

Spatio-temporal variations of water quality in Nova Ponte Reservoir, Araguari River Basin, Brazil

Cristiano Christofaro, Mônica M. Diniz Leão, Sílvia M. A. C. Oliveira, Déborah T. Viana, Camila C. Amorim and Marcela D. Carvalho

ABSTRACT

Multivariate techniques, like cluster analysis (CA) and principal component analysis (PCA), were used to evaluate the spatial and temporal variability of surface water quality in a large neotropical hydroelectric reservoir (Nova Ponte Reservoir). The dataset, obtained during the period of 1995–2011 from the Energy Company of Minas Gerais State (CEMIG), contains 14 parameters surveyed quarterly in 14 sites, at different depths. The CA grouped the 14 sites in three main groups: lotic sites, half of the photic zone sites and bottom sites. Statistical tests showed that only three parameters (total dissolved solids, nitrate and chemical oxygen demand) have no significant difference between cluster groups. The PCA results showed temporal changes of the water quality in all groups, illustrating modifications to the importance of the parameters over time. PCA also revealed the major causes of water deterioration from 1995 to 2005 were related to agricultural and livestock activities. Currently, the water quality parameters related to organic pollution are also highlighted. Generally, this study shows that possible optimization of the monitoring network should consider temporal variation of water quality parameters.

Key words | environmental statistic, trend analysis, water monitoring, water quality

Cristiano Christofaro (corresponding author)
Department of Forestry Engineering,
Universidade Federal dos Vales do Jequitinhonha e
Mucuri,
Rodovia MG 367 n. 5000,
39100-000 Diamantina,
MG,
Brazil
E-mail: cristiano.christofaro@ufvjm.edu.br

Mônica M. Diniz Leão
Sílvia M. A. C. Oliveira
Déborah T. Viana
Camila C. Amorim
Department of Environmental and Sanitary
Engineering,
Universidade Federal de Minas Gerais,
Antônio Carlos Avenue, 6627,
31270-901 Belo Horizonte,
MG,
Brazil

Marcela D. Carvalho
Cemig Geração e Transmissão,
Av. Barbacena, 1200,
30190-131 Belo Horizonte,
MG,
Brazil

INTRODUCTION

The quality of surface waters reflects the interaction between natural agents (lithology, weathering, precipitation, and erosion) and anthropogenic agents (urbanization, industrial and agricultural activities, and water consumption) in the drainage area (Simeonov *et al.* 2003; Shrestha & Kazama 2007). Many studies have proven the connection between water quality and land use (Basima *et al.* 2006; Bottino *et al.* 2013).

Because of spatial and temporal variations in the concentrations of chemical species, a monitoring program must provide reliable and representative estimation of the water quality. A monitoring program will include water samples obtained from many sites at a regular frequency during a long interval of time, thus resulting in a large data matrix (Bartram & Ballance 1996). The resulting large datasets are often difficult to interpret, and the identification

of the possible causes that can influence the occurrence and concentration of a specific parameter is reasonably difficult (Simeonov *et al.* 2003).

Multivariate techniques like cluster analysis (CA) and principal component analysis (PCA) are widely used in water quality studies (Wang *et al.* 2014). CA is a group of multivariate techniques used to recognize discontinuous subsets in an environment which is discrete, but perceived as continuous (Borcard *et al.* 2011). CA is also carried out to identify any analogous behavior among different sampling stations or among measured variables in a dataset from a monitoring program (Mendiguchía *et al.* 2007; Shrestha & Kazama 2007).

PCA is a method for the visualization of complex data by dimensionality reduction (Hair *et al.* 2005). PCA can be

used to reduce the variable numbers with minimal loss of total variance (Zhao *et al.* 2012). In recent years, the PCA technique has been applied in various water quality studies (Singh *et al.* 2004; Krishna *et al.* 2009; Tobiszewski *et al.* 2010; Gatica *et al.* 2012; Zhao *et al.* 2012). The method can be applied to evaluate both spatial and temporal gradients, forming groups along the series (Borcard *et al.* 2011). However, PCA is rarely applied to the temporal trends in water quality data.

The use of multivariate techniques and data reduction is almost mandatory to achieve satisfactory results (Shrestha & Kazama 2007). Multivariate exploratory analyses are an important tool, which can characterize and evaluate river and lake water quality. It can provide evidence of the temporal and spatial variations caused by natural processes or pollutants (Singh *et al.* 2004; Wang *et al.* 2014). Multivariate methods can also be used to optimize the number and the respective locations of monitoring sites, thus reducing

datasets and costs (Simeonov *et al.* 2003; Shrestha & Kazama 2007; Wang *et al.* 2014).

In the present study, a large data matrix, obtained during a 17-year (1995–2011) monitoring program, is subjected to different multivariate statistical techniques to evaluate the spatial and temporal dynamics of the water quality of a neotropical hydroelectrical reservoir (Nova Ponte Reservoir). The objective is to extract the most information from the smallest number of analyses and to evaluate the property of the multivariate models generated.

MATERIALS AND METHODS

Study area

The Nova Ponte Hydroelectric Plant, located in Minas Gerais state (Brazil), in the Araguari River Basin (Figure 1), has a

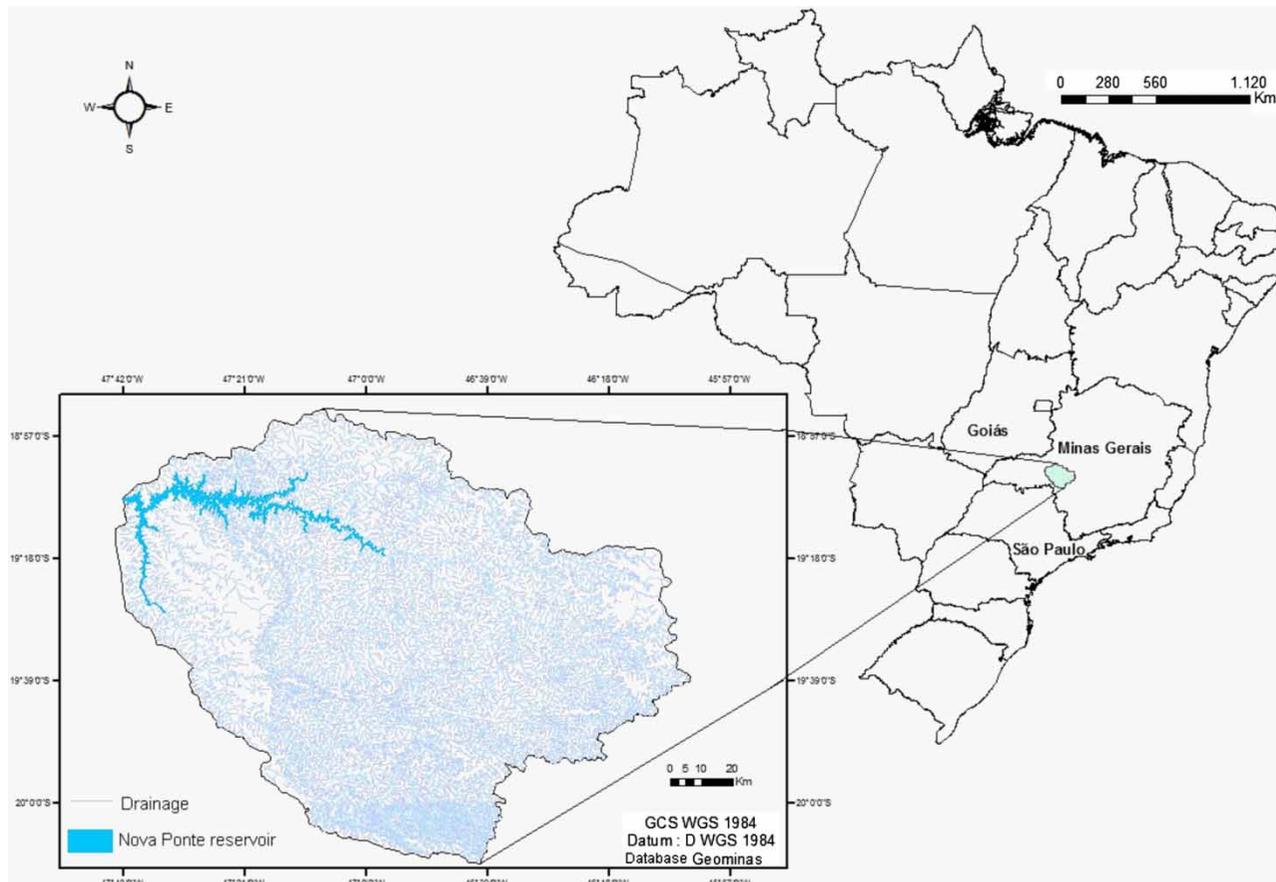


Figure 1 | Nova Ponte Reservoir location, Minas Gerais State, Brazil.

generation capacity of 510 MW and 443 km² of water surface. The region is mainly occupied by annual crops such as soybeans, coffee and sugarcane with a high anthropization degree and a few natural vegetation areas of Cerrado and semideciduous forest in riparian areas. The soils have low natural fertility; therefore, liming and fertilization to maximize agricultural potential is necessary. The main activities are: fertilizer industry, metallurgy, slaughter houses, poultry and fishing. The mining activity is due to niobium, titanium and phosphate sources (IGAM 2010). In this area are two of the largest cities in the state, Uberlândia and Araxá, which are strong sources of sanitary pollution.

Monitoring water quality data

Water quality data from 14 stations in lotic and lentic conditions and different depths (HPZ – half of the photic zone; B – bottom samples), from 1995 to 2011, were selected (Table 1 and Figure 2). The datasets include 14 water quality parameters, monitored quarterly: temperature, turbidity, electrical conductivity, pH, dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, Fe, P, NO₃, ammonia nitrogen (NH₃-N), total dissolved solids (TDS), total suspended solids (TSS) and total solids (TS). The water quality parameters with units and basic statistics are summarized in Table 2.

Table 1 | Basic description of the 14 monitoring stations of Nova Ponte Reservoir, Minas Gerais, Brazil

Station code	Description	Condition	Depth
NP006	Araguari River	Lotic	S
NP021	Quebra Anzol River	Lotic	S
NP025	Capivara River	Lotic	S
NP110	Downstream reservoir	Lotic	S
NP140	Reservoir, near dam	Lentic	HPZ, B
NP150	Reservoir, next to the Araguari River	Lentic	HPZ, B
NP170	Reservoir, next to the Quebra Anzol River	Lentic	HPZ, B
NP180	Reservoir, next to the Santo Antônio River	Lentic	HPZ, B
NP200	Reservoir, next to the Capivara River	Lentic	HPZ, B

S = Superficies; HPZ = Half Photic Zone; B = Bottom.

Multivariate statistical methods

The multivariate analyses of water quality data were performed through CA and PCA. In both cases, all data were normalized to avoid the effects of scale in the classification. Data that were lower than the detection limit were substituted by the limit of detection. Both techniques require arbitrary choices and the application of confirmatory methods by the users (Borcard *et al.* 2011). For example, in CA, users must choose between different clustering methods and then choose the number of clusters. PCA techniques require the selection of optimal numbers of components. However, the choices of components are typically based on subjective criteria, and the appropriateness tests for multivariate models are rarely performed in studies of water quality. In this study, statistical criteria to reduce the identified subjectivity in multivariate models are presented below.

CA was performed using the Euclidean distance to measure similarity between samples. Once CA is a heuristic procedure, the choice of the clustering method can influence the final composition of the groups (Borcard *et al.* 2011). The methods of clustering ‘single’, ‘complete’, ‘average’ and ‘ward’ were tested by the cophenetic correlations. The cophenetic distance between two objects in a dendrogram is the distance at which the two objects become members of the same group (Borcard *et al.* 2011). The cluster method with the highest cophenetic correlation was chosen. The optimal number of groups was defined by the Rousseeuw quality index and confirmed in a silhouette plot of the memberships (Borcard *et al.* 2011). Kruskal–Wallis test was carried out to test for significant differences between groups for each parameter ($p < 0.05$). Boxplots were generated to show the variability of each water quality parameter on the groups and to compare the concentrations with the Brazilian water quality standard (CONAMA, NEC 2005).

The PCA test was performed on the groups of the CA, considering the semi-annual mean calculated for each parameter. To examine the suitability of the data for PCA, the Kaiser-Meyer-Olkin (KMO) (also MSA) index and Bartlett’s test were performed (Hair *et al.* 2005). The KMO index compares the values of correlations between variables and those of the partial correlations. If the KMO index is high (>0.5), the PCA can act efficiently; if KMO is low (<0.5), the PCA is not relevant. For each CA group the KMO index was calculated

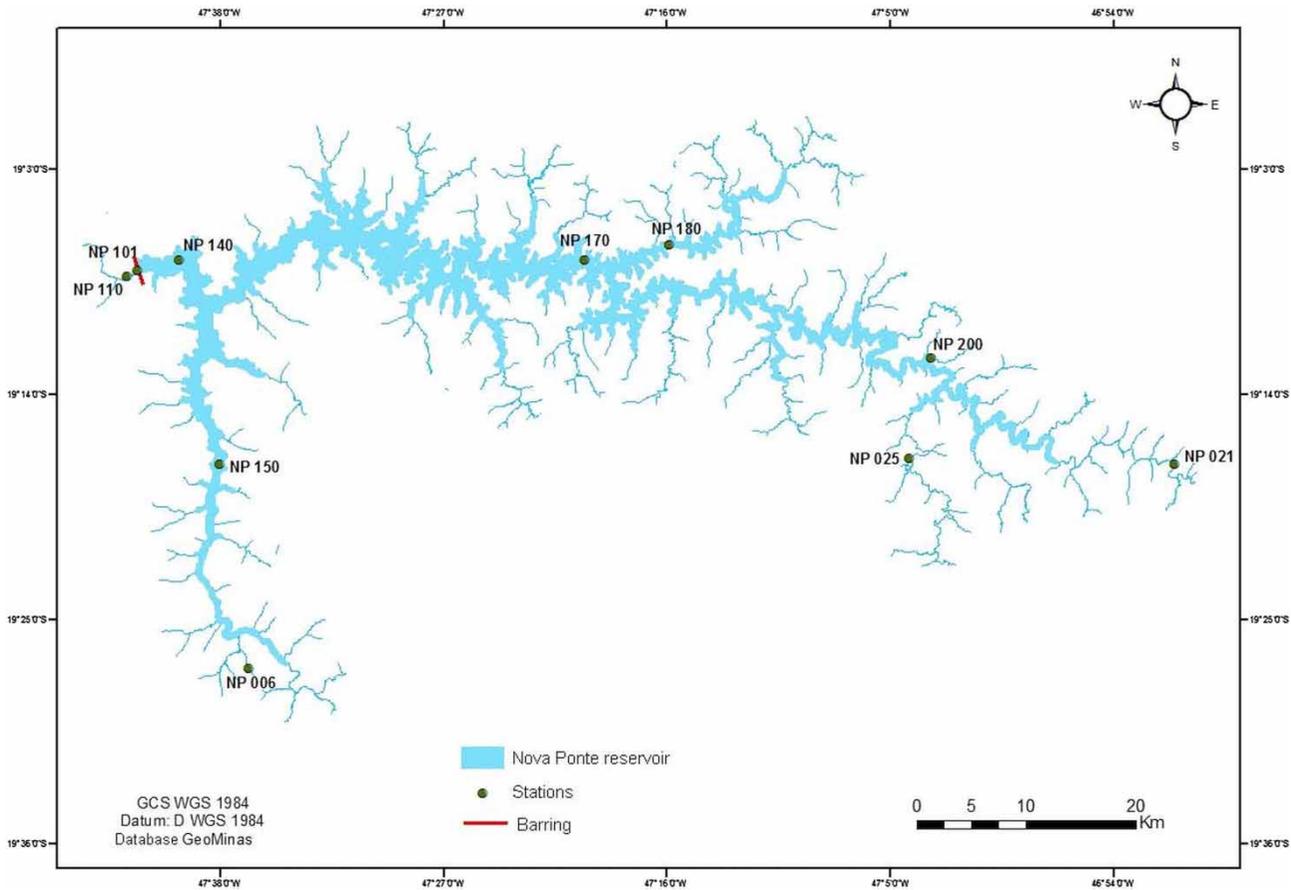


Figure 2 | Location of the sampling stations in the Nova Ponte Reservoir, Minas Gerais, Brazil.

Table 2 | Water quality parameters and descriptive statistics of the Nova Ponte Reservoir, Minas Gerais, Brazil

Parameters	Unit	Abbreviation	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Temperature	°C	T	15.30	21.40	23.20	23.22	25.40	30.50
Turbidity	NTU	Turb	0.50	2.00	5.00	19.97	13.00	918.00
Electrical conductivity	$\mu\text{S}\cdot\text{cm}^{-1}$	EC	8.00	20.00	24.00	31.55	32.00	250.00
pH		pH	5.30	6.50	6.80	6.89	7.20	9.00
Alkalinity	$\text{mg}\cdot\text{l}^{-1}$	Alk	20.00	20.00	20.00	22.12	21.00	54.00
Biochemical oxygen demand	$\text{mg}\cdot\text{l}^{-1}$	BOD	2.00	2.00	2.00	2.39	2.00	17.80
Chemical oxygen demand	$\text{mg}\cdot\text{l}^{-1}$	COD	5.00	5.00	6.50	12.98	12.00	323.00
Iron	$\text{mg}\cdot\text{l}^{-1}$	Fe	0.10	0.10	0.50	2.06	2.00	42.50
Phosphorus	$\text{mg}\cdot\text{l}^{-1}$	P	0.01	0.01	0.02	0.21	0.05	15.00
Nitrate	$\text{mg}\cdot\text{l}^{-1}$	NO_3	0.10	0.10	0.10	0.76	0.20	40.00
Ammonium	$\text{mg}\cdot\text{l}^{-1}$	NH_3	0.10	0.10	0.10	0.50	0.30	18.00
Total dissolved solids	$\text{mg}\cdot\text{l}^{-1}$	TDS	3.00	15.00	30.00	36.62	45.00	208.00
Total suspended solids	$\text{mg}\cdot\text{l}^{-1}$	TSS	3.00	3.00	6.00	29.06	20.00	1,030.00
Total solids	$\text{mg}\cdot\text{l}^{-1}$	TS	3.00	27.00	39.00	66.91	73.00	1,088.00

and the variables with individual KMO < 0.5 were removed. The PCA was done when the KMO of each variable and the overall were > 0.5. Biplots were generated to show the temporal variation of water quality parameters and samples over the monitoring. The variables with vectors longer than the radius of the draw circle have a greater than average contribution (Borcard *et al.* 2011). The distribution of samples on the biplots was analysed to find the temporal tendencies of the water quality parameters.

RESULTS AND DISCUSSION

Site similarity and dissimilarity

The result of CA is shown in Figure 3. The average clustering method with the highest cophenetic correlations (0.80) was chosen after comparing the different methods. All the 14 sampling sites in the study area were grouped into three clusters: cluster I (sites NP170HPZ, NP140HPZ, NP180HPZ, NP150HPZ, NP200HPZ, NP110), cluster II (sites NP025, NP200B, NP006, NP021) and cluster III (sites NP170B, NP180B, NP140B, NP150B). These clusters of sampling stations indicated that the HPZ, B and lotic samplings had similar water quality and can be grouped. The statistical distributions of the water quality parameters for each cluster are shown in Figure 4. Except for NO₃, TDS and COD, all

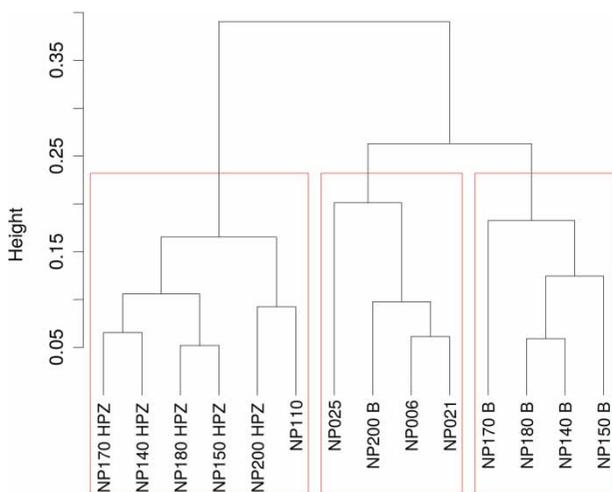


Figure 3 | Dendrogram showing clustering of sampling stations in Nova Ponte Reservoir, Minas Gerais, Brazil.

parameters showed significantly different concentration values between clusters ($p < 0.05$, Kruskal–Wallis test).

Group I is formed predominantly by superficial sampling in lotic stations, except for NP200B. The inclusion of the NP200 point in this group may be explained by their position further upstream of the reservoir (Figure 2), so presenting characteristics that are closer to the lentic points. Group I presented the highest median values for Turbidity, pH, COD, P, TSS and TS (Figure 4), indicating the highest anthropogenic influence, predominantly by diffuse pollution sources. The observed values can be explained by surface runoff from areas where the predominant activities are agriculture and livestock. Group II is formed predominantly by samples obtained at half the photic zone (HPZ) (Figure 3), except for one lotic station in the discharge area of the reservoir (NP110) (Figure 2). Once this point is located after the discharge of the reservoir, there may be an availability of deeper water on the surface. Group II presented the highest value for T, but the lowest median concentrations for Turbidity, Fe, P, TSS and TS (Figure 4), indicating lower levels of anthropogenic influence. Group III is formed only by bottom samples (Figure 3), and showed the highest median concentrations for EC, BOD, Fe, NH₃ and Alkalinity (Figure 4), indicating the contribution of sediments and an intermediate anthropic influence. Except for Fe in groups I and III, only outlier samples showed parameter levels above those established by the Brazilian Environmental Standards (CONAMA, NEC 2005) (Figure 4). The higher levels of Fe in lotic and bottom stations indicate the deposition of this metal in sediment and should be investigated.

The CA results revealed that this approach is useful in offering reliable classifications of surface waters in the whole region and optimizing the design of a future spatial sampling strategy (Simeonov *et al.* 2003). Thus, for quick spatial assessments of water quality, the number of the sampling site could be optimized, and only a representative site sampled from each cluster could be sufficient to determine the water quality of the entire region (Simeonov *et al.* 2003).

Data structure determination and temporal analysis

PCA was performed on each CA group to identify linkages between the parameters. After the KMO criteria, ten parameters were used in group I, 11 in group II and ten in

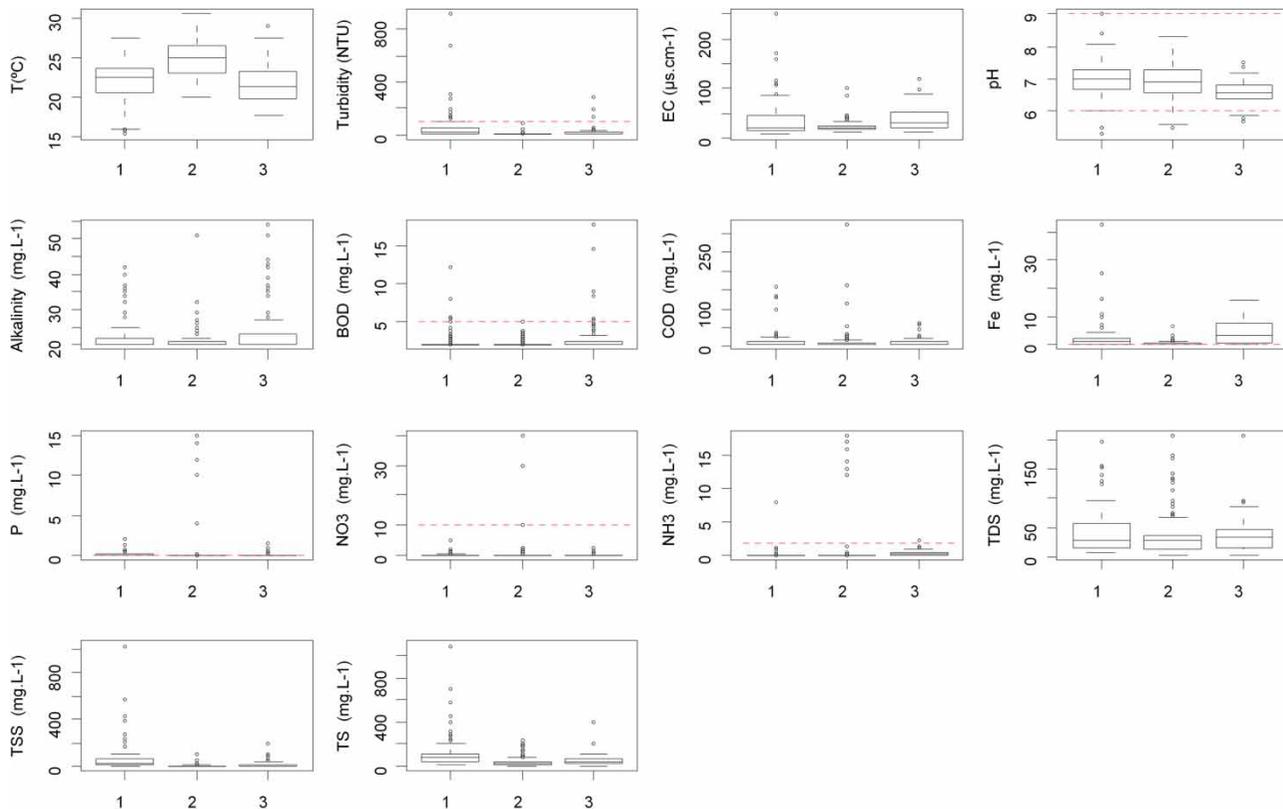


Figure 4 | Variation of the monitored parameters in the three groupings after hierarchical CA in Nova Ponte Reservoir, Minas Gerais, Brazil. The horizontal dotted lines represent the limits of the Brazilian water quality standards (CONAMA, NEC 2005) for the reservoir.

group III, of the original set of 14 parameters, for the PCA analysis. Based on the eigenvalues <1 , four main components were found in groups I and II, and three main components in group III, explaining 85%, 73% and 76% of the total variance, respectively.

The temporal variation in water quality parameters was also assessed by PCA biplots. The biplots (Figure 5) show temporal changes of quality parameters for each group. In group I, there are two atypical sampling periods: 2002-2, with a stronger contribution of TDS and Alk; and 1995-1, with a moderate contribution of BOD and NH_3 . The Turbidity has greatest values for the 2003-2 and 2009-2 samples. In group I, the measured values of the parameters are higher in the rainy season relative to the dry season. The most recent samples were distributed on the opposite side of most of the variables, indicating a general improvement in the water quality in the monitoring period.

In group II, between 1995 and 1999, TS, TDS, NO_3 , NH_3 and COD have the higher concentrations. NH_3 has

higher concentration in the 1998-2 sample, indicating an increased contribution of sewage and industrial sources in this period. Between 2002-1 and 2003-2 there is an increasing concentration of Alkalinity and EC, indicating the influence of diffuse sources. The most recent samples were distributed on the opposite side of most of the variables, indicating a general trend of improving water quality in the evaluation period, except for 2011-1, when Alkalinity increases again. In group III, during 1995–1999 water quality was influenced by EC, TDS, TS and TSS. The four parameters are related to diffuse sources. Between 2001 and 2005 the importance of the parameters was reduced, with a moderate increase in concentration of COD and Alkalinity. The observed alteration may be related to a greater contribution of industrial pollution. The most recent samples (2000-2 to 2009-2) were set on the opposite side of most of the variables, except T and NO_3 , indicating a general improvement in water quality in the monitoring period.

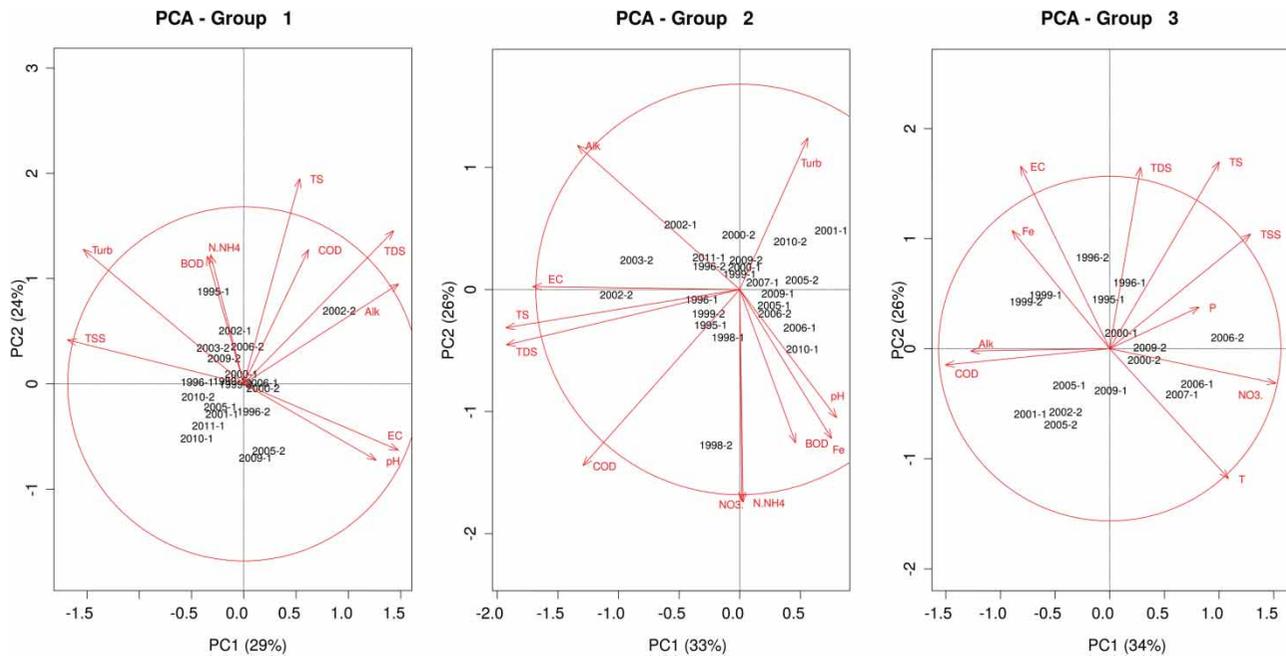


Figure 5 | Principal components biplots for each cluster (groups) of water quality parameters in Nova Ponte Reservoir, Minas Gerais, Brazil.

The PCA grouped the sampling points into some main groups, related to temporal changes. Overall, the analysis indicates that the stations showed an improving trend during the study period, with some atypical periods related to specific quality parameters. Concentrations of solids and nutrients are factors to be observed and should be monitored frequently in all stations and groups. In this study, the PCA technique applied to studies of temporal variations allows a satisfactory assessment of trends of the parameters over time. [Gatica *et al.* \(2012\)](#) demonstrate the applicability of PCA in temporal characterizations of monitoring data, however emphasizing seasonal differences. The use of discriminant analysis to show differences between seasons is also frequent ([Singh *et al.* 2004](#); [Shrestha & Kazama 2007](#); [Yang *et al.* 2010](#)). In our approach, both seasonal and trend information could be characterized in PCA biplots. Although it is an exploratory analysis, these results show that a possible optimization of the monitoring network should consider the temporal variations and the importance of water quality parameters per time. The applied technique can be used as a screening analysis, and a trend analysis ([Hirsch *et al.* 1982](#)), by parameters, can be used to confirm the results.

CONCLUSIONS

Monitoring of water quality generates complex data that require the application of multivariate techniques to facilitate their interpretation. In this study, CA enabled the classification of samples into three major groups consisting of samples of depth: lotic and lentic environments. The information extracted from CA was used to characterize groups according to the water quality parameters, as well as define possible predominant anthropogenic influences. The application of the PCA technique in the groups defined in CA allowed a significant reduction of the data, in groups II and III. About 75% of the variation was explained by the first four principal components, and for group III about 85% of the variability was explained by the first three principal components. The PCA allowed this study to infer the temporal importance of quality parameters in the CA groupings of the 14 monitoring stations and also the specific periods of water quality changes, indicating an improvement in water quality over the 17-year research period. Thus, the multivariate techniques used proved to be useful in the interpretation of monitoring data of the water quality of the reservoir, allowing a better understanding of their temporal and spatial variations.

ACKNOWLEDGEMENTS

The authors thank the Energy Company of Minas Gerais State (CEMIG), the National Electric Energy Agency (ANEEL), the Foundation for Research Support of the State of Minas Gerais – FAPEMIG, the Coordination for the Improvement of Higher Education Personnel – CAPES, and the National Council of Technological and Scientific Development – CNPq for their financial support.

REFERENCES

- Bartram, J. & Balance, R. 1996 Introduction. In: *Water Quality Monitoring: A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes* (J. Bartram & R. Balance, eds). CRC Press, New York, USA, pp. 1–8.
- Basima, L. B., Senzanje, A., Marshall, B. & Shick, K. 2006 Impacts of land and water use on plankton diversity and water quality in small man-made reservoirs in the Limpopo Basin, Zimbabwe: a preliminary investigation. *Physics and Chemistry of the Earth* **31** (15–16), 821–831.
- Borcard, D., Gillet, F. & Legendre, P. 2011 Cluster analysis. In: *Numerical Ecology with R*. Springer, New York, USA, pp. 53–114.
- Bottino, F., Calijuri, M. C. & Murphy, K. J. 2013 Temporal and spatial variation of limnological variables and biomass of different macrophyte species in a neotropical reservoir (São Paulo, Brazil). *Acta Limnologica Brasiliensia* **25** (4), 387–397.
- CONAMA, National Environmental Council 2005 Resolução nº 357/2005. *Environmental Guidelines for Water Resources and Standards for the Release of Effluents* (in Portuguese). Imprensa Oficial, Brasília, Brazil.
- Gatica, E. A., Almeida, C. A., Mallea, M. A., Corigliano, M. C. & González, P. 2012 Water quality assessment, by statistical analysis, on rural and urban areas of Chocancharava River (Río Cuarto), Córdoba, Argentina. *Environmental Monitoring and Assessment* **184** (12), 7257–7274.
- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. 2005 *Multivariate Data Analysis*, 5th edn. Prentice-Hall, Upper Saddle River, NJ, USA.
- Hirsch, R. M., Slack, J. R. & Smith, R. A. 1982 Techniques of trend analysis for monthly water quality data. *Water Resources Research* **18** (1), 107–121.
- Instituto Mineiro de Gestão das Águas – IGAM 2010 *Monitoramento das Águas Superficiais na Bacia do Rio Paranaíba 1998–2009 (Monitoring of Surface Waters in the Paranaíba River Basin 1998–2009)*. Belo Horizonte-MG, Brazil.
- Krishna, A. K., Satyanarayanan, M. & Govil, P. K. 2009 Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: a case study from Patancheru, Medak District, Andhra Pradesh, India. *Journal of Hazardous Materials* **167** (1–3), 366–373.
- Mendiguchía, C., Moreno, C. & García-Vargas, M. 2007 Evaluation of natural and anthropogenic influences on the Guadalquivir River (Spain) by dissolved heavy metals and nutrients. *Chemosphere* **69** (10), 1509–1517.
- Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling & Software* **22** (4), 464–475.
- Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsas, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, Th. 2003 Assessment of the surface water quality in Northern Greece. *Water Research* **37** (17), 4119–4124.
- 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** (18), 3980–3992.
- Tobiszewski, M., Tsakovski, S., Simeonov, V & Namieśnik, J. 2010 Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. *Chemosphere* **80** (7), 740–746.
- Wang, Y. B., Liu, C. W., Liao, P. Y. & Lee, J. J. 2014 Spatial pattern assessment of river water quality: implications of reducing the number of monitoring stations and chemical parameters. *Environmental Monitoring and Assessment* **186** (3), 1781–1792.
- Yang, Y. H., Zhou, F., Guo, H. C., Sheng, H., Liu, H., Dao, X. & He, C. J. 2010 Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environmental Monitoring and Assessment* **170** (1–4), 407–416.
- Zhao, Y., Xia, X. H., Yang, Z. F. & Wang, F. 2012 Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. *Procedia Environmental Sciences* **13** (18th Biennial ISEM Conference on Ecological Modelling for Global Change and Coupled Human and Natural System), 1213–1226.

First received 25 October 2016; accepted in revised form 21 March 2017. Available online 4 April 2017