

The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil

Giovanna Moura Calazans ·
Carolina Cristiane Pinto ·
Elizângela Pinheiro da Costa · Anna Flávia Perini ·
Sílvia Corrêa Oliveira 

Received: 27 December 2017 / Accepted: 19 July 2018 / Published online: 28 July 2018
© Springer Nature Switzerland AG 2018, corrected publication August/2018

Abstract This study sought to evaluate and propose adjustments to the water quality monitoring network of surface freshwaters in the Paraopeba river basin (Minas Gerais, Brazil), using multivariate statistical methods. A total of 13,560 valid data were analyzed for 19 water quality parameters at 30 monitoring sites, over a period of 5 years (2008–2013). The cluster analysis grouped the monitoring sites in eight groups based on similarities of water quality characteristics. This analysis made it possible to detect the most relevant monitoring stations in the river basin. The principal components analysis associated with non-parametric tests and the analysis of violation of the standards prescribed by law, allowed for identifying the most relevant parameters which must be maintained in the network (thermotolerant coliforms, total manganese, and total phosphorus). The discharge of domestic sewage and industrial wastewater, that from mining activities and diffuse pollution from agriculture and pasture areas are the main sources of pollution responsible for the surface water quality deterioration in this basin. The BP073 monitoring site presents the most degraded water quality in the Paraopeba river basin. The monitoring sites BP094 and BP092 are located geographically close and they measure similar water quality, so a possible assessment of the need to maintain

only one of the two in the monitoring network is suggested. Therefore, multivariate analyses were efficient to assess the adequacy of the water quality monitoring network of the Paraopeba river basin, and it can be used in other watersheds.

Keywords Cluster analysis · Principal components analysis · Network monitoring assessment · Brazilian watershed

Introduction

The monitoring of surface water quality is of extreme importance for understanding the current situation of water resources and the main changes that have occurred over time. It also allows for detecting the temporal and spatial trends of watersheds. Systematic and periodic monitoring of the water resource conditions allows for planning interventions for improvements, identifying clandestine releases, subsidizing inspection, environmental licensing and the formulation of environmental policies (Finotti et al. 2009; Bilgin and Konanç 2016).

However, the large number of water quality parameters sampled at multiple monitoring stations results in a high cost to the environmental agency. Therefore, strategies are needed to guide the optimization of the monitoring network, such as the identification of relevant parameters and stations, which must be maintained. In this sense, several methodologies are used for the

G. M. Calazans · C. C. Pinto · E. P. da Costa · A. F. Perini ·
S. C. Oliveira (✉)
Universidade Federal de Minas Gerais - Escola de Engenharia -
Campus Pampulha, Av. Antônio Carlos, 6627, Bloco 1, sala 4525,
Belo Horizonte, MG 31270-901, Brazil
e-mail: silvia@desa.ufmg.br

evaluation and resizing of surface water quality monitoring networks (Dong et al. 2015) such as modeling (Paliwal et al. 2007), numerical models and algorithms (Chen et al. 2012; Lee et al. 2014; Park et al. 2006; Strobl and Robillard 2008), fuzzy logic (Huang et al. 2010; Xu et al. 2012; Wu et al. 2010; Zhao et al. 2011), artificial neural networks (Khalil et al. 2011), entropy method (Almeida 2013; Markus et al. 2003; Soares 2001), and proposal of new methods (Telci et al. 2009; Wang et al. 2006).

The relevance of this work is in the use of multivariate statistical techniques for the proposed adaptation of the water quality monitoring network of a river basin in Brazil. These techniques have been applied to water quality monitoring networks throughout the world with satisfactory results (Gamble and Babbar-Sebens 2012; Shrestha and Kazama 2006; Vieira et al. 2012; Zhang et al. 2009).

The water quality monitoring network of the Paraopeba river basin (Minas Gerais, Brazil), the focus of this study, is located in an important economic center of the Minas Gerais state. Over the last decade, some water quality parameters have presented elevated frequencies of violation to local legislation permitted values, such as fecal contamination indicator, phosphorus, and manganese (Costa et al. 2017). Thus, the several economic activities that are developed on this area together with the intense urbanization may be responsible for different impacts on the water resources.

The monitoring network of the Paraopeba river basin has been active since 1997, and its main objectives are (1) to know and to evaluate the surface water quality, disseminating it to users and guiding the establishment of quality goals; (2) to provide subsidies for the planning of water resources management, and (3) to verify the effectiveness of environmental control actions implemented and to propose priorities for action. In almost 20 years of the monitoring program, no evaluation or resizing tests were conducted for this basin. The high cost of the monitoring program is an obstacle to its maintenance and expansion. Thus, optimization of the network may reduce costs by eliminating unnecessary sampling or less important parameters and equivalent monitoring points. Furthermore, some resources may be applied to other areas in the basin or to the measurement of parameters of greatest need. Thus, the results of this work contribute to the planning and integrated management of water resources in the basin and can serve as a guide for work to adapt networks in other river basins.

Materials and methods

Study area

The Paraopeba river basin is an important sub-basin of the São Francisco river basin. The São Francisco river basin is one of the most important in Brazil due to the large water contribution in six states and the Federal District, being the largest river in Brazil. The Upper São Francisco region, in which the Paraopeba river basin is located, is the most degraded region in relation to surface water quality. The drainage area of the Paraopeba river basin is 12,054 km², which corresponds to 2.5% of the state of Minas Gerais, and it is slightly larger than the land area of countries like Jamaica and Qatar. The Paraopeba river has a length of 510 km and almost 2.5 million people live in the basin, which covers 48 municipalities (Fig. 1).

The uses of water in the basin vary from the dilution of domestic and industrial effluents to more noble ones, such as water supply, fish farming, and recreation. The use and occupation of soil in the Paraopeba river basin is quite diverse, which generated large impacts on surface waters due to different anthropogenic pressures. This basin also includes one of the most important economic centers of Minas Gerais that forms the second industrial pole of the country. There are also activities such as cattle raising, iron and steel, food and automobile industries, mining, sand and clay extraction, extensive animal husbandry, and eucalyptus monocultures.

Monitoring of the surface water quality in the Paraopeba river basin has been carried out since 1997 by the Minas Gerais Water Management Institute (IGAM), through the Project “Minas Waters”. From the implementation of the network in 1997 until 2013, 1509 water samples were collected, and 62 water quality parameters were monitored at 33 monitoring sites, with a quarterly and semi-annual frequency sampling, obtaining a total of 59,860 valid data points. Due to the extensive and complex database, previous data selection was required for the application of multivariate statistics.

Selection of the data for application of multivariate statistical techniques

Preparation of the database for application of statistical techniques consisted of the selection of parameters and monitoring stations while minimizing the missing data.

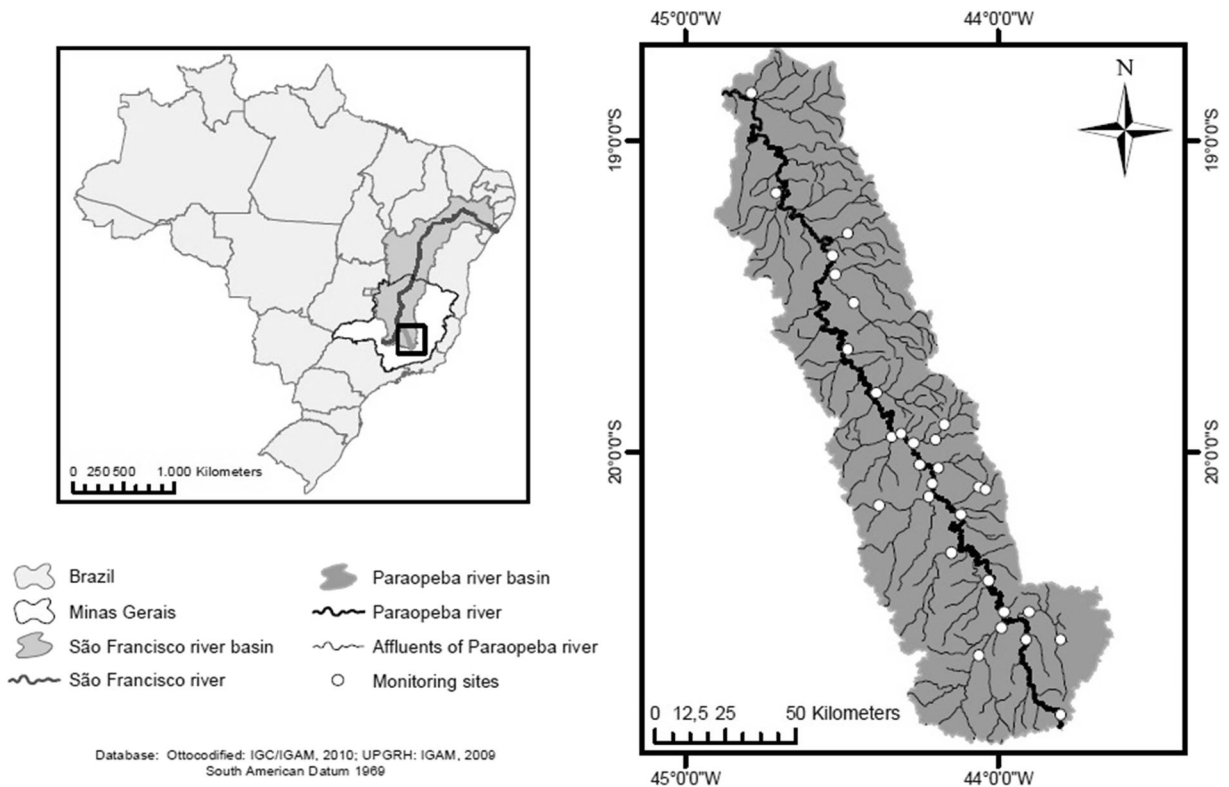


Fig. 1 Localization of Paraopeba river basin, Brazil

Monitoring sites were selected that presented continuous quarterly sampling. Parameters with more than 10% missing data were disregarded. Parameters whose environmental and sanitary significance was low for the interpretation of water quality were also disregarded. Parameters with more than 90% censored data (concentrations below the limit of detection of the analytical method of measurement) were not included, as they did not express variation in water quality during the monitoring period. Therefore, these parameters were highlighted as the least relevant for monitoring in this basin. The monitoring database selected covered 720 samples of 19 parameters in 30 monitoring sites, resulting in an analysis of 13,560 valid data points from the period of 2008 to 2013. The selected parameters included ammonia-nitrogen ($N-NH_4^+$), chemical oxygen demand (COD), chlorophyll-a (*Cl-a*), dissolved iron (Fe), dissolved oxygen (DO), electrical conductivity (EC), 5-day biochemical oxygen demand (BOD_5), nitrate nitrogen ($N-NO_3^-$), organic nitrogen (N_{org}), pH, temperature (TEMP), thermotolerant coliforms (*T. colif.*), total chloride (Cl^-), total manganese (Mn), total phosphorus (P), total solids (TS), total suspended solids

(TSS), true color (COL), and turbidity (TURB). The descriptive statistics (minimum, maximum, median, mean, and standard deviation) of the water quality parameters were calculated for each monitoring site. Distribution of the data was verified by applying the Shapiro-Wilk normality test at a 5% significance level.

Data treatment and multivariate statistical techniques

Cluster analysis

Cluster analysis was used to group the 30 monitoring sites according to the spatial similarity of water quality. The results showed high homogeneity within the group, and a high heterogeneity between the groups (Lattin et al. 2011). The hierarchical agglomeration and euclidean distance was used as a measure of similarity. Prior to the test the data was standardized. This evaluation sought to identify areas where the water quality is similar in the river basin and this statistical technique is commonly used to access the water quality in watersheds throughout the world (Zhang et al. 2009; Gamble and Babbar-Sebens 2012; Vieira et al. 2012; Guigues

et al. 2013; Bhuiyan et al. 2016; Bilgin and Konanç 2016]. In order to identify significant differences between the concentrations of water quality parameters for the groups formed by cluster analysis (CA), non-parametric statistical tests were applied (Mann-Whitney (two groups) or Kruskal-Wallis (more than two groups)), followed by the multiple comparison test (when applicable), at the significance level (α) of 5%. Therefore, it was possible to identify which parameters were responsible for differentiating the groups formed by the CA.

Principal component analysis

The principal components analysis (PCA) is a complementary analysis, which seeks to reorganize and show the multivariate data in order to reduce its size, facilitating its visualization and analysis (Mingoti 2005; Abdi and Williams 2010). This reduction occurs when there is substantial redundancy between the data, and therefore, a reduced number of elements can often explain most of the information (İşçen et al. 2009). In this study, the PCA was used to reduce the complexity of the monitoring database, identifying the main parameters responsible for explaining the greater variability of water quality and the different sources of pollution acting in each group formed by the CA.

Evaluation and proposal for optimization of the monitoring network

The violation percentage of the water quality parameters was calculated in relation to the standards set in COPAM (State Environmental Policy Council) and CERH (State Water Resources Council) Normative Deliberation 01/2008 (MINAS GERAIS 2008), legislation that provides environmental guidelines on water bodies in the state of Minas Gerais, Brazil. The water quality parameters that presented 50% or more violation in the analyzed period, called “critical parameters,” were considered the most important parameters, as well as the parameters which explained the greater variability of water quality, as indicated by the PCA. The CA allowed for identifying differences in water quality monitoring sites, guiding their selection in the monitoring network. The parameters which had more than 90% censored data were considered the least relevant for the monitoring network.

Results and discussion

Descriptive statistics

The results of descriptive statistics analysis are presented in Table 1. It is possible to note a large amplitude for most parameters, resulting in high values of standard deviation and skewness. This asymmetry is common in environmental data and it was confirmed by the Shapiro-Wilk test ($p < 0.05$). In addition, this variation reinforces the importance of the spatial analysis of the concentration of these parameters in the Paraopeba river basin.

Cluster analysis and the Kruskal-Wallis test

For grouping of the 30 surface water quality monitoring sites of the Paraopeba river basin, the Euclidean distance of 21.27 was considered, which resulted in the formation of eight groups as shown in the dendrogram in Fig. 2.

The dendrogram allows for identifying the monitoring sites that present high homogeneity in water quality (monitoring sites in the same group) and those with high heterogeneity (monitoring sites belong to different groups). The monitoring sites of groups 1, 2, 3, 4, 5, and 6 (BP073, BP071, BP098, BP084, BP080, BP086, BP069, and BP069) can be considered the most relevant in the Paraopeba river basin because they have unique characteristics, which are different from most other monitoring sites. Not coincidentally, these sites represent the bodies of water that obtained the highest frequencies of “very bad” and “bad” on the water quality index (WQI) between the monitoring sites of the Paraopeba river basin in the period from 1997 to 2012 (IGAM 2013). Most of these monitoring sites are located in the industrial pole region, in the city of Betim, where water quality is mainly affected by domestic sewage and industrial effluents. The groups formed by the CA were identified geographically in Fig. 3.

The differences between the groups formed by the CA can be confirmed by the results of the Kruskal-Wallis non-parametric test, whose objective was to find significant differences in the concentrations of the parameters analyzed. The results of the multiple comparison test also confirm that the monitoring stations BP073 and BP071 (groups 1 and 2, respectively) are the stations whose surface water quality is more degraded in the Paraopeba river basin. These stations have the

Table 1 Descriptive statistics of water quality parameters in the Paraopeba river basin, from 2008 to 2013

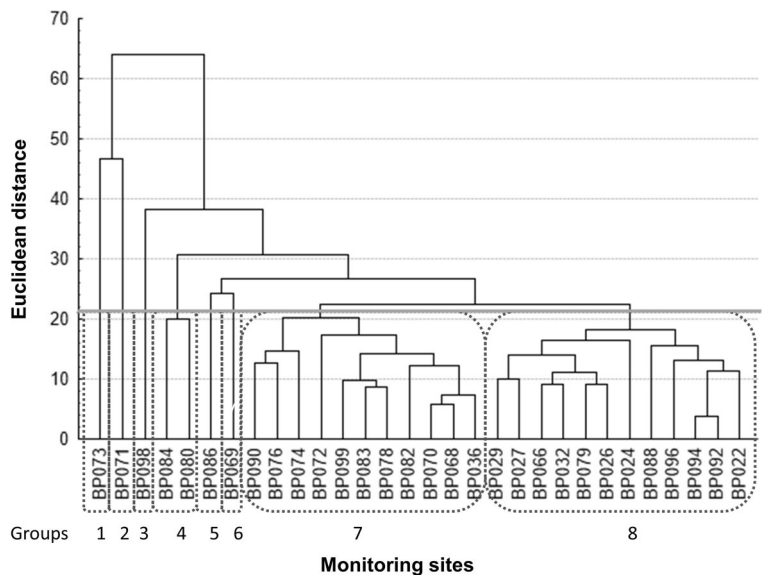
Parameters/unit	MIN	MAX	MED	MEAN	SD	Skewness	Kurtosis
N-NH ₄ ⁺ (mg L ⁻¹)	0.10	21.4	0.1	1.1	3.2	4.2	17.4
COD (mgDO L ⁻¹)	5.0	273.0	13.0	19.1	25.8	5.2	36.1
Cl- <i>a</i> (µg L ⁻¹)	0.006	723.3	2.79	9.6	35.6	13.4	239.3
Fe (mg L ⁻¹)	0.03	1.8	0.2	0.2	0.2	3.0	13.0
DO (mg L ⁻¹)	0.20	9.5	7.0	6.5	1.9	-1.6	2.0
EC (µS cm ⁻¹)	16.1	1483	70.7	113.7	132.3	3.8	22.1
BOD ₅ (mgDO L ⁻¹)	2.0	138.0	2.0	4.2	9.2	8.1	86.8
N-NO ₃ ⁻ (mg L ⁻¹)	0.01	9.4	0.3	0.6	0.9	4.7	32.6
N _{org} (mg L ⁻¹)	0.10	11.0	0.4	0.6	0.7	7.1	82.1
pH	5.7	8.5	6.8	6.8	0.5	0.2	-0.3
TEMP (°C)	17.0	35.2	24.2	24.1	3.1	0.2	-0.4
T. colif. (MPN.100 mL ⁻¹)	2.0	160,000	5000	29,913	49,962	1.9	2.1
Cl ⁻ (mg L ⁻¹)	0.3	230.0	3.13	7.45	15.11	6.6	71.7
Mn (mg L ⁻¹)	0.01	2.8	0.1	0.3	0.4	3.0	10.3
P (mg L ⁻¹)	0.01	5.9	0.1	0.2	0.4	6.6	67.4
TS (mg L ⁻¹)	24.0	2474.0	107.0	158.6	179.5	5.4	49.1
TSS (mg L ⁻¹)	1.0	2239.0	32.0	76.4	158.5	6.5	63.9
COL (HC)	5.0	1704.0	29.0	91.3	172.0	4.5	26.6
TURB. (NTU)	1.2	2512.0	27.0	75.0	162.0	7.0	79.1

MIN minimum, MAX maximum, MED median, SD standard deviation

highest significant concentrations for Cl⁻, Cl-*a*, EC, COD, P, N-NO₃⁻, N_{org}, and TS and the lowest concentrations of DO in relation to the other stations (Fig. 4). Concerning domestic sewage contamination, the concentration of T. colif was significantly lower for groups

3, 7, and 8. Most of the T. colif observations in group 1 focused on the detection limit value of the analytical method, which is 160,000 MPN.100 mL⁻¹. Because of this, BP073 is highly contaminated by domestic sewage, and it has presented the highest frequencies (100%) of

Fig. 2 Dendrogram showing the clustering of sampling sites in the Paraopeba river basin



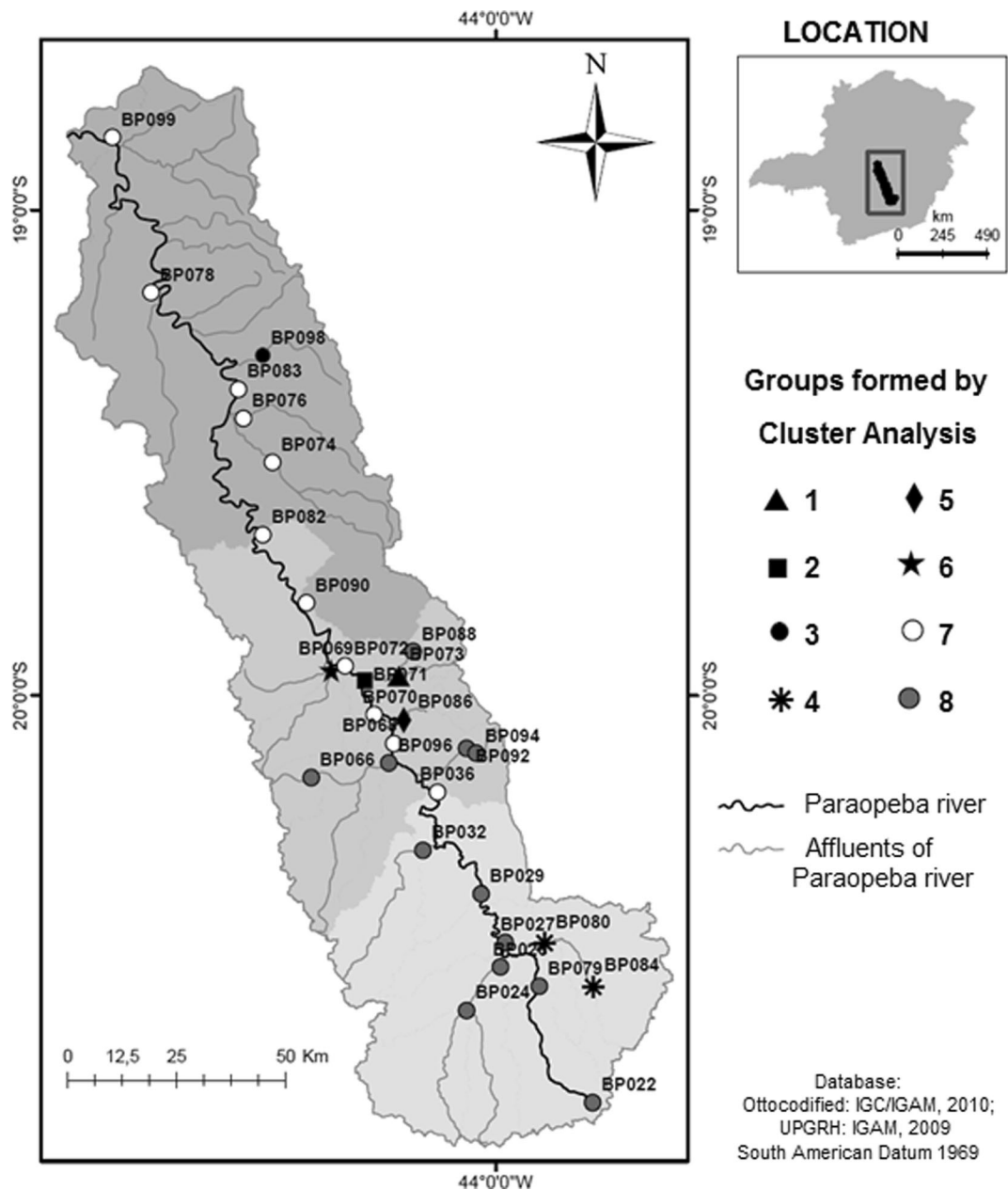


Fig. 3 Identification of the groups of monitoring sites in the Paraopeba river basin formed by the cluster analysis

“very bad” and “bad” classifications of the water quality index (WQI) in the Paraopeba river basin from 1997 to 2012, according IGAM (IGAM 2013).

From this analysis, it can be inferred that if there is a need to interrupt or permanently stop the water monitoring at some sites, where these sites must be present in group 7 or 8 because these groups have a larger number of monitoring sites in which the water quality is similar, and they present the best water quality in the basin.

Principal component analysis

The results of the PCA explained 63.1 to 82.9% of the surface water quality variability of each group formed by the CA, retaining five principal components (PC), considering factors with eigenvalues greater than one (Hair et al. 2005). The correlations between variables and the PC of each group formed by the CA are presented in Table 2.

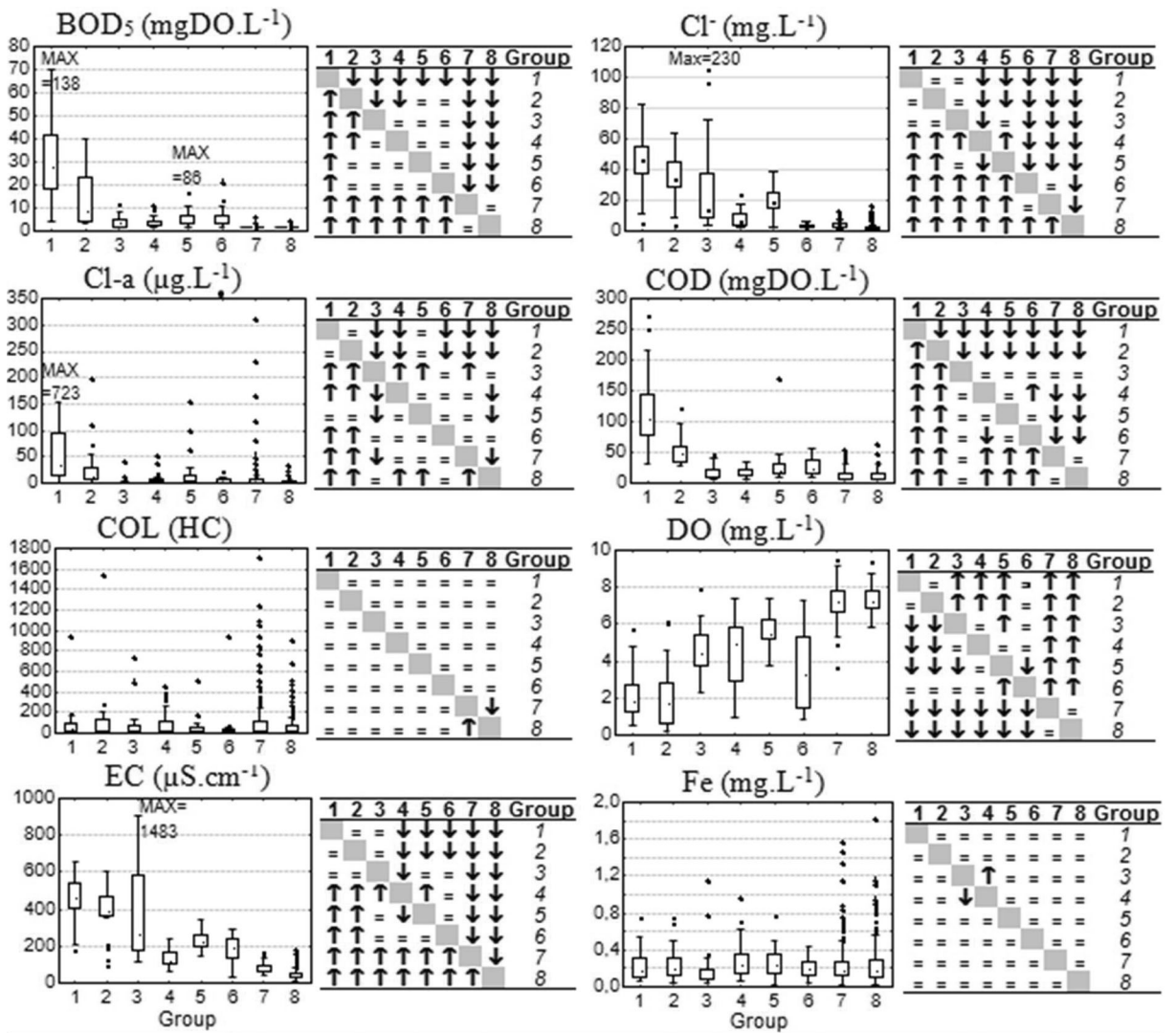


Fig. 4 Box-plot of the concentrations of the water quality parameters and results of the Kruskal-Wallis test and the multiple comparisons test ($p < 0.05$), considering the groups formed by the cluster analysis

It can be observed that groups 1, 2, 3, 4, 5, and 6 are similar in relation to the variables that best correlated with the first factors. The variables that best explain the variability of the water quality of these monitoring sites are related to the presence of organic matter, mainly from domestic sewage and solids in the water body. The monitoring site BP073 (group 1) receives domestic sewage and industrial effluents from the city of Betim, a large industrial region (IGAM 2014). Station BP071 (group 2) is located just downstream of monitoring site BP073, so it is also affected by the water quality of the

previous site. EC is also highly correlated with the first factors in these groups, indicating a high concentration of dissolved solids in the water body. Groups 1, 2, and 3 present a higher concentration of total solids in relation to the other monitoring sites as presented by the non-parametric Kruskal-Wallis test results.

The monitoring sites belonging to group 4 (BP084 and BP080) are located on the same tributary in the beginning of the river basin. In addition to the parameters related to organic matter and solids, a high correlation of dissolved iron is highlighted in this group, which

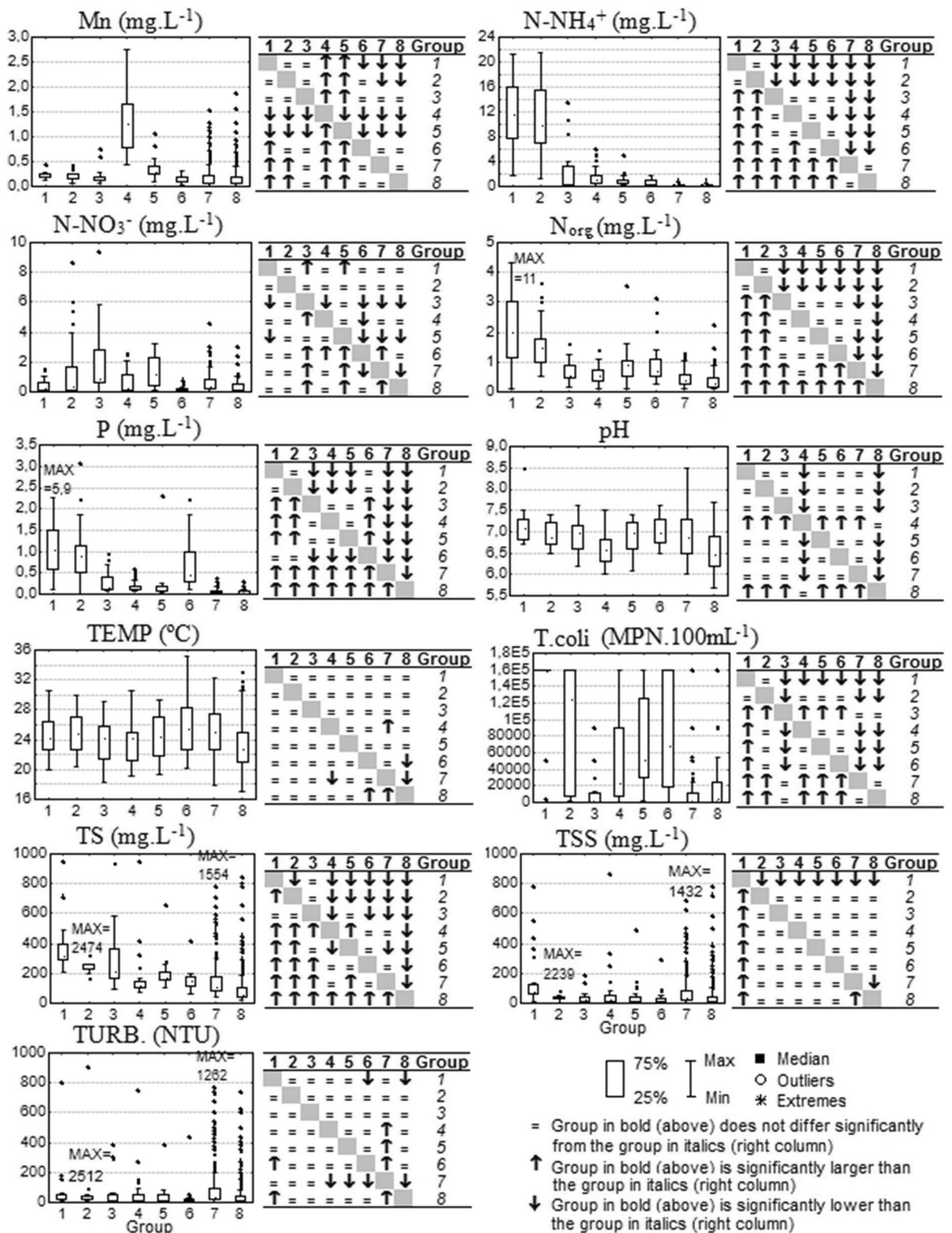


Fig. 4 (continued)

Table 2 Principal components analysis for the monitoring sites in the Paraopeba river basin, considering groups formed by the cluster analysis

Group 1						Group 2					
Variables	PC1	PC2	PC3	PC4	PC5	Variables	PC1	PC2	PC3	PC4	PC5
Cl ⁻	0.697	0.481	-0.545	-0.343	0.249	Cl ⁻	0.197	-0.874	-0.016	-0.069	0.356
Cl- α	0.107	-0.324	-0.285	-0.433	0.388	Cl- α	0.392	0.455	0.231	-0.249	-0.213
T. colif.	0.085	0.703	-0.511	0.199	-0.022	T. colif.	0.484	0.260	-0.791	0.057	0.359
EC	0.475	0.693	-0.501	-0.567	0.139	EC	0.122	-0.920	0.068	-0.145	0.362
COL	0.013	-0.653	0.564	0.298	-0.683	COL	0.525	0.721	-0.253	0.666	-0.516
BOD ₅	0.875	0.013	-0.133	0.017	0.327	BOD ₅	0.790	-0.061	-0.005	0.041	0.250
COD	0.922	0.161	0.026	0.112	0.469	COD	0.910	0.164	-0.151	0.434	-0.126
Fe	0.369	0.500	0.061	0.364	0.079	Fe	0.733	0.156	-0.293	0.379	-0.416
P	0.638	0.110	-0.002	0.303	0.558	P	0.426	-0.637	0.072	0.205	0.342
Mn	0.620	-0.139	-0.214	-0.370	0.295	Mn	0.621	0.546	-0.057	0.171	-0.234
N-NO ₃ ⁻	-0.242	0.406	0.287	-0.208	0.110	N-NO ₃ ⁻	-0.561	-0.215	0.110	0.339	-0.318
N-NH ₄ ⁺	0.424	0.769	-0.220	-0.315	0.018	N-NH ₄ ⁺	0.339	-0.784	-0.236	-0.043	0.333
N _{org}	0.638	-0.075	0.225	0.410	0.641	N _{org}	0.679	0.282	0.096	0.223	-0.527
DO	-0.733	-0.471	0.559	0.344	-0.523	DO	-0.279	0.586	-0.101	0.213	-0.514
pH	0.039	0.397	0.262	-0.507	0.285	pH	-0.633	-0.383	0.610	0.067	-0.128
TSS	0.294	-0.659	0.812	0.269	-0.209	TSS	0.476	0.867	-0.253	0.567	-0.582
TS	0.420	-0.571	0.771	0.157	-0.236	TS	0.515	0.837	-0.259	0.593	-0.581
Temp.	0.100	-0.468	-0.456	-0.313	-0.048	Temp.	0.019	0.226	0.565	-0.575	-0.142
Turb.	0.036	-0.681	0.745	0.252	-0.596	Turb.	0.459	0.828	-0.272	0.621	-0.578
Eigenvalue	6.61	3.16	2.80	1.35	1.17	Eigenvalue	6.19	4.68	1.75	1.52	1.26
Variance (%)	34.8	16.6	14.7	7.1	6.2	Variance (%)	32.6	24.6	9.2	8.0	6.6
Cumulative (%)	34.8	51.4	66.2	73.3	79.4	Cumulative (%)	32.6	57.2	66.4	74.4	81.0
Group 3						Group 4					
Variables	PC1	PC2	PC3	PC4	PC5	Variables	PC1	PC2	PC3	PC4	PC5
Cl ⁻	0.423	0.280	0.134	0.276	0.532	Cl ⁻	0.498	-0.733	0.057	-0.204	-0.217
Cl- α	0.246	-0.192	0.093	-0.717	0.177	Cl- α	0.091	-0.210	0.049	-0.686	-0.497
T. colif.	-0.326	0.034	-0.292	-0.090	0.649	T. colif.	-0.173	-0.170	-0.082	0.143	-0.746
EC	0.904	0.269	-0.303	0.123	0.050	EC	0.633	-0.627	-0.015	0.010	-0.221
COL	-0.180	0.793	-0.205	0.235	-0.088	COL	-0.418	0.270	-0.708	-0.267	0.011
BOD ₅	0.821	0.119	-0.551	0.238	-0.073	BOD ₅	0.684	0.102	-0.424	0.014	-0.035
COD	0.779	0.453	-0.454	-0.064	-0.415	COD	0.338	0.023	-0.587	0.347	0.430
Fe	0.491	-0.014	-0.411	0.464	-0.060	Fe	0.738	0.271	0.017	0.255	0.145
P	0.837	0.221	-0.764	0.279	0.024	P	0.781	0.082	-0.339	0.278	0.053
Mn	0.652	0.307	-0.605	-0.059	-0.432	Mn	0.157	0.469	-0.462	0.462	-0.236
N-NO ₃ ⁻	0.346	-0.609	-0.095	0.012	0.032	N-NO ₃ ⁻	-0.060	-0.773	0.261	0.100	-0.016
N-NH ₄ ⁺	0.893	0.055	-0.592	0.062	-0.287	N-NH ₄ ⁺	0.855	0.186	-0.136	0.291	0.109
N _{org}	0.796	-0.068	-0.260	-0.090	-0.158	N _{org}	0.547	-0.510	-0.010	0.387	0.170
DO	-0.691	-0.148	0.271	0.195	0.319	DO	-0.611	-0.685	-0.025	-0.293	0.046
pH	0.674	-0.115	-0.190	0.032	-0.157	pH	-0.027	-0.460	-0.096	0.155	-0.071
TSS	-0.368	0.728	-0.268	0.099	0.065	TSS	-0.544	-0.055	-0.811	0.182	0.143
TS	0.638	0.561	-0.222	0.133	0.137	TS	-0.489	-0.148	-0.838	0.156	0.118
Temp.	-0.445	0.271	0.467	-0.616	-0.347	Temp.	-0.578	0.205	0.199	-0.134	-0.238
Turb.	-0.369	0.747	-0.186	0.153	-0.001	Turb.	-0.531	-0.001	-0.830	0.145	0.109

Table 2 (continued)

Eigenvalue	7.30	3.98	1.76	1.24	1.14	Eigenvalue	5.69	3.55	2.42	1.32	1.15
Variance (%)	38.4	20.9	9.2	6.5	6.0	Variance (%)	29.9	18.7	12.7	6.9	6.0
Cumulative (%)	38.4	59.4	68.6	75.2	81.1	Cumulative (%)	29.9	48.6	61.4	68.3	74.3
Group 5						Group 6					
Variables	PC1	PC2	PC3	PC4	PC5	Variables	PC1	PC2	PC3	PC4	PC5
Cl ⁻	0.183	0.255	<i>- 0.613</i>	- 0.408	- 0.406	Cl ⁻	<i>0.676</i>	0.354	- 0.042	- 0.219	- 0.398
Cl- α	0.080	- 0.283	<i>- 0.653</i>	0.221	0.105	Cl- α	- 0.120	0.164	<i>- 0.616</i>	- 0.434	0.250
T. colif.	0.227	0.261	0.223	- 0.083	0.465	T. colif.	<i>0.504</i>	- 0.307	- 0.340	0.382	0.059
EC	0.244	0.444	<i>- 0.325</i>	- 0.242	- 0.457	EC	<i>0.854</i>	0.294	0.180	- 0.038	- 0.059
COL	0.480	<i>- 0.719</i>	0.341	0.038	<i>- 0.613</i>	COL	<u>0.012</u>	<u>- 0.823</u>	0.428	- 0.452	- 0.135
BOD ₅	<u>0.763</u>	0.305	0.262	<i>0.684</i>	0.277	BOD ₅	<i>0.612</i>	<u>- 0.299</u>	- 0.172	0.474	0.072
COD	<u>0.765</u>	0.313	0.244	<i>0.650</i>	0.381	COD	<u>0.830</u>	- 0.054	- 0.171	0.250	0.025
Fe	<u>- 0.424</u>	0.424	0.409	0.167	0.195	Fe	<u>0.505</u>	0.167	0.553	0.075	0.209
P	<i>0.745</i>	0.323	0.283	<i>0.652</i>	0.314	P	<u>0.759</u>	0.057	- 0.046	0.196	0.425
Mn	<i>0.507</i>	<i>- 0.720</i>	0.349	- 0.328	- 0.338	Mn	<u>0.367</u>	- 0.456	<i>0.540</i>	0.332	- 0.316
N-NO ₃ ⁻	- 0.265	0.444	<i>0.523</i>	0.037	- 0.013	N-NO ₃ ⁻	0.024	0.396	<i>0.678</i>	0.398	0.141
N-NH ₄ ⁺	<u>0.821</u>	0.277	0.448	<i>0.601</i>	0.122	N-NH ₄ ⁺	<u>0.755</u>	0.200	0.122	0.074	0.306
N _{org}	<u>0.782</u>	0.411	0.257	<i>0.569</i>	0.235	N _{org}	<u>0.608</u>	- 0.130	- 0.136	0.322	0.037
DO	<u>- 0.206</u>	- 0.018	- 0.354	<i>0.670</i>	0.045	DO	<u>- 0.855</u>	- 0.214	- 0.111	- 0.273	0.172
pH	- 0.163	0.201	0.263	0.359	- 0.258	pH	<u>- 0.071</u>	0.350	0.337	0.193	<i>0.581</i>
TSS	<i>0.653</i>	<i>- 0.690</i>	0.409	0.011	- 0.498	TSS	- 0.054	<u>- 0.866</u>	0.332	- 0.380	0.031
TS	<i>0.691</i>	<i>- 0.583</i>	0.295	- 0.039	<i>- 0.617</i>	TS	0.480	<u>- 0.655</u>	0.375	- 0.435	- 0.084
Temp.	0.220	<i>- 0.501</i>	0.306	0.175	0.266	Temp.	<i>- 0.588</i>	- 0.383	- 0.126	- 0.057	0.111
Turb.	<i>0.611</i>	<i>- 0.685</i>	0.459	0.039	<i>- 0.522</i>	Turb.	0.004	<u>- 0.843</u>	0.414	- 0.421	- 0.152
Eigenvalue	5.91	3.50	2.52	1.84	1.52	Eigenvalue	6.99	<u>2.75</u>	2.26	1.60	1.15
Variance (%)	31.1	18.4	13.2	9.7	8.0	Variance (%)	36.8	14.5	11.9	8.4	6.0
Cumulative (%)	31.1	49.5	62.8	72.5	80.5	Cumulative (%)	36.8	51.3	63.1	71.6	77.6
Group 7						Group 8					
Variables	PC1	PC2	PC3	PC4	PC5	Variables	PC1	PC2	PC3	PC4	PC5
Cl ⁻	- 0.419	<i>0.613</i>	0.162	- 0.113	- 0.265	Cl ⁻	<i>0.522</i>	<i>0.644</i>	- 0.289	0.135	0.247
Cl- α	- 0.194	0.112	0.287	- 0.172	0.231	Cl- α	0.334	0.295	- 0.382	0.355	- 0.012
T. colif.	0.254	0.061	<i>- 0.503</i>	- 0.388	0.252	T. colif.	0.170	- 0.113	- 0.088	<i>- 0.535</i>	- 0.338
EC	<i>- 0.439</i>	<i>0.517</i>	0.300	- 0.260	- 0.051	EC	<i>0.515</i>	<i>0.689</i>	- 0.321	0.121	0.299
COL	<u>0.760</u>	0.194	0.042	- 0.005	0.140	COL	0.434	<i>- 0.641</i>	- 0.018	0.292	- 0.090
BOD ₅	<u>0.268</u>	0.224	- 0.125	<i>- 0.508</i>	0.430	BOD ₅	0.177	0.338	- 0.352	- 0.124	- 0.034
COD	<i>0.651</i>	0.437	- 0.060	0.094	0.080	COD	<i>0.698</i>	- 0.021	- 0.154	0.336	- 0.304
Fe	0.031	0.001	- 0.436	- 0.245	- 0.420	Fe	0.107	- 0.211	0.088	<i>- 0.593</i>	0.052
P	0.395	0.478	- 0.110	0.067	- 0.256	P	<i>0.650</i>	- 0.072	- 0.055	0.148	- 0.406
Mn	<u>0.784</u>	0.166	- 0.101	0.408	- 0.090	Mn	<u>0.797</u>	- 0.070	0.118	0.344	- 0.126
N-NO ₃ ⁻	<u>- 0.380</u>	0.317	0.029	0.349	- 0.193	N-NO ₃ ⁻	<u>0.153</u>	0.495	0.417	- 0.303	0.153
N-NH ₄ ⁺	- 0.124	0.435	- 0.473	- 0.397	0.040	N-NH ₄ ⁺	0.341	<i>0.619</i>	- 0.410	- 0.150	0.177
N _{org}	0.189	<i>0.723</i>	- 0.149	0.109	0.166	N _{org}	<i>0.530</i>	0.437	0.103	0.067	- 0.067
DO	- 0.400	- 0.094	<i>- 0.519</i>	0.306	0.461	DO	- 0.473	0.155	<i>0.519</i>	0.080	- 0.353
pH	- 0.324	0.003	0.061	0.189	<i>0.655</i>	pH	0.142	0.350	0.486	0.133	0.329
TSS	<u>0.895</u>	0.372	- 0.135	0.278	0.109	TSS	<i>0.738</i>	- 0.489	0.300	0.434	- 0.415
TS	<u>0.887</u>	0.427	- 0.117	0.234	0.138	TS	<u>0.804</u>	- 0.416	0.240	0.463	- 0.371
Temp.	<u>0.396</u>	- 0.130	0.483	- 0.184	- 0.029	Temp.	<u>0.396</u>	- 0.180	- 0.218	0.060	<i>0.557</i>
Turb.	<u>0.903</u>	0.402	- 0.105	0.245	0.094	Turb.	<i>0.687</i>	<i>- 0.571</i>	0.241	0.364	- 0.347
Eigenvalue	<u>5.16</u>	2.41	1.64	1.42	1.36	Eigenvalue	5.70	2.93	1.60	1.28	1.14
Variance (%)	27.2	12.7	8.6	7.5	7.1	Variance (%)	30.0	15.4	8.4	6.7	6.0
Cumulative (%)	27.2	39.8	48.5	55.9	63.1	Cumulative (%)	30.0	45.4	53.9	60.6	66.6

Italicized values represent high correlation (≥ 0.5)

Double underlined values represent strong correlation (≥ 0.75)

can be explained by its proximity to mining activity. The BP084 site is the only one in the basin in which the iron concentration is above the limit allowed by the law in more than half of the samples.

The PC1 of group 5 (BV086) also highlights a correlation with total manganese. According to the Kruskal-Wallis non-parametric test, this group presents a higher concentration of manganese in relation to the other

monitoring stations in the basin, except for group 4, which presents a higher concentration than group 5. Total manganese may be the result of mining activities, but it is also important to consider that the basal concentration of this mineral in the river is not known, whose concentration can be highly altered during the rainy seasons. The BP069 site constitutes group 6 and the high negative correlation with dissolved oxygen indicates the high degree of deterioration of the watercourse. This monitoring

site receives the impacts from most of the sewage from the cities of Juatuba and Mateus Leme, where there is a maximum sewage collection of 65% (IGAM 2013).

Group 7, consisting of 11 monitoring sites, has a strong correlation with manganese and parameters related to the presence of dissolved and suspended solids in water (COL, COD, TSS, TS, and Turb.), indicating that the main source of pollution in these areas is the transport of solids to the river, mainly from non-biodegradable

Table 3 Surface water quality standards set in Normative Deliberation COPAM/CERH-MG 01/2008

Parameter	Unit	Class			
		1	2	3	4
Biochemical oxygen demand	mg L ⁻¹ O ₂	3	5	10	–
Chlorophyll-a	µg L ⁻¹	10	30	60	–
Density of cyanobacteria	cel mL ⁻¹	20,000	50,000	100,000	–
Dissolved copper	mg L ⁻¹ Cu	0.009	0.009	0.013	–
Dissolved iron	mg L ⁻¹ Fe	0.3	0.3	5.0	–
Dissolved oxygen	mg L ⁻¹ O ₂	6	5	4	2
Free cyanide	mg L ⁻¹ CN	0.005	0.005	0.022	–
Nitrate	mg L ⁻¹ N	10.0	10.0	10.0	–
Nitrite	mg L ⁻¹ N	1	1	1	–
Oils and greases		0	0	0	–
pH		6 to 9	6 to 9	6 to 9	6 to 9
Total ammoniacal nitrogen pH ≤ 7.5	mg L ⁻¹ N	3.7	3.7	13.3	–
T. ammoniacal nitrogen 7.5 < pH < 8.0	mg L ⁻¹ N	2	2	5.6	–
T. ammoniacal nitrogen 8.0 < pH < 8.5	mg L ⁻¹ N	1	1	2.2	–
T. ammoniacal nitrogen pH ≥ 8.5	mg L ⁻¹ N	0.5	0.5	1	–
Thermotolerant coliforms	org 100 mL ⁻¹	200	1000	4000	–
Total arsenic	mg L ⁻¹ As	0.01	0.01	0.033	–
Total barium	mg L ⁻¹ Ba	0.7	0.7	1.0	–
Total boron	mg L ⁻¹ B	0.5	0.5	0.75	–
Total cadmium	mg L ⁻¹ Cd	0.001	0.001	0.01	–
Total chloride	mg L ⁻¹ Cl	250	250	250	–
Total chromium	mg L ⁻¹ Cr	0.05	0.05	0.05	–
Total dissolved solids	mg L ⁻¹	500	500	500	–
Total lead	mg L ⁻¹ Pb	0.01	0.01	0.033	–
Total manganese	mg L ⁻¹ Mn	0.1	0.1	0.5	–
Total nickel	mg L ⁻¹ Ni	0.025	0.025	0.025	–
Total phosphorus	mg L ⁻¹ P	0.1	0.1	0.15	–
Total selenium	mg L ⁻¹ Se	0.01	0.01	0.05	–
Total suspended solids	mg L ⁻¹	50	100	100	–
Total zinc	mg L ⁻¹ Zn	0.18	0.18	5	–
True color	mgPt L ⁻¹	–	75	75	–
Turbidity	NTU	40	100	100	–

organic materials. Note that in this group the concentration of biodegradable organic matter is smaller than other groups previously analyzed. These monitoring sites are located predominantly on the main river channel, in the lower course, downstream from the region where the main contaminations of domestic sewage and industrial effluents are introduced. Therefore, this better condition of water quality is probably due to dilution of the contaminants and pollutants with increase of the river flow.

Correlations similar to those identified for group 7 were found in characterization of the water quality of the 12 monitoring sites in group 8. However, for most of the analyzed parameters, the concentrations in group 7 are significantly higher than that in group 8. Therefore, the stations of group 8 can be considered those of best water quality in the basin. These monitoring sites are located mainly in the upper course of the river basin and in tributaries of the main river.

Therefore, it can be observed that association of the results of CA, PCA and non-parametric Kruskal-Wallis test can lead to an understanding of the sources of pollution in different regions of the river basin, as well as identify areas most critical to water quality. These techniques allow an in-depth study of the superficial water quality of the river basin, orienting studies on the resizing of water quality monitoring networks.

Analysis of violation of the water quality parameters with regard to the standards set in ND COPAM/CERH-MG 01/08 and analysis of censured data

The concentrations of water quality parameters measured in the monitoring program were compared to

the standards sets in ND COPAM/CERH-MG 01/2008 (Table 3). It was verified that T.colif. was considered a critical parameter in 82% of the monitoring sites analyzed. That is, the discharge of untreated domestic sewage into the water body is still the main problem with regard to water quality in the Paraopeba river basin. It is important to emphasize that this fact does not necessarily result from the lack of investments in sewage treatment in this period, but from the fact that investments made were not sufficient to substantially reduce the sewage volume. Additionally, the type of treatment employed may also not be effective in recovering the water quality of the basin. In this respect the parameters total manganese (critical in 67% of the monitoring stations) and total phosphorus (critical in 30% of the monitoring stations) are highlighted. These results coincide with those obtained by the PCA.

Analyzing the number of critical parameters in the monitoring sites of the Paraopeba river basin, it is verified that the monitoring sites belonging to the first groups of the CA (1, 2, 3, 4, 5, and 6) were those that presented the most critical parameters, as shown in Table 4. Thus, the analysis of critical parameters coincides with the groups formed by the CA, demonstrating the efficiency of the multivariate technique for water quality analysis.

Optimizing the Paraopeba river basin surface water quality monitoring network

The parameters total chromium, total cadmium, total selenium, dissolved copper and total boron have 90%

Table 4 Number of critical parameters for each of the monitoring sites in the Paraopeba river basin

	Number of critical parameters										
	<u>8</u>	<u>5</u>	<u>4</u>	<u>3</u>	2			1			0
Monitoring sites	<u>BP073</u>	<u>BP069</u> <u>BP084</u>	<u>BP098</u>	<u>BP071</u> <u>BP080</u> <u>BP086</u>	BP024 BP027 BP029 BP036	BP066 BP068 BP070 BP072	<u>BP074</u> <u>BP076</u> BP079 <u>BP083</u>	BP022 BP026 BP032 <u>BP078</u>	<u>BP082</u> BP088 <u>BP090</u> BP092	BP096 <u>BP099</u>	BP094

Italic and underlined: monitoring sites belong to groups 1 to 6 (cluster analysis)

Upright and underlined: monitoring sites belong to group 7

Upright and no underline: monitoring sites belong to group 8

Table 5 Summary table of the analysis of violations of the standards defined in ND COPAM/CERH-MG 01/08, for each monitoring site of the Paraopeba river basin

Parameters	BP022	BP024	BP026	BP027	BP029	BP032	BP036	BP066	BP068	BP069	BP070	BP071	BP072	BP073	BP074	BP076
As (mg L ⁻¹)	C	N	C	N	N	C	C		C	C	N	C		C		C
Ba (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
B (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
Cd (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
Pb (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
CN (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
Cl ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Cl ⁻ a (µg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Cu (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
T. colif (MPN.100 mL ⁻¹) ¹)	V	V	V	V	V	V	V		V	V	V	V	V	V	V	V
COL (HC)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Cr (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
BOD ₅ (mgDO L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
D.yvano (cel mL ⁻¹)	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-
Fe (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
P (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Mn (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Ni (mg L ⁻¹)	C	N	N	N	N	N	N		N	N	N	N	N	N	N	N
N-NO ₃ ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
N-NO ₂ ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
N-NH ₄ ⁺ (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
OG	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
DO (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
pH	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Se (mg L ⁻¹)	C	C	C	C	C	C	C		C	C	C	C	C	C	C	C
TDS (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
TSS (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
TURB. (NTU)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N
Zn (mg L ⁻¹)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N

Parameters	BP078	BP079	BP080	BP082	BP083	BP084	BP086	BP088	BP090	BP092	BP094	BP096	BP098	BP099
As (mg L ⁻¹)		C		C	C	N	C	C	C	C	C	C	C	
Ba (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	N	N	N	
B (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Cd (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Pb (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C
CN (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Table 5 (continued)

Parameters	BP078	BP079	BP080	BP082	BP083	BP084	BP086	BP088	BP090	BP092	BP094	BP096	BP098	BP099
Cl ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Cl-a (µg L ⁻¹)					N	N								
Cu (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C
T. colif (MPN.100 mL ⁻¹) ¹)		V	V	V	V	V	V		V	V	V	V	V	V
COL (HC)		N						N	N	N				
Cr (mg L ⁻¹)	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BOD ₅ (mgDO L ⁻¹)	N	C		C	C				C	C	C	N		C
D.vyano (cel mL ⁻¹)	-	-	-	-	-	-	N	N	-	-	-	-	-	N
Fe (mg L ⁻¹)			V		N	V	V	N		N			V	
P (mg L ⁻¹)		N	V		V	V	V	V		N	N		V	
Mn (mg L ⁻¹)	V	V	V		V	V	V	V		N	N		V	
Ni (mg L ⁻¹)		N	N		C	N	C	C		C		C		
N-NO ₃ ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	N	N	N	N
N-NO ₂ ⁻ (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	N	N	N	N
N-NH ₄ ⁺ (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	C	N	N	C	N
OG	C	C	C		C	C	C		C	C	C	C	C	C
DO (mg L ⁻¹)	N	N			N	V			N	N	N	N	V	N
pH	N	N	N		N	N	N		N	N			N	
Se (mg L ⁻¹)	C	C	N	C	C	C	C	C	C	C	C	C	C	C
TDS (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	N	N	N	N
TSS (mg L ⁻¹)					N									
TURB. (NTU)					N									
Zn (mg L ⁻¹)	N	N	N	N	N	N	N	N	N	N	C	N	N	N

As total arsenic, *Ba* total barium, *B* total boron, *Cd* total cadmium, *Pb* total lead, *Cu* dissolved copper, *Ni* total nickel, *Se* total selenium, *Cr* total chromium, *N-NO₂⁻* nitrite, *TDS* total dissolved solids, *CN* total cyanide, *D.cyan* density of cyanobacteria, *OG* oils and greases, *Zn* total zinc

- parameter not measured

V 50% or more of violation of legal standards - critical parameter (most relevant parameter)

C 90% or more of censored data - parameter of minor relevance

N parameter did not violate legal standards (from July 2008 to December 2013)—parameter of minor relevance

or more data censored in more than 85% of the monitoring sites evaluated. Therefore, they can be considered the least relevant in monitoring the quality of surface waters in the Paraopeba river basin, and their sampling frequency may be reduced. BP092, BP094, and BP090 are monitoring sites that have the highest number of parameters sampled with 90% or more of censored data (12, 13, or 14), considering the monitoring period from 2008 to 2013, indicating low variability of water quality. The results also suggest an evaluation of the maintenance of BP092 and BP094 in the monitoring network of the Paraopeba river basin. These monitoring sites are located geographically close; they have the lowest Euclidean distance of clustering according to CA, have low number of critical parameters, and have a large number of censored data. So, it is suggested for the state agency—IGAM—to evaluate the relevance of the maintenance of both stations in the monitoring network. Perhaps only one of them would be sufficient to evaluate the background concentration for some of the elements on the basin and serve as a basis for comparison of water quality with other monitoring sites more impacted by anthropogenic activities. Therefore, a permanently stop on water monitoring at BP094 should be indicated due to its geographical proximity to BP092, to the similarity of water quality on both stations and because there were no violations to standards in more than 13 years of monitoring on BP094 site.

Table 5 summarizes the results obtained by the violation analysis and censored data, presenting the most important water quality parameters for the monitoring site (50% or greater violation of legal standards), “V”; the least relevant (90% or more of censored data), “C”; and especially those which did not have violations during the study period (N), for each one of the 30 monitoring sites in the Paraopeba river basin. For the elaboration of this table we considered all parameters sampled in the monitoring network (including parameters not considered in multivariate analysis) which have legal standards in ND COPAM/CERH-MG 01/2008, totaling 29 water quality parameters.

Conclusions

The optimization and resizing of a surface water quality monitoring network for a river basin should be done with caution and planning. This work showed that the association of multivariate statistical techniques with

non-parametric tests and analysis of violating the legal standards allows for identifying similar monitoring sites regarding water quality, prioritizing the maintenance of monitoring stations in the monitoring network, identifying the main parameters that explain the water quality variability and the main sources of pollution in different regions of the river basin. This association was efficient to generate a proposal for optimization and resizing of the monitoring network.

The parameters thermotolerant coliforms, total manganese, and total phosphorus were considered the most relevant for the characterization of water quality in the Paraopeba river basin. On the other hand, total chromium, total cadmium, selenium, total dissolved copper, and total boron were considered less relevant. The discharge of domestic sewage and industrial wastewater are the main sources of pollution responsible for the surface water quality deterioration in the basin. The monitoring site BP073 has shown the most intense degradation, while the BP094 site is responsible for monitoring of the best water quality of the basin. It is suggested for IGAM an assessment of the need for maintenance of BP094 in the monitoring network, since it is located geographically close to BP092, which also has similar water quality, and both may be expressing the same information. In this way, the resources spent on BP094 monitoring could be allocated to areas of greater need in the basin. It is our wish that this work may guide the decision-making of public managers in the Paraopeba river basin, as well as works in other river basins around the world.

Acknowledgments Thanks to the Institute of Water Management of Minas Gerais (IGAM) and its technical team for providing the monitoring data and for the constant support and service.

Funding information Thanks to the National Council of Technological and Scientific Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES), and the Foundation of Support Research of the State of Minas Gerais (FAPEMIG) for the financial support to execute this project.

References

- Abdi, H., Williams, L. J. (2010). Principal component analysis – overview. *WIREs Computational Statistics*, (2), 433–459. <https://doi.org/10.1002/wics.101>. Wiley.
- 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 18 June 2017.
- Bhuiyan, M. A. H., Bodrud-Doza, M., Towfiqul Islam, A. R. M., Rakib, M. A., Rahman, M. S., & Ramanathan, A. L. (2016). Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis. *Environment and Earth Science*, 75(1020), 1–23. <https://doi.org/10.1007/s12665-016-5823-y>.
- Bilgin, A., & Konanç, M. U. (2016). Evaluation of surface water quality and heavy metal pollution of Coruh River Basin (Turkey) by multivariate statistical methods. *Environment and Earth Science*, 75(1029), 1–18. <https://doi.org/10.1007/s12665-016-5821-0>.
- Chen, Q., Wu, W., Blanckaert, K., Ma, J., & Huang, G. (2012). Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses. *Journal of Environmental Management*, 110, 116–124. <https://doi.org/10.1016/j.jenvman.2012.05.024>.
- 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, 2–16.
- Dong, J., Wang, G., Yan, H., Xu, J., & Zhang, X. (2015). A survey of smart water quality monitoring system. *Environmental Science and Pollution Research*, 22(7), 4893–4906. <https://doi.org/10.1007/s11356-014-4026-x>.
- Finotti, A. R., Finkler, R., Silva, M. D., & Cemim, G. (2009). Monitoramento de Recursos Hídricos em Áreas Urbanas. Caxias do Sul. RS: Educ. 272p.
- Gamble, A., & Babbar-Sebens, M. (2012). On the use of multivariate statistical methods for combining in-stream monitoring data and spatial analysis to characterize water quality conditions in the White River Basin, Indiana, USA. *Environmental Monitoring and Assessment*, 184, 845–875.
- Guigues, N., Desenfant, M., & Hance, E. (2013). Combining multivariate statistics and analysis of variance to redesign a water quality monitoring network. *Environmental Science: Processes and Impacts*, 15(9), 1692–1705. <https://doi.org/10.1039/C3EM00168G>.
- Hair, J. F. JR., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2005). *Análise Multivariada de dados*. Bookman: São Paulo, 5ª ed.
- Huang, F., Wang, X., Lou, L., Zhou, Z., & Wu, J. (2010). Spatial variation and source apportionment of water pollution in Qiantang river (China) using statistical techniques. *Water Research*, 44(5), 1562–1572. <https://doi.org/10.1016/j.watres.2009.11.003>.
- Instituto Mineiro de Gestão Das Águas – IGAM (2014). Monitoramento da Qualidade das Águas Superficiais no Estado de Minas Gerais: relatório trimestral (2º trimestre de 2014). Belo Horizonte. <https://goo.gl/uBwUjf>. Accessed 10 June 2016.
- Instituto Mineiro de Gestão das Águas (IGAM) (2013). Identificação de municípios com condição crítica para a qualidade de água na bacia do rio Paraopeba. Belo Horizonte. http://www.igam.mg.gov.br/images/stories/ARQUIVO_SANEAMENTO/estudo-saneamento-rio-paraopeba.pdf. Accessed 10 June 2016.
- İşçen, C. F., Altın, A., Şenoğlu, B., & Yavuz, H. S. (2009). Evaluation of surface water quality characteristics by using multivariate statistical techniques: a case study of the Euphrates river basin, Turkey. *Environmental Monitoring and Assessment*, 151(1–4), 259–264. <https://doi.org/10.1007/s10661-008-0267-9>.
- Khalil, B., Ouarda, T. B. M. J., & St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology*, 405(3–4), 277–287. <https://doi.org/10.1016/j.jhydrol.2011.05.024>.
- Lattin, J., Carroll, J. D., & Green, P. E. (2011). *Análise de Dados Multivariados*. São Paulo: Cengage Learning 455 p.
- Lee, C., Paik, K., Yoo, D. G., & Kim, J. H. (2014). Efficient method for optimal placing of water quality monitoring stations for an ungauged basin. *Journal of Environmental Management*, 132, 24–31. <https://doi.org/10.1016/j.jenvman.2013.10.012>.
- Markus, M., Knapp, H. V., & Tasker, G. D. (2003). Entropy and generalized least square methods in assessment of the regional value of streamgages. *Journal of Hydrology*, 283, 107–121. [https://doi.org/10.1016/S0022-1694\(03\)00244-0](https://doi.org/10.1016/S0022-1694(03)00244-0).
- Minas Gerais (2008). State Environmental Policy Council (COPAM) and State Water Resources Council (CERH). Normative Deliberation n. 01, from May 05, 2008. Provides for the classification of water bodies and environmental guidelines for their classification as well as establishes the conditions and standards for effluent discharge and other measures.
- Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG 297 p.
- Paliwal, R., Sharma, P., & Kansal, A. (2007). Water quality modeling of river Yamuna (India) using QUAL2E-UNCAS. *Journal of Environmental Management*, 83, 131–144. <https://doi.org/10.1016/j.jenvman.2006.02.003>.
- 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(3), 289–297.
- Shrestha, S., & Kazama, F. (2006). Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river, Japan. *Environmental Modeling and Software*, 22, 464–475. <https://doi.org/10.1016/j.envsoft.2006.02.001>.
- Soares, P. F. (2001). Projeto e avaliação de desempenho de redes de monitoramento de qualidade de água utilizando o conceito de entropia. Escola Politécnica da Universidade de São Paulo. [http://www.ib.usp.br/limnologia/textos/tde\[1\].pdf](http://www.ib.usp.br/limnologia/textos/tde[1].pdf). Accessed 20 June 2017.
- 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. <https://doi.org/10.1016/j.jenvman.2007.03.001>.
- Telci, I. T., Nam, K., Guan, J., & Aral, M. M. (2009). Optimal water quality monitoring network design for river systems. *Journal of Environmental Management*, 90(10), 2987–2998. <https://doi.org/10.1016/j.jenvman.2009.04.011>.
- Vieira, J. S., Pires, J. C. M., Martins, F. G., Villar, V. J. P., Boaventura, R. A. R., & Botelho, C. M. S. (2012). Surface water quality assessment of Lis River using multivariate statistical methods. *Water, Air, and Soil Pollution*, 223, 5549–5561. <https://doi.org/10.1007/s11270-012-1267-5>.

- Wang, C., Wang, Y., & Wang, P. (2006). Water quality modeling and pollution control for the eastern route of south to north water transfer project in China. *Journal of Hydrodynamics*, 18(3), 253–261. [https://doi.org/10.1016/S1001-6058\(06\)60001-2](https://doi.org/10.1016/S1001-6058(06)60001-2).
- Wu, M., Wang, Y., Sun, C., Wang, H., Dong, J., Yin, J., & Han, S. (2010). Identification of coastal water quality by statistical analysis methods in Daya Bay, South China Sea. *Marine Pollution Bulletin*, 60(6), 852–860. <https://doi.org/10.1016/j.marpolbul.2010.01.007>.
- Xu, H. S., Xu, Z. X., Wu, W., & Tang, F. F. (2012). Assessment and spatiotemporal variation analysis of water quality in the Zhangweinan river basin, China. *Procedia Environmental Sciences*, 13, 1641–1652. <https://doi.org/10.1016/j.proenv.2012.01.157>.
- Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yan, Q., Wang, Y., & Ye, F. (2009). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment*, 152, 123–131. <https://doi.org/10.1007/s10661-008-0301-y>.
- Zhao, G., Gao, J., Tian, P., Tian, K., & Ni, G. (2011). Spatial-temporal characteristics of surface water quality in the Taihu Basin, China. *Environment and Earth Science*, 64(3), 809–819. <https://doi.org/10.1007/s12665-011-0902-6>.