

I-190 - APLICAÇÃO DA ANÁLISE DE COMPONENTES PRINCIPAIS PARA UMA ANÁLISE EXPLORATÓRIA DOS DADOS DE QUALIDADE DA ÁGUA DA REPRESA DE NOVA PONTE/MG

Déborah Tavares Viana⁽¹⁾

Bióloga pela Pontifícia Universidade Católica de Minas Gerais. Mestranda do Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG.

Jéssyca Irene Guimarães

Graduanda em Engenharia Ambiental pela Universidade Federal de Minas Gerais.

Ana Laura Cerqueira Trindade

Bióloga pela Universidade Federal de Minas Gerais. Mestre pelo Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG.

Mônica Maria Diniz Leão

Engenheira Química (UFMG, 1980) e de Segurança do Trabalho (UFMG, 1999); Doutora em Engenharia Ambiental (INSA, Toulouse, FR, 1984); Professora associada do Depto. Eng. Sanitária e Ambiental (DESA) da Escola de Engenharia da UFMG.

Silvia Maria Alves Corrêa Oliveira

Engenheira Eletricista, Mestre e Doutora em Saneamento, Meio Ambiente e Recursos Hídricos pela Escola de Engenharia da UFMG. Professora Adjunta do Depto. Eng. Sanitária e Ambiental (DESA) e coordenadora do Curso de Graduação em Engenharia Ambiental da UFMG.

Endereço⁽¹⁾: Avenida Antônio Carlos, 6627- Escola de Engenharia - UFMG - Bloco 1, Sala 4606 - CEP 31270-901 - Brasil - Tel: (31)34091947 - e-mail: deborahviana@gmail.com

RESUMO

Um dos maiores problemas para eficácia dos monitoramentos é a falta de utilização dos dados, devido a essas dificuldades a análise multivariada começa a ser difundida na área de qualidade da água nos anos noventa. Para realização deste trabalho os dados do monitoramento da qualidade da água de 1995 a 2011 foram obtidos através da Companhia Energética de Minas Gerais. Foram selecionados 5 pontos lânticos distribuídos no reservatório. Foi realizada a análise estatística descritiva, incluindo o cálculo de mediana, percentil 10% e percentil 90%, assimetria e curtose. A Análise de Componentes Principais foi realizada para identificar os agrupamentos existentes entre os parâmetros. Baseado no *Scree Plot* e nos autovalores (foram considerados apenas os maiores que 1) foram encontrados 6 componentes principais que explicam 66% da variância total dos dados. A análise de componentes principais encontrou 6 componentes principais e nos mostrou que 3 desses grupos foram os mais significativos para a análise de qualidade da água. Mostrou também que há parâmetros não estão influenciando muito as variações da qualidade da água devido a seus valores censurados ou pouco variáveis. Também foi possível identificar que os grupos formados pelas componentes principais estão muito ligados aos grupos de sólidos, íons e nutrientes, mostrando que são as principais fontes de variação na qualidade da água. Na análise de tendência Mann-Kendall o reservatório mostrou estabilidade na maior parte dos parâmetros (sem tendência). Os parâmetros que apresentaram uma tendência em mais do que em uma estação sempre seguiram o mesmo padrão (aumento ou diminuição), indicando que o reservatório apresenta tendências espaciais semelhantes. A análise de tendência indica o possível aumento da cor nas águas do reservatório e a diminuição da condutividade elétrica e dos SDT. Embora seja uma análise preliminar, estes resultados mostram que uma eventual otimização da rede de monitoramento deverá considerar essas fontes.

PALAVRAS-CHAVE: Monitoramento, Análises Estatísticas, Reservatório, Qualidade da Água.

INTRODUÇÃO

O monitoramento da qualidade da água é uma operação muito importante para assegurar a preservação dos cursos d'água. O ambiente aquático é essencial para assegurar o refúgio de animais, a biodiversidade, ecossistemas ripários e ainda tem grande valor cênico. Em geral, um programa de monitoramento inclui coletas frequentes nos mesmos pontos de amostragem e análise em laboratório de grande número de parâmetros, resultando em uma matriz de grandes dimensões e complexa interpretação (ANDRADE et al, 2007). Um dos maiores problemas para eficácia dos monitoramentos é a falta de utilização dos dados, pois

sua armazenagem e análise eficiente acabam não sendo realizados (CARLETON et al, 2005). Muitas dúvidas são levantadas sobre a melhor maneira de armazenar e analisar os dados de tantos pontos e parâmetros. Devido a essas dificuldades a análise multivariada começa a ser difundida na área de qualidade da água nos anos noventa e é muito explorada na década seguinte (VEGA et al, 1998; SIMEONOV et al, 2003; SINGH ET al, 2004; ASTEL et al, 2007). As análises multivariadas se mostram eficaz para caracterização e avaliação da qualidade de rios e lagos, pois evidenciam variações temporais e espaciais causadas por processos naturais ou poluentes (Jayakumar E Siraz, 1997; VeGa et al., 1998; Helena et al., 2000; Singh et al., 2004). Para estudos de series temporais e espaciais as análises do tipo Análise de Agrupamentos (AA), também conhecida como Análise de Cluster (AC), Análise dos Componentes Principais (ACP), e Análise de Discriminantes (AD) são muito utilizadas (LI et al, 2007; ZHOU et al 2007; ZHANG et al, 2011). Cada uma das análises apresenta uma série de vantagens e desvantagens, as quais devem ser estudadas para cada tipo de estudo.

A análise de agrupamentos constitui uma metodologia numérica multivariada, com o objetivo de propor uma estrutura classificatória, ou de reconhecimento da existência de grupos, objetivando, mais especificamente, dividir o conjunto de observações em um número de grupos homogêneos, segundo algum critério de homogeneidade. Simeonova e Einax (2002) salientam a utilidade da AA como uma forma de detectar similaridades e dissimilaridades em um grupo grande de objetos. Assim, como resultados são obtidos as similaridades entre objetos internos e alta heterogeneidade entre os grupos de objetos, podendo ser visualmente analisados pelos dendogramas (ZHANG, 2010).

O método de ACP é usado para identificar quais variáveis explicam a maior variância, portanto contém informações sobre um pequeno grupo de dados que englobam grande número de variáveis (LINDMAN, 1974). A ACP reduz o número de variáveis agrupando diversas em um só grupo, as componentes principais (CP). Ou seja, o método ACP é uma ferramenta de reconhecimento que explica a variância de um grande número de dados de um conjunto de variáveis correlacionadas com um pequeno conjunto de variáveis independentes (SIMEONOV et al, 2003). Essa transformação, em outro conjunto de variáveis, ocorre com a menor perda de informação possível, sendo que esta também busca eliminar algumas variáveis que possuem pouca informação. A metodologia do método da ACP consiste em extrair autovalores e autovetores de uma matriz de covariância de variáveis originais para produzir novas variáveis ortogonais que são combinações lineares das variáveis originais.

A Análise de Discriminantes, assim como a ACP, é muito utilizada para encontrar as variáveis mais significativas para assim aperfeiçoar o monitoramento. Essa forma de análise é encontrada muitas vezes associada a AC para verificar as variáveis mais significativas associadas com os agrupamentos encontrados no dendograma (ZHOU et al, 2007).

A aplicação de diferentes técnicas de estatística multivariadas, como a análise de agrupamento (ou cluster), a análise discriminante e a análise dos componentes principais ajudam na interpretação de dados matriciais complicados levando a um melhor entendimento temporal e espacial de dados de qualidade de água. Além de identificar possíveis fatores que influenciam o sistema em análise e ainda, oferecer ferramentas para aprimorar o monitoramento dos recursos hídricos (ZHANG, 2010).

A identificação da tendência temporal de um determinado conjunto de dados ambientais é importante tanto para a comunidade científica quanto para os órgãos de controle ambiental, uma vez que essa informação pode embasar atuações preventivas e corretivas (Cristofaro & Leao, 2009). Esse tipo de análise associado a ocupação do solo é uma parte vital para entender as variações espaços-temporais da qualidade da água (CHANG 2008). Para a análise de tendência existem diversos testes estatísticos, no caso desse trabalho o mais indicado é o teste Mann-Kendall de tendência, pois ele é um teste não paramétrico robusto o suficiente para trabalhar com outliers, dados censurados e faltantes. O teste MK determina se há uma tendência de acordo com o tempo, indicando se os valores de Y aumentam ou diminuem de acordo com X que nesse caso é o tempo.

MATERIAIS E MÉTODOS

A Usina Hidrelétrica Nova Ponte, localizada no município de Nova Ponte, é um barramento do rio Araguari, um dos afluentes do rio Paranaíba. A UHE Nova Ponte iniciou sua operação no ano de 1994, com capacidade de geração de 510 MW e possuindo um reservatório com uma área de 443 km² de espelho d'água (ARAXÁ AMBIENTAL, 2010), sendo a maior usina do rio Araguari.

Para realização deste trabalho os dados do monitoramento da qualidade da água de 1995 a 2011 foram obtidos através da Companhia Energética de Minas Gerais. O Sistema de Informação da Qualidade da Água dos Reservatórios da Cemig – SISÁGUA- foi acessado no endereço < www2.cemig.com.br/sag>. Foram selecionados 5 pontos lânticos (coletados na metade da zona fótica) distribuídos no reservatório (Fig.1 e Tab.2).

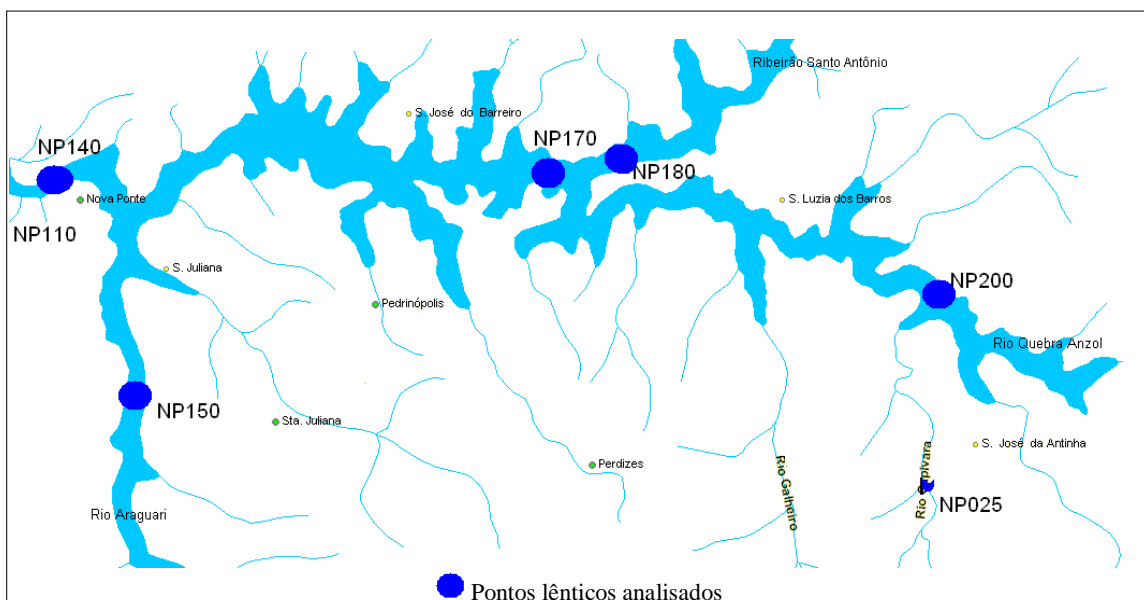


Figura 1: Localização dos pontos de coleta no reservatório de Nova Ponte.

Para a seleção de parâmetros, foram calculadas as porcentagens de dados faltantes para cada parâmetro e assim selecionados os parâmetros mais completos, sendo assim viáveis para a análise. Foram então selecionados 15 parâmetros: temperatura, pH, Condutividade Elétrica (CE), cor, turbidez (NTU), sólidos dissolvidos totais (SDT), sólidos suspensos totais (SST), Sólidos totais (ST), demanda química de oxigênio (DQO), demanda bioquímica de oxigênio (DBO), nitrogênio amoniacal (N-NH₃), oxigênio dissolvido (OD), coliformes termotolerantes, ferro total (Fe), fósforo total (P). O banco de dados final reuniu 154 coletas. Foi realizada a análise estatística descritiva, incluindo o cálculo de mediana, percentil 10% e percentil 90%, assimetria e curtose (Tab.1).

Tabela 1: Estatística descritiva para os parâmetros utilizados do banco Siságua.

| Parâmetro | N | Mediana | Percentil 10% | Percentil 90% | Assimetria | Curtose |
|-------------------|-----|---------|---------------|---------------|------------|---------|
| Temperatura | 151 | 25 | 22 | 27 | -4 | 27 |
| NTU | 152 | 2 | 1 | 6 | 8 | 67 |
| CE | 154 | 23 | 18 | 31 | 5 | 36 |
| pH | 153 | 7.0 | 6.4 | 7.8 | 0.3 | -0.4 |
| OD | 154 | 7.4 | 5.6 | 8.3 | -1.0 | 1.6 |
| Cor | 122 | 3.0 | 1.0 | 12.7 | 4.2 | 24.7 |
| Coliformes | 123 | 5 | 1 | 130 | 6 | 47 |
| DBO | 138 | 2.0 | 2.0 | 2.0 | 4.0 | 16.0 |
| DQO | 128 | 6.0 | 5.0 | 19.0 | 7.8 | 67.4 |
| Fe | 135 | 0.20 | 0.10 | 0.64 | 5.30 | 37.09 |
| P | 131 | 0.01 | 0.01 | 0.06 | 5.54 | 30.21 |
| N-NH ₃ | 135 | 0.10 | 0.10 | 0.41 | 4.25 | 16.85 |
| STD | 132 | 29 | 10 | 68 | 6 | 50 |
| SST | 125 | 3 | 3 | 10 | 5 | 32 |
| ST | 120 | 32 | 14 | 81 | 3 | 9 |

Tabela 2: Descrição dos pontos lânticos do reservatório de Nova Ponte

| Estação | Descrição | Curso d'água | Latitude | Longitude |
|---------|---|------------------------|-------------|-------------|
| NP140 | Ponto no corpo principal do reservatório, entre o barramento e o braço do rio Quebra Anzol. | Rio Araguari | 19° 07' 41" | 47° 40' 59" |
| NP170 | Reservatório, na ponte velha do rio Quebra Anzol, a jusante do braço do ribeirão Santo Antônio. | Rio Quebra Anzol | 19° 07' 26" | 47° 20' 00" |
| NP180 | Reservatório, no braço do ribeirão Santo Antônio. | Ribeirão Santo Antônio | 19° 06' 54" | 47° 16' 43" |
| NP200 | Reservatório, no braço do Rio Quebra Anzol à jusante do braço do rio Capivara | Rio Quebra Anzol | 19° 12' 50" | 47° 02' 11" |
| NP150 | Reservatório, no braço do rio Araguari. | Rio Araguari | 19° 17' 28" | 47° 38' 08" |

Para a ACP os dados foram estandardizados de acordo com a escala Z (média 0 e desvio padrão 1) para eliminar as interferências das diferentes escalas dos parâmetros analisados e não resultar agrupamentos errôneos ou enviesados. As análises foram conduzidas no software STATISTICA - data analysis software system - StatSoft Inc. (2007) versão 8.0 e XLSTAT - Addinsoft SARL (2013).

RESULTADOS E DISCUSSÃO

A Análise de Componentes Principais foi realizada para identificar os agrupamentos existentes entre os parâmetros. Baseado no *Scree Plot* e nos autovalores (foram considerados apenas os maiores que 1) foram encontrados 6 componentes principais que explicam 66% da variância total dos dados. As cargas de cada variável foram analisadas de acordo com Liu et al, 2003 que propõe: forte >0.75; moderada 0.75-0.50; fraca 0.50-0.30. A primeira componente explica 18% da variância e teve cargas fortes para turbidez e ST, e cargas moderadas para cor, SDT e SST. Portanto essa componente representa principalmente as variáveis ligadas a presença de sólidos em suspensão na água, possivelmente por escoamento superficial.

A segunda componente, explicando 13% da variância, teve carga moderada para CE, DQO, SDT e representa principalmente os íons dissolvidos na água. A terceira componente, com 11% da variância, teve carga moderada para N-NH₃, P e OD, representando provavelmente as cargas de poluição vindas por arraste dos fertilizantes no solo uma vez que se trata de uma região agrícola. A inclusão do OD pode estar ligada a especiação desses compostos em ambiente redutor ou oxidante. A quarta componente, que explica 9% da variância, obteve carga moderada apenas para a variável pH o que pode indicar que o pH não está influenciando o meio, já que é pouco variável no reservatório.

A quinta e a sexta componente, com 7,5% e 6,5% da variância respectivamente, obtiveram cargas moderadas para Fe e DBO. Mostrando que esses parâmetros não estão ligados aos demais, provavelmente pelos seus valores censurados. Portanto pode-se concluir que as componentes mais importantes para essa PCA foram as componentes 1, 2 e 3 que explicam 42% da variância total. Temperatura e coliformes não apresentaram cargas altas nem moderadas para nenhuma das componentes.

A ACP permitiu classificar os parâmetros em três grupos principais, indicando a possível fonte de alteração da qualidade da água: escoamento superficial com arraste de partículas suspensas e também de nutrientes. Embora seja uma análise preliminar, estes resultados mostram que uma eventual otimização da rede de monitoramento deverá considerar essas fontes. É possível ver também no gráfico da componente 1 versus a componente 3 (fig.2) que existem dois principais grupos formados na ACP. Um caracterizado pelo arraste de partículas suspensas e íons, um pelo escoamento de nutrientes. Foi escolhida a representação das componentes 1 e 3 por serem as mais representativas e mais conclusivas.

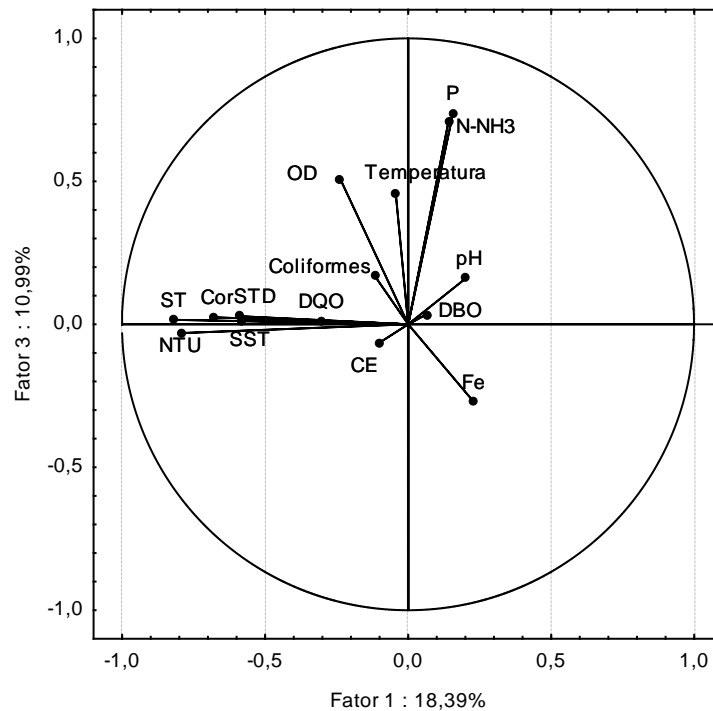


Figura 2: Projeção das cargas nos planos das componentes principais 1x3.

Na análise de tendência Mann-Kendall o reservatório mostrou estabilidade na maior parte dos parâmetros (sem tendência). Os parâmetros que apresentaram uma tendência em mais do que em uma estação sempre seguiram o mesmo padrão (aumento ou diminuição), indicando que o reservatório apresenta tendências espaciais semelhantes. Parâmetros que tendem a aumentar ou diminuir em uma estação a montante também tendem de forma semelhante em estações a jusante, o que fica mais claro na tabela X.

| Estação | Parâmetro | Tendência |
|---------|-----------|-----------|
| NP140 | CE | - |
| | OD | - |
| | Cor | + |
| NP150 | CE | - |
| | SDT | - |
| NP170 | CE | - |
| | Cor | + |
| NP180 | SDT | - |
| | ST | - |
| NP180 | Cor | + |

Figura 3: Resultados da análise de tendência MK para o reservatório de Nova Ponte.

A análise de tendência indica o possível aumento da cor nas águas do reservatório e a diminuição da condutividade elétrica e dos SDT. É possível observar nessa tabela que os parâmetros que apresentam tendência são ligados principalmente aos íons dissolvidos (CE, SDT) e aos sólidos suspensos (cor). Esse resultado corrobora os resultados obtidos na ACP, mostrando que os parâmetros mais representativos do reservatório e também os parâmetros que estão sofrendo maior variação (mudança temporal para aumento ou diminuição) são os parâmetros diretamente ligados ao escoamento superficial.

CONCLUSÕES

A análise de componentes principais encontrou 6 componentes principais e nos mostrou que 3 desses grupos foram os mais significativos para a análise de qualidade da água. Mostrou também que há parâmetros não estão influenciando muito as variações da qualidade da água devido a seus valores censurados ou pouco variáveis. Também foi possível identificar que os grupos formados pelas componentes principais estão muito ligados aos grupos de sólidos, íons e nutrientes, mostrando que são as principais fontes de variação na qualidade da água.

Os testes de ACP associados com testes de tendências temporais mostraram-se eficientes para identificar os fatores que devem ser foco nas ações ambientais. Esse resultado é interessante para auxiliar nas tomadas de decisões relativas ao monitoramento periódico da qualidade da água, mostrando que as concentrações de sólidos e nutrientes são fatores a serem observados e devem ser monitorados frequentemente.

É importante também salientar que os resultados dos testes estatísticos têm limitações e devem ser analisados de forma crítica. Os agrupamentos e as tendências encontradas devem ser confrontados com estudos aprofundados e análise dos dados de ocupação do solo para verificar a confiabilidade do teste. Os testes apresentam limitações e o banco de dados também, uma vez que a série histórica apresenta falhas e outliers. A identificação do teste mais adequado deve ser analisada para minimizar o Máximo possível essas falhas, mas elas ainda assim não deixam de existir, limitando as conclusões.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro concedido pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDRADE, E. *et al.* Seleção dos Indicadores da Qualidade das Águas Superficiais pelo Emprego da Análise Multivariada. **Revista de Engenharia Agrícola**, v.27, n.3, p.683-690, 2007.
2. ARAXÁ AMBIENTAL LTDA. **Relatório do Monitoramento da Qualidade das Águas – UHE Nova Ponte**. Araxá, 2010. 83p. (Relatório de Monitoramento)
3. ASTEL, A. *et al.* Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. **Water Research**, Bulgária, n. 41, p.4566-4578, 2007.
4. CARLETON, Christian *et al.* A relational database for the monitoring and analysis of watershed hydrologic functions: I. Database design and pertinent queries. **Computers & Geosciences**, Califórnia, v. 31, n. , p.393-402, 2005.
5. HELENA, B. *et al.* Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis. **Water Research**, v.34, p.807–816, 2000.
6. HELSEL, D. R.; FRANS, L. M. **Regional Kendall test for trend**. 2006. Disponível em:<<http://pubs.acs.org/doi/abs/10.1021/es051650b?prevSearch=helsel&searchHistoryKey=>>>. Acesso: 09 junho 2009.
7. JAYAKUMAR, R. e SIRAZ, L. Factor analysis in hydro geochemistry of coastal aquifers. **Environmental Geochemistry**, v.31, p.174–177, 1997.
8. LIU, Chen-Wuing. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. **The Science Of The Total Environment**, Taiwan, n. 313, p.77-89, 2003.
9. LINDMAN, Harold. **Analysis of variance in complex experimental designs**. EUA: W. H. Freeman, 1974. 352 p.
10. SIMEONOV, V. *et al.* Assessment of the surface water quality in Northern Greece. **Water Research**, Grécia, n. 37, p.4119-4124, 2003.
11. SINGH, K. *et al.* Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. **Water Research**, v.38, p.3980–3992, 2004.
12. VEGA, M. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. **Water Research**, v.32, p.3581–3592, 1998.
13. ZHANG, Xuan *et al.* Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. **Environ Monit Assess**, China, n. 173, p.17-27, 2011.
14. ZHOU, Feng *et al.* Application of Multivariate Statistical Methods to Water Quality Assessment of the Watercourses in Northwestern New Territories, Hong Kong. **Environ Monit Assess**, China, n. 132, p.1-7, 2007.