Assessment of water quality using principal component analysis: a case study of the açude da Macela – Sergipe – Brazil

Carlos Alexandre Borges Garcia[1]. Helenice Leite Garcia[1]. Maria Caroline Silva Mendonça[1]. Anamália Ferreira da Silva[1]. José do Patrocínio Hora Alves[1]. Silvânio Silvério Lopes da Costa[1]. Geovanny Oliveira Araújo[2] and Igor Santos Silva[1]

[1]Universidade Federal de Sergipe. Av. Marechal Rondon. s/n - Bairro Rosa Elze - São Cristóvão - SE – Brazil; e-mail: cgarcia@ufs.br.

[2]Universidade Federal da Bahia. Rua Barão de Jeremoabo. 147 Campus Universitário de Ondina. Salvador - BA - Brasil.

ABSTRACT

The quality of surface water is a very sensitive issue and it is a great environmental concern worldwide. In recent years. there has been an increase in awareness and concern about water pollution across the globe. Thus. new approaches towards achieving sustainable water resources management have been developed internationally. In present study multivariate statistical approaches are used; interpretation of large and complex data matrix obtained during a monitoring of the Açude da Macela. which is located in Itabaiana. Sergipe. Brazil and used for human consumption and irrigation of vegetables. Samplings were done on selected sites for two years (2010–2012) across in the reservoir width with a view to monitor changes caused by anthropogenic sources. Sampling. preservation and transportation of the samples to the laboratory were done in concordance with standard methods.

Keywords: Water quality. PCA. Açude da Macela

1. INTRODUCTION

The quality of surface water is a very sensitive issue and it is a great environmental concern worldwide. It is critical for long-term economic development. social welfare. and environmental sustainability. In recent years. there has been an increase in awareness and concern about water pollution across the globe. Thus. new approaches towards achieving sustainable water resources management have been developed internationally.

In view of this. the water quality index (WQI) is considered a key element in the sound management of water resources. as it can be used to simplifying expressions of a complex set of pollution variables in the rivers. streams and lakes of both developed and developing countries. Generally. WQI is a dimensionless number which combines multiple water quality factors into a single number by normalizing values to subjective rating curves and enabling easy interpretation of monitoring data. Conventionally. normalization of variables such as dissolved oxygen (DO). pH. nutrients (nitrogen and phosphorus). etc.. has been used to evaluate the quality of water separately. depending upon the designated water uses of the water body and local preferences (Chaturvedi *et al*. 2010).

However. analysis. including a number of parameters grouped according to common features. can provide partial information about the overall water quality. The incorporation of different parameters on a single number makes interpretation through traditional approaches even more difficult. Although the computational mathematical modeling of the water quality river or reservoir is useful to evaluate the overall quality. the application of the models are often limited by the prior knowledge of hydrodynamics and extensive validation (Lobato et al. 2015).

A variety of water quality indices have been designed to judge out the overall water quality within a particular area promptly and efficiently. Some examples of these are the US National Sanitation Foundation Water Quality Index (NSFWQI). the Environmental Institute of Paraná (RWQI). Oregon Water Quality Index (OWQI) to evaluate the general water quality of Oregon's stream; and the Central Pollution Control Board of India (CPCB-WQI). These indices measure the quality of water from its reservoirs due to the factors that influence them and. therefore. the parameters adopted for each assume weights and different importance levels in the evaluation of water quality.

As the indicator has the function to simplify. other techniques have been applied to identify some informative content which may not have been properly analysed. thus influencing the safe management of water resources. So. some multivariate statistical techniques have been used to assist the monitoring of water quality. formulating a rapid response to aquatic pollution. Among these. the Principal Component Analysis (PCA) is an analytical methodology used commonly in the scientific community as it allows reducing the dimensionality of a data set. while maintaining the characteristics of variables which contribute most to this variation.

In this context. the aim of this study is to evaluate the water quality of the Açude da Macela using the Principal Component Analysis to assist interpretation and extraction of the most important parameters for the assessment of variations in water quality of this reservoir.

2. MATERIALS AND METHODS

2.1 Study Area

This case study will concern the Açude Macela resevoir. which is used for human consumption and irrigation of vegetables.

The perimeter of Açude da Macela is located in the urban area of the municipality of Itabaiana. as shown in Figure 1. The reservoir has a storage capacity of 2.710.000 m³ and was designed to provide irrigation water for 156 hectares and be an area used for the production of vegetables that supplies the markets of Itabaiana and the surrounding region. The water from the dam presents improper quality for irrigation. possibly due to the agricultural evictions (pesticides). which are drained into this area through the rivers or ground water. and domestic and industrial sewage discharged into the reservoir without any treatment.

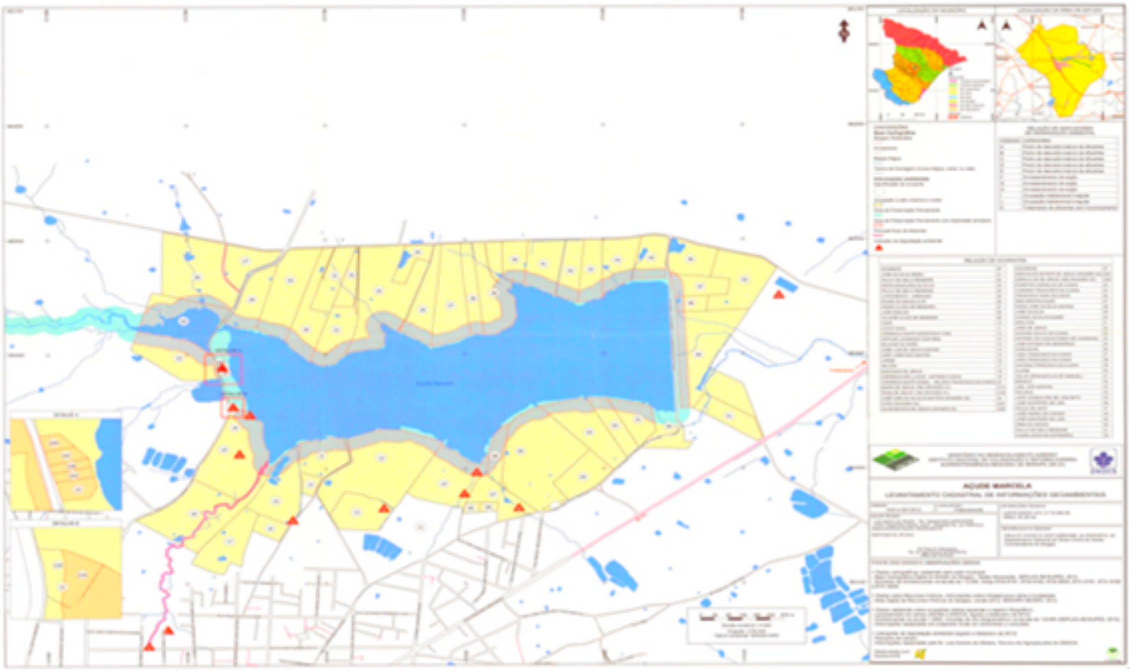

Figure 1- Açude da Macela reservoir

Source: DNOCS (2012)

2.2 Sampling

Samplings were done on selected sites for two years (2010–2012) across in the reservoir width with a view to monitor changes caused by anthropogenic sources. Sampling. preservation and transportation of the samples to the laboratory were done in concordance with standard methods (APHA. 2012). Eleven physico-chemical parameters have been determined by prescribed standard methods.

2.3 Data Treatment

Before data can be analyzed. it was performed the standardization of these due to the different ranges of each environmental variable. since the dimensions used to compute the distance between the objects should be of a similar magnitude. To reduce the dimensionality of the data set and minimize the loss of information. it was carried out the sorting the data set. The raw data was converted into dimensionless values represented by mean equals zero and variance equals one. subtracting the average of each variable data set and dividing them by the standard deviation.

Then. the linear correlation between variables was assessed using Pearson product-moment correlation coefficient. which measures the degree of correlation and direction of it (positive or negative).

Principal component analysis (PCA) aims to find combinations for certain variables to determine indices which describe the variation in the data with minimal loss of information. According Manly (2005). the goal of this analysis is to take X variables X1. X2. .... XP. finding combinations of these indices to produce Z1. Z2. .... Zp. which are uncorrelated in order of importance and which describe the variation in the data. The lack of correlation means that indexes are measuring different dimensions of the data. and the order is such that Var (Z1) ≥ Var (Z2) ≥. . . ≥ Var (Zp). where Var (Z1) denotes the variance (Z1). Z indices are. then. the principal components. The eigenvalues of the main components are a measure of their associated variances. and the sum of them coincides with the total number of variables. The analyzes were determined with the aid of R Project for Statistical Computing.

2.4 Application of Water Quality Indexes

From the physical. chemical and biological parameters obtained. it was possible to calculate the water quality indexes. described in Table 1 for the Açude da Macela. Since the quality parameters required for the calculation of WQI had not been analyzed. its application was impossible. Moreover. it is important to emphasize the WQI proposed is for rivers without dams. which is not the case of the studied ecosystem.

Table 1 – Indexes used to assess the water quality of Açude da Macela.

Index	Parameters	*wi*	Equação	Classification
O-WQI (Dunnette. 1979; Cude. 2001)	Temperature DO pH Ammonia + nitrate Total phosphorus total solids chlorophyll *a*		$\sqrt{\frac{N}{\sum_{i=2}^{n}\frac{1}{q_i^2}}}$	0-25 = very bad 26-50 = bad 51-70 = reasonable 71-90 = good 91-100 = excellent
WQIR (IAP. 2014)	DO Deficit Total P total inorganic N chlorophyll a Secchi depth COD Cyanobacteria Time residences Depth	17 12 8 15 12 12 8 10 6	$IQAR = \frac{\sum(q_i w_i)}{\sum w_i}$	0 - 1.50 = Not Impacted 1.51 - 2.50 = bit run down 2.51 - 3.50 = moderately degraded 3.51 - 4.50 = polluted 4.51 – 5.50 = very polluted > 5.51= extremely polluted
CPCB-IQA (Sarkar & Abbasi. 2006)	DO pH chlorophyll *a*	0.31 0.22 0.28	$\sum_{i=1}^{N} q_i w_i$	0-38 = very bad 38-50 = bad 50-63 = reasonable/good 63-100 = good/excellent

Source: Abassi *et al* (2012).

Where: N - number of water quality variables; q*i* - quality class of water in relation to the variable "*i*"; w*i* - weights calculated for the variable "*i*".

It is important to mention that. for the calculation of indexes O-WQI. WQIR and CPCB-WQI. some variables can be deleted. as well as others can be inserted. such as chlorophyll which is indicative of the presence of algae. therefore. it can be included in the calculation of the indexes. The water quality parameters used in the index calculations were analyzed according to their arithmetic mean. which were obtained from the R Project for Statistical Computing.

Starting from the understanding that pollution is a dynamic process. some environmental variables can be considered critical from a certain time. and. because of that. the WQIR can be modified. or new environmental variables can be incorporated into the index or may occur replacement of its variables.

Thus. as it can be seen in Table 2. from the principal component analysis. new weights were assigned to the variables considered critical according to their contributions to the determination of each principal component. Then. it was performed the arithmetic mean calculation of percentage contribution of each variable and the result was an approximate figure for assigning new weights used in calculating the WQIR-m.

Table 2- Weights obtained through the contribution (%) of each parameter for PCs.

Parameters	w*i*
Chlorophyll *a*	14.000
Colour	8.000
$N\text{-}NO_3$	6.000
$N\text{-}NO_2$	12.000
$N\text{-}NH_4$	10.000
DO	5.000
P-PO4	4.000
pH	10.000
Conductivity	7.000
SS	11.000
Total solids	7.000
Water temperature	6.000

The WQIR-m was determined following the same calculation method proposed by the Environmental Institute of Paraná. as well as their classification. The result of this calculation was compared to the other indices in order to minimize subjectivity and improve the reliability of the final assessment.

3. RESULTS AND DISCUSSION

Through an analysis of the correlation matrix shown in Table 3. it was possible to verify the association between the environmental variables - we were able to obtain an overview of the data set. thus enabling us to identify the variables which have greater significance to this study. Adopting a 5% level of significance. it is noted that. according to the Pearson coefficient. out of 12 pairs between the data set. only those variables with correlations in module equal or higher than r = 0.50 are significant (Triola. 1999).

Table 3- Correlation matrix.

Variáveis	Chlor. a	Colour	N-NO_3	N-NO_2	N-NH_4	DO	P-PO_4	pH	Cond	SS	Total solids	W temp.
Chlor. a	**1**											
Colour	**0.770**	**1**										
N-NO_3	**0.527**	0.006	**1**									
N-NO_2	**-0.526**	**-0.524**	0.431	**1**								
N-N H_4	**-0.527**	**-0.565**	**-0.517**	-0.288	**1**							
DO	**0.641**	-0.231	0.498	-0.098	-0.398	**1**						
P-PO4	**-0.730**	**-0.770**	-0.475	**-0.594**	**0.728**	**-0.569**	**1**					
pH	-0.205	-0.424	0.406	0.305	-0.173	-0.150	**-0.566**	**1**				
Cond	-0.311	-0.152	0.488	0.352	0.241	-0.033	**0.563**	**-0.575**	**1**			
SS	**0.667**	**–0.607**	0.392	-0.307	-0.203	0.374	-0.406	-0.194	**0.652**	**1**		
Total S	-0.203	-0.395	-0.205	-0.209	-0.043	-0.137	0.135	0.179	**0.505**	-0.395	**1**	
W temp.	-0.189	-0.123	0.156	-0.005	0.083	-0.236	0.131	0.063	0.229	-0.196	-0.034	**1**

Only some parameters showed significant correlation relations. High correlation can be observed between chlorophyll a and colour. nitrate. nitrite. ammonia nitrogen. dissolved oxygen. phosphate and suspended solids. indicating an eutrophication process. The variable colour is also correlated with the nitrogenous compounds. phosphates and suspended solids. indicating the presence of ions in solution. such as iron and manganese; as well as decomposition of organic matter in water by algae or by discharges from industrial and household waste. It is also possible to observe the negative correlation of the dissolved oxygen with pH and water temperature as the solubility of oxygen decreases as the oxygen is used to decompose the organic matter present in the medium. It is possible the seasonal fluctuations may be responsible for this kind of correlation.

Table 4 – Descriptive statistics

Parameters	Mean	Desvio Padrão	Amostras
Conductivity (mS cm^{-1})	1.67	0.63	123
Colour (Pt-Co)	21.48	13.28	123
pH	8.56	0.39	123
Total solids (mg L^{-1})	1058.02	339.64	123
Suspended solids (mg L^{-1})	23.06	24.79	123
Dissolved oxygen (mg L^{-1})	5.32	2.85	123
N-NH_4 (µg L^{-1})	48.19	39.66	123
N-NO_2 (µg L^{-1})	146.92	146.31	123
N-NO_3 (µg L^{-1})	1255.96	615.88	123
P-PO_4 (mg L^{-1})	415.14	200.30	123
Chlorophyll a (µg L^{-1})	59.28	63.20	123
Water Temperature (ºC)	28.23	1.82	123

In Table 4. the basic descriptive statistics are presented. Chlorophyll a is used as an indicator of phytoplankton biomass in aquatic environments because it is a pigment found in all plant groups and other autotrophic organisms. Its concentration helps in the interpretation of results of physical and chemical analysis.

also it is indicative of the physiological state of phytoplankton and the degree of eutrophication of aquatic environment.

According to Tundisi *et al* (2002). the extent of eutrophication of lakes. rivers and reservoirs is based on the concentrations of nitrogen. phosphorus and chlorophyll a. which is indicative of the degree of eutrophication of aquatic environment. According to the parameters analyzed in this study. it is observed that phosphate. nitrogen compounds and chlorophyll a concentrations are above the limits set by CONAMA Resolution 357/05 to the Açude da Macela.

Table 5 shows the principal components (PCs) and their eigenvalues. and the percentage of variance of each PC. Figure 2 shows the scree plot of the eigenvalue for each component. yielding five PCs eigenvalues> 1. adding 72% of the total variance in the dataset. The Scree Plot shows a marked marked change of slope from the first to the second eigenvalue because PC1 is responsible for 25.57% of the water quality variation in the weir Macela. PC2 is responsible for 13.60%. PC3 for 12.36%. PC4 for 10.61% and PC5 for 10.05%. showing the first five principal components account for 72.18% of the variation in water quality of Açude the Macela.

Table 5- Total variance explained

Component	Eigenvaluesr	% of variance	cumulative.%
1	3.068601	25.57167	25.57167
2	1.631287	13.59406	39.16573
3	1.483142	12.35952	51.52525
4	1.273158	10.60965	62.1349
5	1.20525	10.04375	72.17865
6	0.917259	7.643825	79.82248
7	0.641212	5.343437	85.16591
8	0.523913	4.365939	89.53185
9	0.476204	3.968366	93.50022
10	0.340936	2.841131	96.34135
11	0.328466	2.73722	99.07857
12	0.110572	0.921431	100

Component loading and communalities for each variable in four selected component before varimax rotation were described in Table 6. and after Varimax rotation in Table 7. The commonalities provide an index for the efficiency of the set reduced components as well as the degree of contribution each of the variables selected for the four components. The first PC accounting for 25.27% of the total variance was correlated positively with nitrate and chlorophyll a. while nitrite. ammonia nitrogen and phosphate presented negative contributions to this variation.

Nitrogen is one of the most important elements for maintaining the life of aquatic ecosystems. It can be found in ammoniacal. nitrite and nitrate forms - nitrate is the main form of nitrogen found in the waters. as it is a major source for primary producers. When in high concentrations. the oxidation of their species can consume a lot of oxygen. stimulating the growth of algae. The ammonia nitrogen is derived from the organic matter decomposition process and. in large quantities. can cause mortality of fish. The organic nitrogenous substances undergo decomposition to nitrate. passing through ammonia and. for this reason. its presence indicates recent pollution.

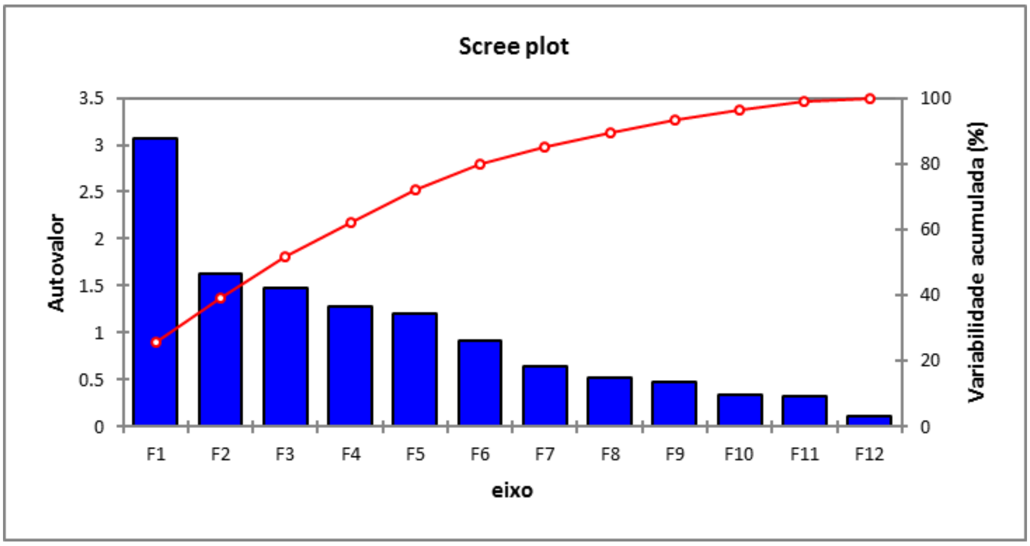

Figure 2- Scree plot of the eigenvalue for each component.

The second PC was highly correlated with electrical conductivity. suspended solids and total solids. indicating that the fluid medium has high levels of salinity and dissolved pollutants (such as NO_3. PO_4. Cl^-. SO_4^{-2}. Na^+ and K^+). Several factors may have contributed to this situation. such as climatic conditions. especially during the dry season. where the increased evapotranspiration rate may have contributed to the increased salinity of water. The third PC was correlated with pH and dissolved oxygen. representing a source of physical and chemical variability. The forth PC was negatively correlated with dissolved oxygen. indicating waste water from domestic and agricultural discharge in the Açude da Macela.

Table 6- Component matrix (a).

Parameters	**Components**			
	1	2	3	4
Conductivity (mS cm^{-1})	0.470	0.788	0.257	-0.156
Colour (Pt-Co)	0.347	0.473	-0.169	-0.216
pH	0.270	-0.818	0.625	0.212
Total solids (mg L^{-1})	0.474	0.718	0.372	0.007
Suspended Solid (mg L^{-1})	-0.448	0.860	0.256	-0.123
Dissolved Oxygen (mg L^{-1})	0.415	-0.479	0.601	-0.661
$N-NH_4$ (µg L^{-1})	-0.742	-0.311	0.016	0.065
$N-NO_2$ (µg L^{-1})	-0.733	-0.421	0.015	0.087
$N-NO_3$ (µg L^{-1})	0.833	-0.559	0.159	-0.160
$P-PO_4$ (mg L^{-1})	-0.658	-0.582	0.269	-0.160
Chlorophyll *a* (µg L^{-1})	0.754	-0.351	0.053	0.681
Water temperature (ºC)	0.327	0.424	-0.277	0.133

Table 8 represents the correlation components matrix (component Score covariance matrix) of varimax rotated four PC. which indicates that there are no correlation between components so each of the components represents a discrete unit from others.

Extraction Method: Principal Component Analysis; a—4 components extracted.

Table 7- Rotated component matrix (a)

Parameters	Components			
	1	2	3	4
Conductivity (mS cm^{-1})	0.231	0.588	0.260	-0.246
Colour (Pt-Co)	0.147	0.773	-0.203	-0.016
pH	0.070	-0.518	0.726	0.412
Total solids (mg L^{-1})	0.123	0.754	0.145	0.103
Suspended Solid (mg L^{-1})	-0.248	0.892	-0.254	-0.327
Dissolved Oxygen (mg L^{-1})	0.115	-0.179	0.684	-0.638
$N\text{-}NH_4$ (µg L^{-1})	-0.812	-0.435	0.315	0.249
$N\text{-}NO_2$ (µg L^{-1})	-0.823	-0.119	0.415	-0.107
$N\text{-}NO_3$ (µg L^{-1})	0.844	-0.255	-0.249	-0.042
$P\text{-}PO_4$ (mg L^{-1})	-0.746	-0.269	0.354	-0.260
Chlorophyll *a* (µg L^{-1})	0.836	-0.124	0.142	0.181
Water temperature (ºC)	0.127	0.745	0.157	0.233

Extraction Method: Principal Component Analysis. Rotation Method: Equamax with Kaiser Normalization; a—rotation converged in 5 iterations.

Table 8- Component score covariance matrix.

Components	1	2	3	4
1	1.000	0	0	0
2	0	1.000	0	0
3	0	0	1.000	1.000
4	0	0	0	0

3.1 Application of Water Quality Indexes

Besides being a challenging task. it is necessary to strike a balance between the determined value. the quality of water and the effectiveness of the index by analyzing some water quality parameters. Even when all those pre-selected variables are considered important as quality indicators. some assume different weights because the final destination of the water. Therefore. the Açude da Macela reservoir was evaluated according to the WQIR (IPA). O-WQI. CPCB-WQI and IQAR-m indexes. in order to minimize the subjectivity and improve the credibility of the evaluation of the quality of this water body.

Through the results obtained by analyzing the physical. chemical and biological parameters. the water quality indexes were determined for the Açude da Macela. whose results are shown in Table 9.

Table 9 – Water Quality Indexes for Açude da Macela

Indexes	Açude da Macela	Classification
WQIR (IAP)	6.00	extremely polluted
O-WQI	2.16	very bad
CPCB-WQI	20.13	very bad
WQIR-m	4.61	very polluted

Observing Table 9 and relating the results to the numerical descriptions of the contents found in Table 2 - it is possible to notice that the quality of Açude da Macela is classified as extremely polluted. very bad. very bad and very polluted. according to the adjusted indices - with respect to the parameters and calculation formula of WQIR (IAP). O-WQI. CPCB-WQI WQIR-m. respectively.

When comparing the calculation result obtained by WQIR (IAP) and WQIR-m. it is possibe to observe that there was a change on the classification. The reservoir left the extremelly polluted rate to very polluted. this may be attributed to the new weights which were obtained for the environmental variables through principal component analysis. ie. the eclipse effect. which is a result of the easing of the negative behavior of an environmental variable compared to the stable behavior of others. may have been reduced or eliminated.

Also it can be noted that all indices were calculated based on different parameters. but these results are consistent with the principal component analysis. Thus. the variables. which presented the worst results of the analyzes were chlorophyll a. the ammonia and nitrite and nitrate. and total phosphorus. which are indicative of eutrophication process. possibly due to domestic waste dumps and industrial and agricultural runoff associated with the application of fertilizers in agriculture.

In this context. with respect to environmental management. monitoring of physical. chemical and biological parameters for assessing the impact of human action on water resources is essential. The use of good management practices and nutrient load reduction through the treatment of domestic. agricultural and industrial effluents contribute to the improvement of water quality as a basis for sustainable development of the region and the state.

4. CONCLUSION

Regarding the quality of water according to the indicators. it was found that the Açude da Macela reservoir is classified as extremely polluted. very bad. very bad and very polluted according to the water quality indexes RWQI. OWQI. CPCB-WQI. WQI-m. respectively. It is important to notice that these results are in accordance with the physico-chemical and biological characteristics of this ecossystem. which had in some variables. such as chlorophyll a. ammonia. nitrite. nitrate and total phosphorus. values well above to the maximum allowable for any use of water. which may be related. possibly. to evictions of domestic. agricultural and industrial effluents.

The principal component analysis was presented as an important tool to explain the variance of the data set of interrelated variables through a smaller set of independent variables. principal components. and has been instrumental in minimizing the eclipse effect. giving an accurate answer to the assessment of water quality of the Açude da Macela reservoir.

The results of this case study show that it is necessary to adopt measures for the control and reduction of nutrients and organic loads in the water to contain the eutrophication process of this reservoir. In this context. it is essential to monitor the physical. chemical and biological parameters in order to assess the impact of human action on this water resource.

REFERENCES

Abbasi. T.; Abbasi. S. A. **Water Quality Indices.** 1ª edição. Elsevier. 384p. 2012.

APHA - American Public Health Association. **Standard Methods for the examination of water and wastewater**. American Water Works Association. Water Environment Federation. 22. ed. Washington. 2012.

Chaturvedi. M.K.. Bassin. J.K. Assessing the water quality index of water treatment plant and bore wells. in Delhi. India. **Environmental Monitoring Assessment**. 163. 449-453. 2010.

Departamento Nacional De Obras Contra As Secas (Dnocs) – Coordenadoria Estadual De Sergipe. **Ficha Técnica Operacional Dos Açudes Cest-Se**. Aracaju/Se. 2014.

Lobato. T.C.; Hauser-Davis . R.A.; Oliveira . T.F.; Silveira . A.M.; Silva. H.A.N.; Tavares . M.R.M.; Saraiva . A.C.F. Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: a case study in the amazon region**. Journal of Hydrology.** 522. p. 674–683. 2015.

Manly. B. J. F. **Métodos Estatísticos Multivariados:** Uma Introdução. 3.ed. Porto Alegre:Bookman. 2005.

Triola. M. F. **Introdução á Estatística.** 7 ed. Rio de Janeiro: Editora LTC. 1999. 410p.

Tundisi. J. G.; Tundisi-Matsamura. T.. 2002. **Lagos e Reservatórios. Qualidade de Água:** O Impacto de Eutrofização. V.3 Disponível em: <http://www.rimaeditora. com.br/lagos.pdf>. Acesso em: 06 set. 2014.