

# **ANÁLISE ESTATÍSTICA MULTIVARIADA DE DADOS DE QUALIDADE DA ÁGUA DA LAGOA DA PARANGABA, FORTALEZA, CEARÁ, BRASIL.**

**Erilany C. M. BARROS; Lyndervan O. ALCÂNTARA; Marcio R. P. SANTOS;  
Hugo L. B. BUARQUE; Raimundo B. GOMES.**

Grupo de Pesquisa em Processos Químicos e Ambientais  
Instituto Federal de Educação, Ciência e Tecnologia do Ceará  
Av. Treze de Maio, 2081, Benfica, 60.040-531, Fortaleza, Ceará  
Telefone/fax: +55 (85) 3307-3647 / +55 (85) 3307-3711,  
E-mails: erilanydmb@hotmail.com, lyndervan@hotmail.com,  
marcio\_xp@hotmail.com, hbruarque@ifce.edu.br.

## **RESUMO**

No presente estudo foram avaliados os dados da qualidade de água da Lagoa da Parangaba, Fortaleza, Ceará, com o intuito de determinar as relações e a influência dos fatores determinantes na qualidade da água, através de técnicas de estatística multivariada. Para isso, foram coletadas amostras bimestrais no período de agosto de 2006 a julho de 2009, em três pontos de distintos da lagoa. O presente estudo objetivou fazer a comparação com os parâmetros físico-químicos e os parâmetros físico-químicos incluindo os seis metais, e juntamente extrair indicadores de qualidade para diferentes pontos de coleta do corpo hídrico, utilizando métodos estatísticos multivariados de dados. A partir da análise de componentes principais foram selecionados igualmente 9 parâmetros para os dois testes. Foram extraídos três componentes principais explicando 64% e 67% O percentual de variância total explicada, respectivamente para os dois procedimentos. As componentes principais foram extraídas, com rotação, utilizando o critério da raiz latente, seguido da obtenção dos escores fatoriais, através do método de Bartlett, os quais foram utilizados como índices de qualidade deles derivados, bem como associá-los com a precipitação pluviométrica do período. Análise de agrupamento foi aplicada para analisar similaridade entre os pontos de amostragem, como também, identificar os distintos testes de agrupamento e a influência dos teores de metais nestes.

**Palavras-chave:** análise de componentes principais, análise de agrupamento, Lagoa da Parangaba.

## **1. INTRODUÇÃO**

O processo de urbanização e as atividades de uso e ocupação no entorno das lagoas urbanas tem gerado sérios problemas ambientais, como o aterramento e poluição de estruturas lacustres. Negligencia-se o fato de que as lagoas interferem no equilíbrio hídrico, micro clima, valorização da paisagem urbana, possuindo também potencial para desenvolvimento das atividades de pesca e lazer (VASCONCELOS et al., 1995).

A Lagoa da Parangaba está localizada no bairro de mesmo nome em Fortaleza, Ceará, e faz parte da bacia hidrográfica do Rio Ceará-Maraguapinho com uma extensão de 303.200 m<sup>2</sup>, uma profundidade média de 2,77 m e um volume de água de 1,2 bilhões de m<sup>3</sup>. É considerada uma das maiores lagoas urbanas de Fortaleza, tanto em profundidade como em volume de água. A Lagoa também é conhecida pela feira que se realiza todos os domingos em área urbanizada na margem da mesma - a “Feira dos Pássaros”.

A avaliação *in loco* das principais fontes de poluição e dos impactos no entorno da lagoa permite observar a presença de entradas pontuais de esgotos, acúmulo de lixo, criação de animais, falta de operação e manutenção nos sistemas de drenagem, aterramento e ocupação irregular da Área de Preservação Permanente, isso vem prejudicando as atividades na Lagoa da Parangaba. Sabe-se também que existe uma grande área de comercialização com rede de esgoto clandestino e fossa-sumidouro, que desemboca neste sistema lacustre.

A quantificação da carga poluidora que aporta a um corpo d'água é um elemento fundamental para qualquer manejo que objetive a conservação e o uso sustentável da água. Através da análise integrada entre os dados de qualidade da água e as características de uso e ocupação do solo em uma bacia hidrográfica, juntamente com a distribuição da população urbana e a disponibilidade de infraestrutura urbana e industrial, torna-se possível definir relações de causa e efeito entre as condições de ocupação da bacia e a qualidade da água (ZIMMERMANN et al., 2008).

Para uma interpretação ecológica da qualidade das águas superficiais e/ou para estabelecer um sistema de monitoramento, é necessário a utilização de métodos simples e que dêem informações objetivas e interpretáveis, partindo para critérios próprios que considerem as características peculiares dos recursos hídricos (PINEDA, SCHÄFER, 1987 apud TOLEDO, NICOLELLA, 2002). Neste aspecto, o uso de índices de qualidade de água é uma tentativa que todo programa de monitoramento de águas superficiais prevê como forma de acompanhar, através de informações resumidas, a possível deterioração dos recursos hídricos ao longo da bacia hidrográfica ou ao longo do tempo (TOLEDO, NICOLELLA, 2002).

Os métodos estatísticos multivariados podem oferecer uma interpretação simples a partir da análise de dados complexos, que ajudaria na melhor compreensão do comportamento do corpo hídrico estudado. Esses métodos também permitem a identificação dos possíveis fatores ou fontes responsáveis pelas variações na qualidade da água do manancial.

Neste contexto, o objetivo do estudo é aplicar técnicas estáticas multivariada, utilizando métodos da Análise de Componentes Principais (ACP) e a Análise de Agrupamento (AA) a partir de um conjunto de dados contendo parâmetros físico-químicos das águas superficiais da lagoa, as quais foram coletadas bimestralmente num período compreendido entre 2006 a 2009. As técnicas aplicadas permitiram determinar as relações e a influência dos fatores determinantes na qualidade da água, bem como comparar os resultados gerais obtidos somente com os parâmetros físico-químicos típicos, como também com a inclusão de teores de diferentes metais pesados.

## 2. MATERIAIS E MÉTODOS

A Lagoa da Parangaba é um das maiores lagoas urbanas de Fortaleza, com uma extensão superficial do espelho líquido superior a 30 ha. Foram selecionados três pontos de amostragem, coletadas bimestralmente em pontos situados na entrada tributária no centro da lagoa e no sangradouro para melhor entender a qualidade de água. Na Figura 1 é ilustrada a localização dos pontos de coleta selecionados. Os dados geográficos (georeferenciamento) dos pontos estão dispostos na Tabela 1.



Figura 1 – Imagem Fotogramétrica da Lagoa da Parangaba com os pontos de coleta.

**Tabela 1 – Localização geográfica dos pontos de coleta da Lagoa da Parangaba.**

Ponto de Coleta	Coordenadas Geográficas
Ponto 1	3°46'140"/38°34'090"
Ponto 2	3°33'977"/38°46'085"
Ponto 3	3°45'870"/38°33'932"

Inicialmente, foram previamente selecionados vinte e cinco parâmetros físicos – químicos incluindo seis metais, no período de agosto de 2006 a fevereiro de 2009. Estes parâmetros foram uniformemente formatados e usados na obtenção de estatísticas básicas (média, desvio padrão), para a compilação dos dados analíticos (parâmetros) foram feita com auxílio da planilha eletrônica EXCEL<sup>®</sup>. Todos os parâmetros foram obtidos seguindo metodologias do Standard Methods for the Examination of Water and Wastewater.

Foram realizados dois testes: um somente com os 25 parâmetros físico-químicos; e outro com mesmos parâmetros físico-químicos, mas incluindo os metais. Em ambos os testes, com o intuito de reduzir a quantidade original de parâmetros, foi empregada a Análise de Componentes Principais, e assim, obter uma melhor compreensão dos fatores que influenciam a qualidade de água. Nestes testes foram analisados a matriz de correlação dos dados, a matriz anti-imagem, o critério de adequação de Kaiser-Meyer-Olkin (KMO), e o percentual de variância das componentes principais, estes foram utilizados como critério de redução e seleção dos parâmetros de qualidade mais relevantes para o estudo.

Assim, o KMO e o índice de adequação do parâmetro na matriz anti-imagem deveriam ser superiores a 0,70 e 0,50, respectivamente, valores considerados adequados para os propósitos da análise fatorial (HAIR JR et al., 2005). Caso contrário, o parâmetro com menor índice de adequação na matriz anti-imagem era excluído do conjunto de dados e o novo conjunto era reavaliado. A extração das componentes principais do conjunto de dados deve explicar maior parte da variância dos dados originais. Para as considerações desse estudo, a extração ACP deve ser superior a 64% da variância dos dados para conjunto avaliado.

Posteriormente, aplicou os parâmetros selecionados na análise de agrupamento, baseado no método de *Ward* usando quadrado distância euclidiana e a norma do z-scores, possibilitando extrair informações de similaridade entre os três pontos amostragem da lagoa, a partir da média dos parâmetros selecionados, em cada ponto, no período.

A análise de agrupamento (AA) agrega os elementos no intuito de formar grupos que sejam semelhantes entre si no que diz respeito às características que neles foram medidas. Os agrupamentos resultantes dos elementos devem então exibir elevada homogeneidade interna (dentro dos agrupamentos) e elevada heterogeneidade externa (entre agrupamentos). Assim, se a classificação for bem sucedida, os elementos dentro dos grupos estarão bem próximos quando representados graficamente e diferentes agrupamentos estarão distantes. (HAIR JR et al., 2005)

### 3. RESULTADOS E DISCUSSÕES

Utilizando a metodologia proposta, dos vinte e cinco parâmetros físico-químicos originais somente 9 (nove) parâmetros foram selecionados através da ACP com um KMO de 0,718, sugerindo que estes são os que exercem maior influência na variação do comportamento do corpo hídrico, haja vista que estes forneceram correlações não hipotéticas, com nível de significância de 5% e índices de adequação para os parâmetros superiores a 0,50. Com o teste incluindo os seis metais considerados, foram também selecionados 9 (nove) parâmetros, para um KMO de 0,725, com o mesmo nível de significância.

A partir da série de dados selecionados foram determinadas estatísticas como: o valor mínimo, valor máximo, média aritmética e desvio padrão, para os nove parâmetros incluindo todos os três pontos de coleta, conforme apresentado na Tabela 2. Os coeficientes de variação altos indicam a inconstância dos valores dos parâmetros, e sua maior variabilidade no sistema lacustre.

**Tabela 2 – Estatística descritiva básica das variáveis estudada dos três pontos de amostragem.**

Parâmetros	Mínimo	Máximo	Média	Desvio Padrão	Cof. de Variação.
ALC (mg CaCO <sub>3</sub> /L)	105,0	276,3	156,3	57,9	37%
S (mg/L)	0,00	3,10	0,90	0,77	86%
NITRATO (mg/L)	0,008	9,758	0,930	2,100	226%
Mn (mg/L)	0,007	0,070	0,030	0,013	44%
SSV (mg/L)	2	480	42	101	240%
OG (mg/L)	0,80	61,00	11,59	11,02	95%
T (°C)	28,0	33,0	29,5	1,2	4%
CLOa (µg/L)	10,00	133,60	41,25	30,00	73%
Ni (mg/L)	0,001	0,022	0,007	0,006	89%
SDT (mg/L)	323	916	502	106	21%
DT(mg CaCO <sub>3</sub> /L)	129,0	211,6	162,8	22,1	14%
pH	7,8	8,9	8,4	0,3	3%
OD(mg/L)	4,3	12,0	7,9	1,7	21%
SO <sub>4</sub> (mg/L)	0,7	67,5	39,1	13,8	35%
TURB(µT)	10	25	16	3	22%

O primeiro teste reduziu o sistema em até oito parâmetros, considerando o conjunto com nove parâmetros o mais significativo segundo todos os critérios estabelecidos. No teste incluindo metais, o conjunto mais significativo possuía oito parâmetros, sendo este o mais significativo conforme os critérios usados. Contudo, para este teste, optamos pelo sistema com nove parâmetros de modo a melhor comparar os resultados de ambos os testes.

As matrizes anti-imagem para os dois procedimentos realizados, mostradas na Tabela 3 e na Tabela 4, apresentaram altos valores na diagonal principal, demonstrando a elevada comunalidade entre os parâmetros considerados.

**Tabela 3 – Matriz anti-imagem obtida para o sistema considerado.**

	T	TURB	SDT	DT	pH	OD	SO <sub>4</sub>	NITRATO	OG
T	<b>0,574</b>								
TURB	0,177	<b>0,706</b>							
SDT	-0,315	-0,059	<b>0,770</b>						
DT	-0,104	-0,186	-0,116	<b>0,828</b>					
pH	-0,239	-0,058	0,016	-0,194	<b>0,712</b>				
OD	0,149	-0,153	-0,256	-0,194	-0,260	<b>0,756</b>			
SO <sub>4</sub>	-0,011	-0,142	-0,187	-0,104	-0,202	-0,048	<b>0,813</b>		
NITRATO	0,239	0,425	0,119	0,252	-0,254	-0,151	-0,094	<b>0,593</b>	
OG	-0,243	-0,097	0,206	-0,063	0,055	0,047	0,202	0,064	<b>0,588</b>

**Tabela 4 – Matriz anti-imagem obtida para o sistema considerado.**

	T	SSV	ALC	S	OG	NITRATO	CLOa	Mn	Ni
T	<b>0,770</b>								
SSV	-0,172	<b>0,526</b>							
ALC	0,195	0,148	<b>0,794</b>						
S	0,291	-0,205	-0,494	<b>0,698</b>					
OG	0,003	-0,572	0,010	-0,006	<b>0,662</b>				
NITRATO	0,198	-0,121	-0,206	0,128	0,095	<b>0,774</b>			
CLOa	0,130	-0,128	0,029	0,302	-0,190	-0,038	<b>0,778</b>		
Mn	0,063	0,189	0,171	0,135	-0,121	0,235	-0,167	<b>0,794</b>	
Ni	-0,112	0,225	-0,134	0,078	-0,063	-0,216	-0,079	0,287	<b>0,705</b>

Seguindo o procedimento, observou que três parâmetros permanecem nos dois testes, sendo este temperatura (T), nitrato (NITRATO) e óleos e graxas (OG). O nitrato está associado com poluição recente, indicando a possível existência de despejos orgânicos recentes no meio hídrico; o teor de óleos e graxas é oriundo de resíduos industriais, esgotos domésticos, efluentes de oficinas e posto de gasolina, e tem como principal problema associado a diminuição da área de contato entre a superfície da água e do ar atmosférico, impedindo, a transferência de oxigênio da atmosfera para a água; a temperatura varia de acordo com regime climático, mas sua elevação em um corpo d'água pode ser provocada por despejos industriais. Ela também desempenha um papel principal de controle no meio aquático, condicionando as influências de uma série de variáveis físico-químicas.

Cada teste extraiu três componentes principais (CP1, CP2 e CP3) a partir da matriz de correlação do sistema analisado. O percentual de variância total explicada obtida para as três componentes foi igual a 64% e 67%, respectivamente. Na Tabela 5 e Tabela 6 observamos ACP rotacionadas pelo Método Varimax, como também com seus respectivos percentuais de variância explicada.

**Tabela 5– Análise Componente Principal rotacionada pelo Método Varimax.**

Componentes	Variância Explicada	pH	OD	SO4	SDT	TURB	NITRATO	DT	T	OG
1	28,27%	<b>0,765</b>	<b>0,709</b>	<b>0,708</b>	<b>0,628</b>	0,227	0,120	0,494	0,252	-0,434
2	20,97%	-0,128	0,175	0,124	0,327	<b>0,829</b>	<b>-0,820</b>	<b>0,569</b>	0,108	0,151
3	14,85%	0,204	-0,149	-0,113	0,217	-0,088	-0,281	0,237	<b>0,858</b>	<b>0,579</b>

**Tabela 6– Análise Componente Principal rotacionada pelo Método Varimax incluindo os metais.**

Componentes	Variância Explicada	S	ALC	T	CLOa	Ni	NITRATO	Mn	SSV	OG
1	26,77%	<b>-0,888</b>	<b>-0,744</b>	<b>0,667</b>	<b>0,579</b>	0,073	-0,247	0,433	0,027	0,181
2	21,06%	0,127	0,417	-0,120	0,004	<b>0,824</b>	<b>0,743</b>	<b>-0,662</b>	-0,071	-0,124
3	19,55%	0,051	-0,135	0,106	0,392	-0,210	-0,008	0,001	<b>0,898</b>	<b>0,850</b>

A partir da extração das componentes principais foi utilizado o critério de Bartlett para estimar coeficientes fatoriais que minimiza a variância de erro, sendo estes adotados como peso na obtenção dos escores fatoriais. Estes fatores, calculados para todas as componentes principais, podem ser usados como indicadores de qualidade de água (IQA), tanto para o teste com metais e sem metais, no período de estudo. Deste modo, os indicadores determinados foram plotados num gráfico juntamente com dados de precipitação pluviométrica compilados de FUNCEME (2009). A seguir, está expressa uma das equações obtidas, Eq. 1, para os IQA, a partir da primeira componente principal.

$$IQA_1 = 0,351Z_1 + 0,278Z_2 + 0,285Z_3 + 0,223Z_4 - 0,029Z_5 + 0,173Z_6 + 0,129Z_7 + 0,101Z_8 - 0,204Z_9 \quad [\text{Eq. 1}]$$

em que  $Z_i$  são variáveis normalizadas e os sub-índices representam as variáveis na mesma ordem dos parâmetros da Tabela 4. Os índices calculados por este procedimento possuem uma média igual a zero e uma variância unitária, com valores compreendidos entre -3 e 3. Pode ser notado que valores muito negativos para o  $IQA_1$  indicam uma elevada presença de óleos e graxas, oriunda provavelmente das oficinas mecânicas presentes no entorno e da lavagem de veículos automotivos.

Na Figura 2 é apresentado o gráfico do  $IQA_1$  (sem metais) e na Figura 3 o gráfico do  $IQA_2$  (considerando os metais), bem como a precipitação pluviométrica no período de estudo. É possível verificar os distintos índices de qualidade tendem a se tornar negativos nos períodos chuvosos, uma provável consequência do arraste de óleos lubrificantes e aditivos automotivos, bem como da lixiviação de baterias nesse período, contaminando a lagoa e influenciando as suas características.

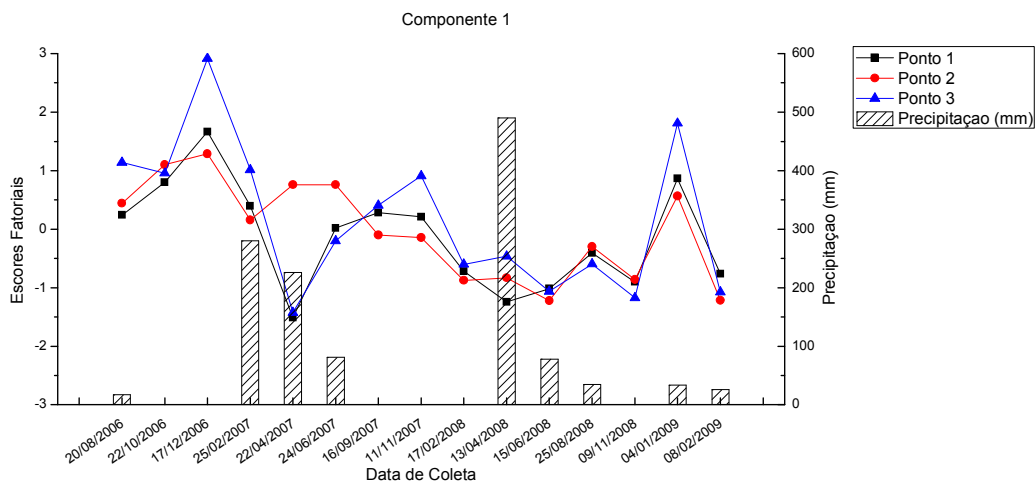


Figura 2 – IQA<sub>1</sub> e dados pluviométricos (CEARÁ, 2009) da Lagoa da Parangaba no período considerado.

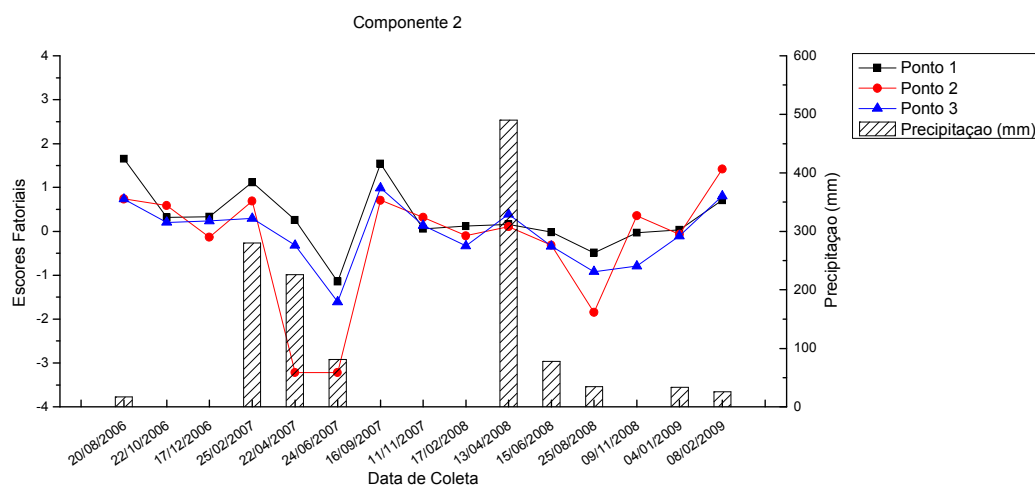


Figura 3 – IQA<sub>2</sub> e dados pluviométricos (CEARÁ, 2009) da Lagoa da Parangaba no período considerado.

Ainda, a partir da análise de agrupamento realizada nos pontos de coleta de amostras, verificaram-se agrupamentos distintos para dois testes (sem metais e com metais). O dendograma apresentado na Figura 4 mostra duas análises de agrupamento, com os respectivos agrupamento 1 (sem metais) e agrupamento 2 (com metais).

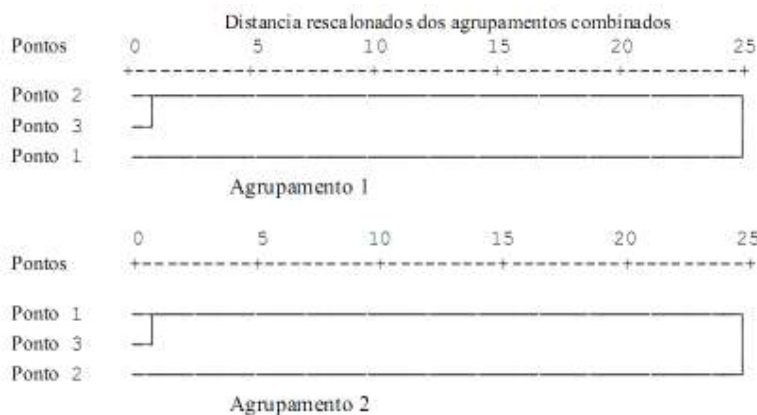


Figura 3 – Dendograma obtido pela AA, utilizando método de Ward e o quadrado da distância euclidiana.

#### 4. CONCLUSÃO

Neste estudo, foram utilizadas técnicas estatísticas multivariadas para determinar o comportamento da qualidade de água da Lagoa da Parangaba. Com o uso da análise de componente principal foi possível reduzir o número de parâmetros originais, selecionando parâmetros determinantes para as variações das características do corpo hídrico estudado. De acordo com os critérios estabelecidos para a seleção e redução das variáveis, também foi possível identificar três componentes principais relacionadas às características relevantes na variação do comportamento do corpo aquático e índices de qualidade deles derivados, bem como associá-los com a precipitação pluviométrica do período.

A partir da AA também foi possível identificar os distintos testes de agrupamento e a influência dos teores de metais nestes. Como também se identifica os distintos índices de qualidade com adição dos metais no estudo. Assim, os métodos de estatísticos multivariados podem ser usados para determinar os fatores que influenciam na qualidade da água e para entender a complexidade da qualidade de um sistema lacustre.

#### 5. REFERÊNCIAS

CEARÁ. Fundação Cearense de Meteorologia e Recursos Hídricos. **Monitoramento: Gráfico de chuvas dos postos pluviométricos**. Disponível em: <http://www.fuceme.br>. Acesso em: jul. 2010.

HAIR JR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. Análise multivariada de dados. 5ª ed. Reimpr. 2007. Porto Alegre: Bookman, 2005.

PINEDA, M.D.; SCHÄFER, A. Adequação de critérios e métodos de avaliação da qualidade de águas superficiais baseada no estudo ecológico do rio Gravataí, Rio Grande do Sul, Brasil. *Ciência e Cultura*, v.39, p.198-206, 1987.

SPSS – Statistical Package For The Social Sciences. Base 13.0 User's Guide. Chicargo: SPSS, 2004.

**Standard methods for the examination of water and wastewater**. 21.ed. Washington, DC, EUA: American Public Health Association (APHA), the American Water Works Association (AWWA), and the Water Environment Federation (WEF), 2005.

TOLEDO, L. G.; NICOLELLA G. Índice de qualidade de água em microbacia sob uso agrícola e urbano. São Paulo, SP, Brasil. *Scientia Agrícola*, v.59, n.1, p.181-186, 2002.

Vasconcelos, F. P.; Coriolano, L. T. e Souza, M.J. (1995). *Análise Ambiental e Sócio-Econômica dos Sistemas Lacustres Litorâneos do Município de Fortaleza*. Relatório de Projeto. Universidade Estadual do Ceará. Fortaleza.

ZIMMERMANN; C. M.; GUIMARÃES; O. M.; ZAMORA; P. G. P. Avaliação da qualidade do corpo hídrico do rio Tibagi na região de Ponta Grossa utilizando análise de componentes principais (PCA). Ponta Grossa, PR, Brasil. *Química Nova*, v. 31, n. 7, p. 1727-1732; 2008.