by biodegradable organic matter and solid loading. Such monitoring sites are located in streams in the MRBH.

Cluster II (monitoring sites SC13, BV149, BV151, BV148, BV152, BV150, BV146, BV156, BV161, BV142, BV140, and BV141)

SC13 is located downstream of the sanitary land-fill in Ribeirão das Neves. As well as the solids contributed by this, algae have also developed in this region. The DO concentrations of this group were significantly higher than those of other groups. The main sources of pollution at these points are agriculture, pasture areas, and algal blooms. Significantly higher concentrations of Cl-*a* and Ph-a were observed.

Some of these monitoring sites are located in regions of transitioning land uses; therefore, they suffer pressures from both domestic sewage and industrial effluent (Sete Lagoas industrial pole), and diffuse pollution from agriculture and the pasture areas of farms in the region (bean, corn, rice, sorghum).

Cluster III (monitoring sites SC19, SC21, BV160, BV137, SC17, BV130, SC16, BV153, BV105, and BV083)

The river in this region faces intense sedimentation and contains a high concentration of biodegradable organic matter, which deteriorates the river's water quality. Some reports published by IGAM confirm these results (IGAM 2013b; IGAM 2014), and pollution sources include diffuse pollution from pasture areas, agriculture, and sand extraction.

Cluster IV

Group 4 contains the highest number of monitoring sites, which are not heavily affected by the sources of pollution mentioned above. Therefore, the water quality of these sites can be considered to be better than others in the Velhas River Basin. The concentrations of water quality parameters in this group were significantly lower than those in the others according to the Kruskal–Wallis test.

Analysis of water quality parameter violations

The concentrations of water quality parameters measured in the monitoring program were compared to the standards sets in ND COPAM/CERH-MG 01/2008 (Supplementary Material: Table S1). Information about the classes of rivers for each monitoring site and the percentage of noncompliance to ND COPAM/CERH-MG 01/2008 for each parameter at each monitoring site can be found in the supplementary material (Table S2 and Table S3). The analysis of violation was conducted for the 36 water quality parameters that have regulatory values, which were monitored in the Velhas river basin from July 2008 to December 2013. Total mercury, sulfide, and total cyanide were not considered as their detection limits were higher than the maximum permissible concentrations. Thermotolerant coliforms, total manganese, and total phosphorus were considered to be the most relevant parameters for monitoring the water quality of the Velhas River Basin. These parameters are critical in over 46% of the analyzed monitoring sites (Fig. 5).

Thermotolerant coliforms were a critical parameter in 75% of the analyzed monitoring sites. Thus, the results indicate that the entire Velhas River Basin is contaminated by domestic sewage. This result is a
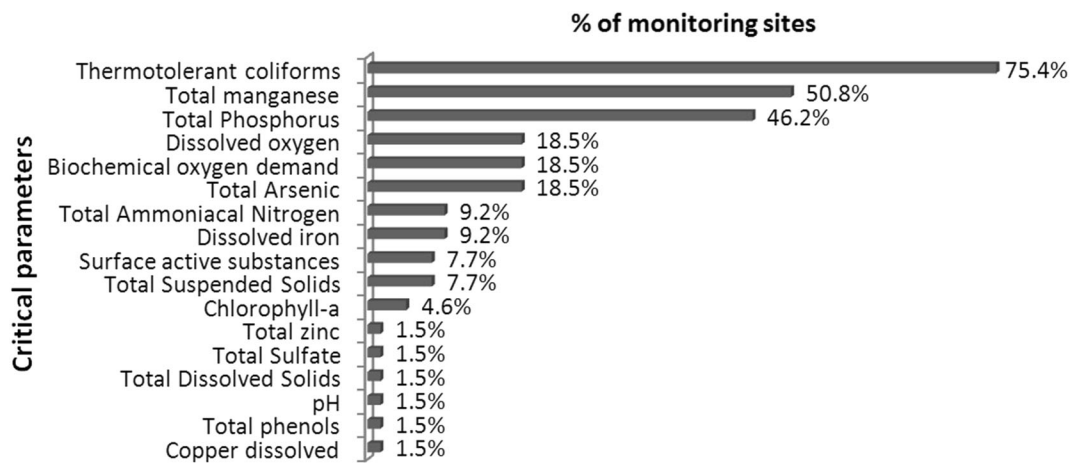
Fig. 5 Percentage of monitoring sites in the Velhas River Basin that violated the analyzed parameters at a rate of 50% or more

consequence of the low rate of or insufficient treatment of sewage in the basin.

The other critical parameters presented in Fig. 5, which are responsible for the degradation of water quality in the Velhas River Basin, are also related to the discharge of domestic sewage, industrial effluent, mining activities, drainage of agricultural areas, and erosion and silting. A similar result was found by Knupp (2007), who evaluated the percentage of violation at the monitoring sites of the Velhas River Basin from 1997 to 2003. This indicates that the main problems previously identified in the basin persist after 10 years of monitoring, which reinforces the need to improve sanitation and the inspection of industrial activities.

Despite the natural occurrence of several metals in the region, mining activities intensify the exposure of these metals to the environment. It is important to highlight that the Upper Velhas Region is characterized by important mineral reserves of iron, manganese, copper, antimony, arsenic, gold, aluminum, and uranium, and it is extensively exploited by mining activities (CBH Velhas 2018). This fact justifies the high percentage of violation of the total manganese. Studies by Costa et al. (2017) and Pinto et al. (2018), in the region of the Velhas river, corroborate the results found.

Among the 65 monitoring sites analyzed, sites SC14 and SC03, followed by SC26, SC25, and AV320, have the highest number of critical parameters (Table 4). In addition to domestic sewage contamination (characteristic of high T. *coli*. and low DO concentrations), the water quality of these monitoring sites is being degraded by effluent from the chemical and mining industries. These monitoring sites are characterized by high levels

of toxic contamination (IGAM 2013b) owing to the high concentrations of metals and other toxic substances in watercourses (dissolved arsenic, free cyanide, dissolved copper, dissolved iron, manganese, sulfide, and total zinc).

The results analyzed by monitoring sites are consistent with those found for the clusters. Monitoring sites belonging to group 1 in the dendrogram had the highest number of critical parameters. Therefore, the analysis of violation is consistent with cluster analysis as the monitoring sites grouped together coincide with the largest violations in the Velhas River Basin, which validated their results.

The spatial analysis of the 65 monitoring sites of the Velhas river basin proved to be valuable, since it was possible to realize a division of these sites in zones with different characteristics and sources of pollution. The main pressure factors acting on the surface water quality of the Velhas river basin are domestic sewage areas of pasture, industrial effluents, mining, and diffuse pollution by agriculture. And, considering the entire Velhas river basin, the most relevant parameter identified was thermotolerant coliform. This is a worrisome result that affects the quality of life in cities and represents real risks to the health of the local population.

In the study by Oliveira et al. (2018), the authors highlighted the fact that untreated or insufficiently treated sewage contamination generates degradation throughout the basin, especially in urban areas, in addition to compromising supply sources.

A temporal analysis was realized in the study by Calazans et al. (2018) using monthly frequency monitoring data on the surface water quality of the Velhas

Table 4 Number of critical parameters (violation of 50% or more) for each monitoring site of the Velhas River Basin during the period from July 2008 to December 2013

	Number of critical parameters								
	8	7	6	5	4	3	2	1	0
Monitoring sites	**SC14**	**SC26**	*SC19*	*SC16*	**SC10**	SC24	SC28	SC39	BV162
	SC03	**SC25**	BV156	SC13	BV142	SC23	*SC17*	SC33	*BV161*
		AV320	**BV154**	*BV160*	BV141	*SC21*	SC12	SC30	BV144
				BV155	*BV130*	*BV153*	BV146	SC27	BV143
				BV062		BV152	BV139	SC22	
						BV150	BV135	BV151	
						BV137	*BV105*	BV149	
						BV063	*BV083*	BV148	
						BV035	BV076	BV147	
						AV300	BV067	BV145	
							BV037	BV140	
							BV013	BV136	
							AV340	BV133	
							AV210	AV250	
							AV080	AV060	
							AV070		
							AV020		
							AV010		
							AV007		

Monitoring sites belong to group 1; group 2; group 3; and group 4 (cluster analysis)

river basin, in the same evaluation period of this work. The results showed that the temporal variation in water quality is partly determined by hydrological conditions, but there are variations between months of the year. The cluster analysis applied grouped the 12 months of the year into two groups. The influence of seasonality was observed in the groups formed, where the months of the dry season and the months of the rainy season belong to different groups.

Conclusions

This study presents the results from the assessment of spatial variations in the surface water quality of the Velhas River Basin, which involved multivariate statistical analysis (cluster analysis (CA)), a Kruskal–Wallis test, and an analysis of violations of legal standards. Sixteen water quality parameters were measured at each monitoring site. CA grouped the 65 monitoring sites into four clusters with similar water quality characteristics. The Kruskal–Wallis test aided in identifying the

parameters responsible for water quality variations. The Upper and Upper–Middle Velhas regions are more affected by the release of industrial effluents and domestic sewage, while the others (Lower–Middle and Lower Velhas regions) are more affected by diffuse pollution and erosion. The results suggest that, among the several pollution sources that affect the surface water quality of the Velhas River Basin, the discharge of domestic sewage that has not undergone adequate treatment predominates throughout the basin. Different surface water quality characteristics of the Velhas River basin are observed in regions upstream and downstream of the Metropolitan Region of Belo Horizonte due to deterioration after this region and differences between land use characteristics. The parameters thermotolerant coliforms, total manganese, and total phosphorus were considered the most relevant for the characterization of water quality in the Velhas River basin. This evaluation can be used as a guideline in decision-making for effective water quality management and could help managers identify the main sources of pollution in different regions to determine their priorities for improving water

quality. The analysis of 65 monitoring sites allowed a broad knowledge of the entire basin. The division of these sites into groups is important for planning and making decision in the river basin. Different measures should be adopted according to regions and sources of pollution, aiming at the sustainable use of water resources.

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

References

Achieng', A. O., Raburu, P. O., Kipkorir, E. C., Ngodhe, S. O., Obiero, K. O., & Ani-Sabwa, J. (2017). Assessment of water quality using multivariate techniques in River Sosiani, Kenya. *Environmental Monitoring and Assessment, 189*, 1–13. https://doi.org/10.1007/s10661-017-5992-5.

Agência Nacional De Águas – ANA. (2013). Cuidando das águas: soluções para melhorar a qualidade dos recursos hídricos, second ed. ANA, Brasília. Accessed 25 June 2017.

Agência Nacional De Águas – ANA. (2015). Conjuntura dos Recursos Hídricos no Brasil: regiões hidrográficas brasileiras, special ed. Superintendência de Planejamento de Recursos Hídricos, Brasília. Accessed 25 June 2017.

Alkarkhi, A., Ahmad, A., Ismail, N., Easa, A., Omar, K. (2008). Assessment of surface water through multivariate analysis. *Journal of Sustainable Development, 1*(3), 27–33.

Almeida, K.C.B. (2013). Avaliação da rede de monitoramento de qualidade das águas superficiais da Bacia do Rio das Velhas utilizando o método da entropia. Universidade Federal de Minas Gerais. http://www.smarh.eng.ufmg.br/defesas/1042 M.PDF. Accessed 01 January 2017.

APHA, AWWA, WEF. (2012). *Standard methods for examination of water and wastewater* (22nd ed.). Washington: American Public Health Association 1360p.

Bhat, S. A., Meraj, G., Yaseen, S., & Pandit, A. K. (2014). Statistical assessment of water quality parameters for pollution source identification in Suknag Stream: an inflow stream of Lake Wular (Ramsar Site), Kashmir Himalaya. *Journal of Ecosystems, 18*, 1–18. https://doi.org/10.1155/2014/898054.

Calazans, G. M., Pinto, C. C., Costa, E. P., Perini, A. F., & Oliveira, S. C. (2018). Using multivariate techniques as a strategy to guide optimization projects for the surface water

quality network monitoring in the Velhas river basin, Brazil. *Environmental Monitoring and Assessment, 190*(12), 2–15. https://doi.org/10.1007/s10661-018-7099-z.

Comitê da Bacia Hidrográfica do Rio das Velhas – CBH Velhas. (2014). Análise integrada, articulação e compatibilização dos interesses internos e externos, cenários e prognósticos: relatório 03, revisão 02 – atualização do plano diretor de recursos hídricos da bacia hidrográfica do rio das Velhas.

Comitê da Bacia Hidrografica do Rio das Velhas – CBH Velhas. (2018). A Bacia. Disponível em < http://cbhvelhas.org.br/a-bacia-hidrografica-do-rio-das-velhas/>. Acesso em: 10 dez. 2018.

Conselho Estadual de Meio Ambiente/Conselho Estadual de Recursos Hídricos (COPAM/CERH-MG). (2008). Deliberação Normativa Conjunta COPAM/CERH-MG n°01, de 05 de maio de 2008. Brazil. http://www.siam.mg. gov.br/sla/download.pdf?idNorma=8151. Accessed 15 June 2017.

Costa, E. P., Pinto, C. C., Soares, A. L. C., Melo, L. D. V., & Oliveira, S. C. (2017). Evaluation of violations in water quality standards in the monitoring network of São Francisco River basin, the third largest in Brazil. *Environmental Monitoring and Assessment, 189*(11), 2–16. https://doi.org/10.1007/s10661-017-6266-y.

Finotti, A. R., Finkler, R., Silva, M. D., & Cemim, G. (2009). *Monitoramento de Recursos Hídricos em Áreas Urbanas* (1st ed.). Rio Grande do Sul: Educs.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. (2009). *Análise Multivariada de dados* (6th ed.). Porto Alegre: Book.

Instituto Mineiro de Gestão das Águas – IGAM. (2013a). Monitoramento da Qualidade das Águas Superficiais no Estado de Minas Gerais: relatório trimestral (2° trimestre de 2013). Accessed 20 May 2018.

Instituto Mineiro de Gestão das Águas – IGAM. (2013b). Identificação de municípios com condição crítica para a qualidade de água na bacia do rio das Velhas. Belo Horizonte. Accessed 16 June 2017.

Instituto Mineiro de Gestão das Águas – IGAM. (2014). Monitoramento da Qualidade das Águas Superficiais no Estado de Minas Gerais (2° trimestre de 2014). Accessed 20 May 2018.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (sixth ed.). New Jersey: Pearson.

Knupp, E.A.N. (2007). Usos de métodos estatísticos para dados de qualidade de águas: estudo de caso, rio das Velhas. Universidade Federal de Minas Gerais. http://www. repositorio.cdtn.br:8080/bitstream/123456789/908/1/Tese_ Eliana_ANonato.pdf. Accessed 21 December 2016.

Lattin, J., Carroll, J. D., & Green, P. E. (2011). *Análise de Dados Multivariados* (1st ed.). São Paulo: Cengage Learning.

Lee, H., Chan, Z., Graylee, K., Kajenthira, A., Martínez, D., & Roman, A. (2014). Challenge and response in the São Francisco River basin. *Water Policy, 16*(S1), 153–200. https://doi.org/10.2166/wp.2014.007.

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. *The Science of Total Environment, 313*, 77–89.

Mendiguchía, C., Moreno, C., Galindo-Riano, M. D., & Garcia-Vargas, M. (2004). Using chemometric tools to assess anthropogenic effects in river water, a case study: Guadalquivir

River (Spain). *Analytica Chimica Acta, 515*(1), 143–149. https://doi.org/10.1016/j.aca.2004.01.058.

Naghettini, M., & Pinto, E. J. A. (2007). *Hidrologia Estatística* (1st ed.). Belo Horizonte: CPRM.

Oliveira, S. C., Pinto, C. C., Soares, A. L. C., Costa, E. P. (2018). Avaliação da qualidade de água em 14 anos de monitoramento da porção mineira da bacia do rio São Francisco. II Simpósio da Bacia Hidrográfica do Rio São Francisco - Desafios da Ciência para um novo Velho Chico, Sergipe, Brasil.

Panda, U. C., Sundaray, S. K., Rath, P., Nayak, B. B., & Bhatta, D. (2006). Application of factor and cluster analysis for characterization of river and estuarine water systems – a case study: Mahanadi River (India). *Journal of Hydrology, 331*(3–4), 434–445. https://doi.org/10.1016/j.jhydrol.2006.05.029.

Park, S. Y., Choi, J. H., Wang, S., & Park, S. S. (2006). Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modelling, 199*, 289–297. https://doi.org/10.1016/j.ecolmodel.2006.06.002.

Phung, D., Huang, C., Rutherford, S., Dwirahmadi, F., Chu, C., Wang, X., Nguyen, M., Nguyen, N. H., Do, C. M., Nguyen, T. H., & Dinh, T. A. D. (2015). Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. *Environmental Monitoring and Assessment, 187*, 1–13. https://doi.org/10.1007/s10661-015-4474-x.

Pinto, C. C., Almeida, K. B., & Oliveira, S. C. (2018). Avaliação espacial da qualidade das águas da calha do rio das Velhas, estado de Minas Gerais. *Tchê Química, 15*(30), 75–86.

Reghunath, R., Sreedhara Murthy, T. R., Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research, 36*(10), 2437–2442.

Schwertmana, N. C., Owensa, M. A., & Adnanb, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis, 47*(1), 165–174. https://doi.org/10.1016/j.csda.2003.10.012.

Secretaria do Estado do Meio Ambiente e Desenvolvimento Sustentável de Minas Gerais – SEMAD. (2006). Aperfeiçoamento do monitoramento da qualidade das águas da Bacia do Alto Curso do Rio das Velhas. Accessed 10 June 2017.

Sistema Nacional de Informações sobre Saneamento - SNIS. (2017). *Diagnóstico dos serviços de água e esgotos - 2015*. Brasília: SNSA/MCIDADES, 212p.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika, 52*, 591–599.

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River Basin, Japan. *Environmental Modeling and Software, 22*(4), 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001.

Simeonova, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., & Kouimtzis, T. (2003). Assessment of the surface water quality in Northern Greece. *Water Research, 37*(17), 4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1.

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. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques – a case study. *Analytica Chimica Acta, 538*, 355–374. https://doi.org/10.1016/j.aca.2005.02.006.

Sojka, M., Siepak, M., Ziola, A., Frankowski, M., Murat-Blażejewska, S., & Siepak, J. (2008). Application of multivariate statistical techniques to evaluation of water quality in the MalaWelna River (Western Poland). *Environmental Monitoring and Assessment, 147*, 159–170. https://doi.org/10.1007/s10661-007-0107-3.

Strobl, R. O., & Robillard, P. D. (2008). Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management, 87*(4), 639–648.

Trindade, A.L.C. (2013). Aplicação de técnicas estatísticas para avaliação de dados de monitoramento de qualidade das águas superficiais da porção mineira da Bacia do Rio São Francisco. Universidade Federal de Minas Gerais. http://www.smarh.eng.ufmg.br/defesas/1037M.PDF. Accessed 18 June 2018.

Trindade, A. L. C., Almeida, K. C. B., Barbosa, P. E., & Oliveira, S. M. A. C. (2017). Tendências temporais e espaciais da qualidade das águas superficiais da sub-bacia do rio das Velhas, estado de Minas Gerais. *Engenharia Sanitária e Ambiental, 22*, 13–24. https://doi.org/10.1590/s1413-41522016131457.

Varol, M., Gökot, B., Bekleyen, A., & Şen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River Basin, Turkey. *Catena, 92*, 11–21. https://doi.org/10.1016/j.catena.2011.11.013.

Vega, M., Pardo, R., Barrado, E., & Deban, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research, 32*(12), 3581–3592. https://doi.org/10.1016/S0043-1354(98)00138-9.

Wunderlin, D. A., Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., & Bistoni, M. A. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. *Water Research, 35*(12), 2881–2894. https://doi.org/10.1016/S0043-1354(00)00592-3.

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. *Environmental Monitoring and Assessment, 173*(1–4), 17–27. https://doi.org/10.1007/s10661-010-1366-y.