

Utilização de modelo baseado em análise por componentes principais para identificação de condições críticas de qualidade de água superficial

O monitoramento da qualidade da água é cada vez mais essencial para a preservação dos recursos hídricos. Os dados coletados das estações de monitoramento são uma fonte rica de informações sobre a qualidade da água, importantes para a proteção do recurso hídrico, a manutenção da capacidade ambiental e o controle da poluição. Incrementar as análises dos dados coletados é uma das formas de melhorar o processo de tomada de decisão sobre esses aspectos. Neste estudo, Análise por Componentes Principais (ACP), associada ao Filtro de Hampel, foi aplicada para obtenção de um modelo utilizado no monitoramento da qualidade da água. O estudo de caso refere-se a uma estação de monitoramento (BV137), localizada na bacia do rio das Velhas (Brasil). O conjunto de dados utilizado, disponibilizado pelo órgão gestor das águas em Minas Gerais - o Instituto Mineiro de Gestão das Águas (IGAM), foi coletado durante 21 anos (1997-2018) e examinado doze parâmetros físicos, químicos e microbiológicos. O modelo ACP, utilizado como sistema de monitoramento, foi capaz de discriminar eventos críticos de poluição e não críticos. Dada uma região de referência para a qualidade da água, os grupos de eventos críticos foram reconhecidos usando os gráficos de scores da ACP (score plots) e associados às fontes de poluição características da área de estudo (descargas de efluentes domésticos e industriais, atividades agrícolas e pecuárias e mineração). O mapa de uso e ocupação do solo da bacia de contribuição da estação de monitoramento também foi utilizado para as análises. Os parâmetros de qualidade da água com maior relevância foram coliformes termotolerantes, sólidos totais, turbidez e arsênio total. Essa aplicação demonstrou o potencial da técnica ACP para a identificação de padrões de comportamento e classificação de cenários críticos de poluição, podendo ser utilizado como sistemas de suporte à decisão em atividades de monitoramento e gerenciamento de qualidade de água.

Palavras-chave: Qualidade da água; Monitoramento; Análise por componentes principais; Visualização de dados.

Use of a principal component analysis based-model to identify critical conditions of surface water quality

Monitoring water quality is increasingly essential for water resource preservation. Data collected from monitoring stations are a rich source of water quality information, which is important for the protection of water resources, the maintenance of environmental capacity and the control of pollution. Increasing the analysis of the collected data is one of the ways to improve the decision-making process about those aspects. In this study, Principal Component Analysis (PCA), associated with the Hampel Filter, was applied to obtain a model used to monitor water quality. The case study refers to a water quality monitoring station (BV137), located in the Velhas river basin (Brazil). The data set used, made available by the water management agency in Minas Gerais - the Instituto Mineiro de Gestão das Águas (Minas Gerais Water Institute, IGAM), was collected over 21 years (1997-2018) and examined twelve physical, chemical and microbiological parameters. The PCA model, used as a monitoring system, was able to discriminate between critical and non-critical pollution events. Given a reference region for water quality, groups of critical events were recognized using the PCA score plots and associated with the sources of pollution characteristic of the study area (discharges of domestic and industrial effluents, agricultural and livestock activities and mining). The land use and occupation map of the monitoring station's contribution basin was also used for the analyzes. The most relevant water quality parameters were thermotolerant coliforms, total solids, turbidity and total arsenic. This application demonstrated the potential of the PCA technique for the identification of behavior patterns and classification of critical pollution scenarios, which can be used as decision support systems in monitoring and water quality management activities.

Keywords: Water quality; Monitoring; Principal component analysis; Data visualization.

Topic: **Engenharia de Recursos Hídricos**

Received: **03/04/2021**

Reviewed anonymously in the process of blind peer.

Approved: **26/04/2021**

Carolina Cristiane Pinto 
Universidade Federal de Minas Gerais, Brasil
<http://lattes.cnpq.br/4037071086115232>
<http://orcid.org/0000-0003-4580-4013>
ccristiane15@hotmail.com

Fábio Palmer Caldeira Parreiras de Faria
Universidade Federal de Minas Gerais, Brasil
<http://lattes.cnpq.br/5561768970044523>
fabioalmercp@gmail.com

Gustavo Matheus de Almeida 
Universidade Federal de Minas Gerais, Brasil
<http://lattes.cnpq.br/3191967289613425>
<http://orcid.org/0000-0002-2898-5177>
galmeyda@deq.ufmg.br



DOI: 10.6008/CBPC2179-6858.2021.004.0024

Referencing this:

PINTO, C. C.; FÁRIA, F. P. C. P.; ALMEIDA, G. M.. Utilização de modelo baseado em análise por componentes principais para identificação de condições críticas de qualidade de água superficial. **Revista Ibero Americana de Ciências Ambientais**, v.12, n.4, p.288-305, 2021. DOI: <http://doi.org/10.6008/CBPC2179-6858.2021.004.0024>

INTRODUÇÃO

A restauração de rios e a melhoria da qualidade das águas são umas das questões ambientais mais importantes do mundo contemporâneo (VOZA et al., 2018). As intervenções em bacias hidrográficas, visando à restauração de mananciais geralmente incluem: técnicas de proteção de nascentes e de áreas de recarga de aquíferos, redução da carga poluidora de fontes pontuais e difusas, recuperação de áreas degradadas, monitoramento da qualidade da água, entre outros (BAPTISTA et al., 2016).

Nesse contexto, os Programas de Monitoramento da Qualidade da Água (PMQA) são de extrema importância, pois eles proveem dados sobre o *status* e as mudanças temporais e espaciais da qualidade da água, podem ajudar na manutenção dos recursos hídricos, no uso da capacidade ambiental e no controle da poluição (CHEN et al., 2012; BEHMEL et al., 2016).

Como a qualidade da água não é uma condição estática, nem pode ser definida pela medição de apenas um parâmetro, além de ser variável no tempo e espaço, requer monitoramento de rotina para detectar padrões espaciais e mudanças ao longo do tempo (UNEP et al., 2008). Os principais elementos do monitoramento da qualidade da água são: medições no local ou ponto de amostragem, a coleta e análise de amostras de água, o estudo e a avaliação dos resultados analíticos e o relato dos resultados (BARTRAM et al., 1996). Com o avanço tecnológico das áreas de instrumentação, informática e bancos de dados, observou-se, ao longo dos últimos anos, a coleta e o acúmulo de grandes massas de dados. O desafio nesse novo cenário, seja no setor público ou privado, é transformar esses dados brutos em informação relevante. Assim, é necessária a utilização de métodos estatísticos para fins de monitoramento e diagnóstico.

O problema ao se trabalhar com dados de qualidade de água é o de reduzi-los a formas utilizáveis para a tomada de decisão (COURTEMANCH, 1994; ZIMMERMAN et al., 1996). Conforme explanado por Bodo (1989), dados de qualidade da água são de natureza complexa e apresentam características como presença de *outliers*, assimetria positiva, distribuição não normal dos dados, dados censurados, padrões sazonais, autocorrelação e dependência de outras variáveis não controláveis. Essas características impedem o uso de métodos estatísticos padrão. De maneira geral, os processos ambientais podem ser caracterizados como complexos, multifacetados, não lineares e dinâmicos (COTHERN et al., 1993). A interpretação e análise dos dados de qualidade de água não são, portanto, simples e requer conhecimentos do processo e de técnicas estatísticas específicas capazes de gerar informações sobre o estado da qualidade da água e auxiliar os gestores e órgãos responsáveis pela gestão do recurso hídrico na tomada de decisão e, conseqüentemente, na proteção das águas, e no seu uso sustentável e racional.

Nesse sentido, a Análise por Componentes Principais (ACP), pertencente à área de estatística multivariada, é uma das estratégias mais comuns para análise exploratória não supervisionada de dados (BALLABIO, 2015). Por ser capaz de lidar com a alta dimensão de sistemas, ao considerar a correlação espacial entre suas variáveis ou parâmetros (GE et al., 2013), a ACP vem sendo amplamente utilizada para a avaliação e interpretação de conjuntos de dados de qualidade de água (VEGA et al., 1998; SINGH et al., 2004; ZHANG et al., 2010; GATICA et al., 2012; MOHAMED et al., 2015; PHUNG et al., 2015; VOZA et al., 2018; CRUZ et al.,

2019).

A ACP aplicada a dados de monitoramento de qualidade de água cresceu consideravelmente, nos últimos anos, com estudos realizados em diferentes países. Porém, conforme relatado por Sergeant et al. (2016), ainda se encontram erros analíticos e descuidos evitáveis em muitos estudos mais atuais de aplicações de ACP em dados de qualidade de água. Em um estudo de revisão realizado por Olsen et al. (2012), sobre a aplicação de PCA a amostras de água, os autores ressaltaram que nos artigos revisados as discussões sobre tratamento de dados antes das avaliações estatísticas eram inadequadas ou até ausentes. Eles verificaram também que comumente não há discussão de como os dados faltantes ou como os valores não detectados foram tratados; em alguns estudos não há informação de como os dados foram padronizados ou se foram padronizados, além da falta de discussão sobre a detecção e exclusão de *outliers*. Portanto, há potencial para melhorias no que diz respeito às aplicações da ACP no monitoramento da qualidade da água. Além disso, a ACP pode ser utilizado com diversos objetivos (BRO et al., 2014). Os estudos de avaliação da qualidade da água geralmente consideram várias estações simultaneamente, dada uma região, sub-bacia ou curso do rio, e baseiam sua análise em *loadings* (pesos) (BENGRÁINE et al., 2003; SINGH et al., 2004; PHUNG et al., 2015; VOZA et al., 2018).

Neste contexto, este estudo tem como objetivo a utilização de ACP, em combinação com Filtro de Hampel, para desenvolver um modelo para o mapeamento e classificação de fontes de poluição, determinadas pelos valores críticos dos parâmetros de qualidade de água monitorados. O presente trabalho propõe uma abordagem local, com foco em uma estação de monitoramento de água. Uma abordagem de qualidade da água local melhora a compreensão de detalhes e aspectos locais e quando associada à abordagem global comumente usada pode levar a uma tomada de decisão mais racional. O estudo de caso refere-se a uma estação de monitoramento localizada na bacia hidrográfica do rio das Velhas, uma das principais sub-bacias da bacia do rio São Francisco (Brasil).

METODOLOGIA

A Figura 1 ilustra a metodologia proposta neste trabalho para o mapeamento e a classificação das fontes de poluição, a partir da análise de ACP. Cada etapa é descrita abaixo.

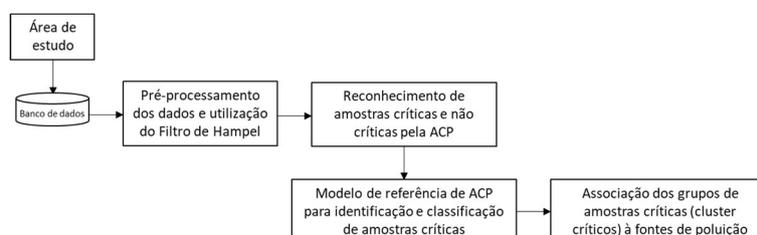


Figura 1: Metodologia adotada neste estudo.

Área de estudo

A bacia do rio das Velhas está localizada no estado de Minas Gerais (Brasil), e seu principal rio - o rio das Velhas - é o maior afluente da bacia do rio São Francisco, com 806,84 km. A bacia compreende uma área

de 27.850 km², com uma população de aproximadamente 4,4 milhões de habitantes. A Região Metropolitana de Belo Horizonte (RMBH), capital do estado de Minas Gerais, ocupa apenas 10% da área da bacia, mas tem mais de 70% dessa população (CBH VELHAS, 2015). De acordo com Costa et al. (2017) e Trindade et al. (2017), essa bacia é a mais poluída e uma das principais responsáveis pela deterioração da qualidade das águas superficiais do rio São Francisco.

A estação de monitoramento BV137, localizada na calha do rio das Velhas, no município de Lagoa Santa (que faz parte da Região Metropolitana de Belo Horizonte), foi definida como estudo de caso. Esta é uma das estações prioritárias para o monitoramento e controle do rio das Velhas (CALAZANS et al., 2018). A Figura 2 mostra a localização da área de estudo e a estação de monitoramento da qualidade da água selecionada.

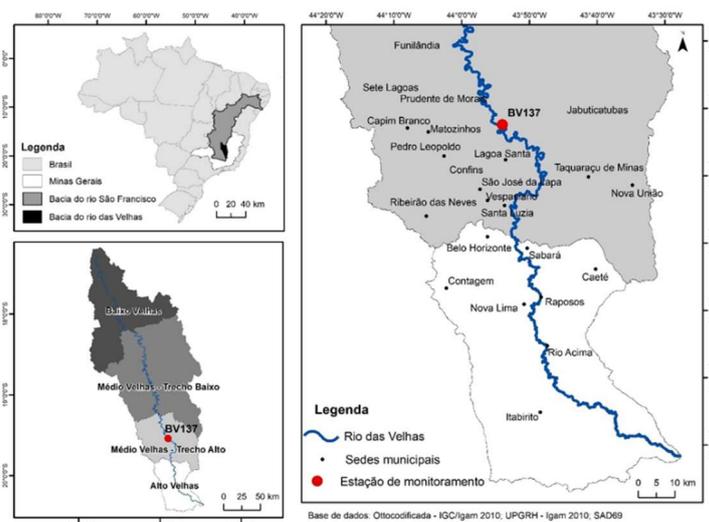


Figura 2: Localização da área do estudo de caso e da estação de monitoramento de qualidade da água superficial.

A bacia do rio das Velhas é dividida em quatro regiões fisiográficas - Alto, Médio Alto, Médio Baixo e Baixo (Figura 2) - em função das características do uso do solo (Quadro 1).

Quadro 1: Regiões da bacia hidrográfica do rio das Velhas e principais características.

Regiões	Área total (km ²)	Principais características	Principais fatores de pressão
Alto	2.739,76	Apresenta maior contingente populacional, com uma expressiva atividade econômica concentrada, principalmente, na RMBH. Nessa região encontra-se o sistema de abastecimento integrado rio das Velhas Compreende toda a região denominada Quadrilátero Ferrífero, responsável pela maior produção nacional de minério de ferro.	Esgotos industriais e domésticos não tratados e os efluentes gerados pelas atividades minerárias
Médio Alto	4.276,01	Apresenta menor concentração populacional que o Alto rio das Velhas	Predomínio das atividades agrícolas e pecuárias.
Médio Baixo	12.204,16	Representa a maior porção dentro da bacia do rio das Velhas	
Baixo	8.630,07	Baixa concentração populacional e nenhum dos municípios tem 100% do território inserido na bacia.	

Fonte: Adaptado CBH Velhas (2015).

A estação de monitoramento BV137 está localizada na região do Médio Alto rio das Velhas, que apresenta menor concentração populacional quando comparada ao trecho Alto, com predominância de atividades agropecuárias (CBH VELHAS, 2015). A irrigação é a atividade mais significativa no trecho médio da bacia, com maior percentual de área mecanizada e, conseqüentemente, o uso de maior volume de

fertilizantes e defensivos e intensificação do potencial de erodibilidade (CAMARGOS, 2005). O levantamento dos fatores de pressão que influenciam a qualidade da água na estação BV137 foi realizado considerando a área de drenagem incremental desta estação.

As classes de cobertura do solo consideradas foram: cobertura natural (afloramento rochoso, vegetação arbustiva, vegetação arbórea, hidrografia) e uso antrópico (agropecuária, área irrigada, queimada, silvicultura, área urbana e mineração) (CBH VELHAS, 2015), conforme apresentado na Figura 3.

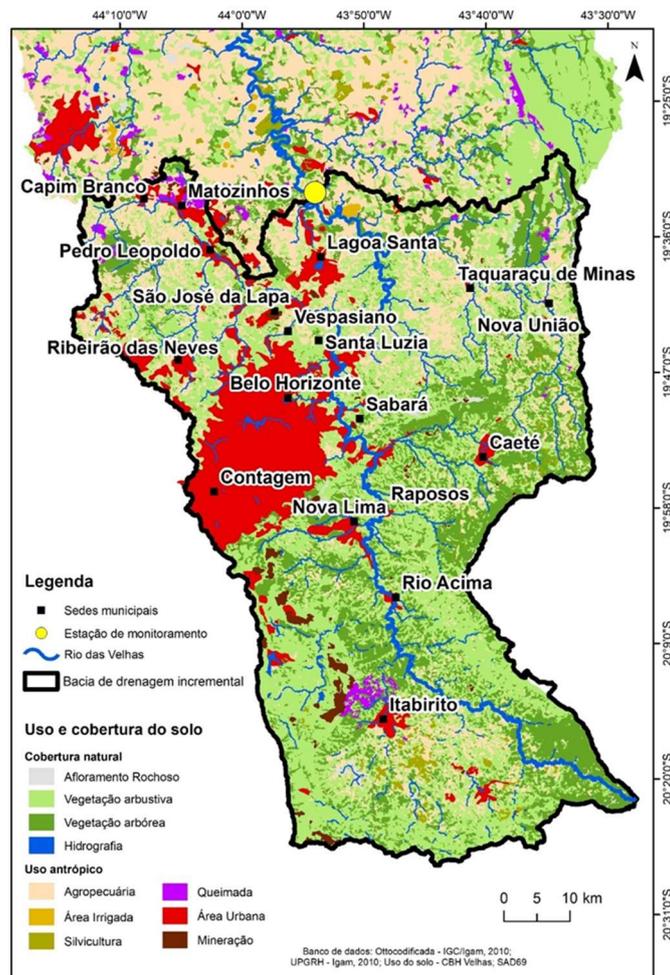


Figura 3: Mapa de uso e ocupação do solo da bacia de drenagem da estação de monitoramento BV137.

Banco de dados

No estado de Minas Gerais, a rede de monitoramento da qualidade da água é gerenciada pelo Instituto Mineiro de Gestão das Águas (IGAM), através de um programa conhecido como Programa Águas de Minas, desde 1997. Neste estudo foi utilizado um banco de dados histórico referente ao monitoramento da qualidade da água na estação BV137, incluindo os principais parâmetros físico-químicos e microbiológicos para a avaliação de uma rede de monitoramento de qualidade de água (Tabela 1).

Tabela 1: Parâmetros físico-químicos e microbiológicos da qualidade da água com respectivo símbolo, unidade e técnica analítica.

Parâmetros	Símbolo	Unidade	Técnica analítica
Arsênio total	As _T	mg.L ⁻¹ As	Espectrometria de absorção atômica com geração de hidreto
Cloreto	Cl ⁻	mg.L ⁻¹ Cl	Colorimétrico
Condutividade elétrica <i>in loco</i>	CE	µS.cm ⁻¹	Condutimétrico

Parâmetros	Símbolo	Unidade	Técnica analítica
Coliformes termotolerantes	Colif. term.	NMP ^a /100 ml	Tubos múltiplos
Demanda bioquímica de oxigênio	DBO	mg.L ⁻¹ O ₂	Titulométrico
Fósforo total	P _T	mg.L ⁻¹ P	Colorimétrico
Nitrato	N-NO ₃ ⁻	mg.L ⁻¹ N	Potenciométrico
Oxigênio dissolvido	OD	mg.L ⁻¹ O ₂	Titulométrico
pH <i>in loco</i>	pH	-	Potenciométrico
Sólidos total	ST	mg.L ⁻¹	Gravimétrico
Temperatura da água	T _{água}	°C	Termométrico
Turbidez	Turb.	NTU	Turbidimétrico

^a Número mais provável.

A amostragem dos parâmetros de qualidade da água, bem como o transporte e as análises laboratoriais foram realizados por um laboratório acreditado pelo Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO). Os resultados do monitoramento estão disponíveis no site do IGAM.

Assim, a base de dados selecionada para análise é composta por 12 parâmetros de qualidade da água, monitorados de 1997 a 2018 (monitoramento inicialmente trimestral e a partir de 2008 mensal), formando uma matriz de dados brutos $X_{[n=148, p=12]}$ (em que n é o número de amostras e p o número de parâmetros), correspondendo a um total de 1.774 dados válidos.

Pré-processamento dos dados

Inicialmente, efetuou-se uma inspeção visual dos dados, a partir de histogramas e gráficos de dispersão e de séries temporais, com o objetivo de identificar registros faltosos e registros errôneos de análises laboratoriais, por exemplo. Com essa análise exploratória, conheceu-se melhor o conjunto de dados e verificou-se sua adequação para a presente proposta de estudo.

Os dados censurados (abaixo ou acima do limite mínimo de detecção do método analítico) foram substituídos pelo valor do limite de detecção do método analítico. No caso de dados faltantes, toda a amostra foi excluída do banco de dados. Em sequência, realizou-se a análise descritiva por parâmetro (cálculo de média, desvio-padrão, mínimo, máximo, percentis 25, 50 e 75 e o coeficiente de variação). A normalidade dos dados foi verificada, para cada parâmetro, pela aplicação do teste de Shapiro-Wilk ao nível de significância (α) de 5% (SHAPIRO et al., 1965).

Num segundo momento, identificaram-se, de modo quantitativo, os dados discrepantes (*outliers*). Estatisticamente os *outliers* podem ser definidos como valores pertencentes a uma população diferente porque se originam de outro processo ou fonte (REIMANN et al., 2008).

O método $\pm 3\sigma$, comumente utilizado para a detecção de *outliers*, por usar a média aritmética simples (\bar{X}) é sensível a dados extremos. Com isso, adotou-se o Filtro de Hampel, que utiliza a mediana (DAVIES et al., 1993; LIN et al., 2007). Nesse caso, calcula-se a estatística MAD (*Median Absolute Deviation*) (Equação 1a), a partir dos desvios absolutos das n observações (x_i ; $i = 1, 2, \dots, n$) em relação à mediana (\tilde{X}), dado um parâmetro em particular. Os limites inferior (Equação 1b) e superior (Equação 1c) são mais estreitos em relação à abordagem tradicional.

$$MAD = 1,4868 \cdot \text{Mediana}\{|x_i - \tilde{X}|\} \quad (1a)$$

$$LS \text{ (Limite Superior)} = \bar{X} + 3 \cdot MAD \quad (1b)$$

$$LI \text{ (Limite Inferior)} = \bar{X} - 3 \cdot MAD \quad (1c)$$

A estatística MAD é um método robusto a valores extremos presentes no conjunto de dados, além de ser tolerável a desvios consideráveis da distribuição normal (REIMANN et al., 2008), sendo, portanto, mais adequado para análise de dados ambientais (SABINO et al., 2014).

O Filtro de Hampel foi aplicado a cada parâmetro e identificou-se os valores atípicos. Para verificar se esses valores poderiam ser considerados ou não erros de medição, cada observação foi analisada, assim como os demais parâmetros coletados na mesma data, as coletas realizadas em datas próximas e qual a condição do tempo na data da coleta.

Na etapa seguinte, em caso de pelo menos um valor atípico para quaisquer parâmetros, a respectiva amostra foi rotulada como crítica. Então, após a aplicação do filtro de Hampel as amostras foram classificadas em dois grupos (críticas e não críticas), e passaram a ser identificadas com um rótulo (*label*) do grupo correspondente. As amostras classificadas como críticas resultam de valores elevados em pelo menos um dos parâmetros de qualidade, sendo um indicativo de situações indesejáveis (por exemplo, lançamento de esgotos sanitários, efluentes industriais, fontes não pontuais de nutrientes, processos erosivos, entre outros). As amostras não críticas resultantes foram usadas como comportamento de referência para os parâmetros de qualidade da água.

Análise por Componentes Principais

A Análise por Componentes Principais (ACP) é uma técnica estatística multivariada. Sua formulação matemática foi fornecida por Pearson (1901) e Hotelling (1933). O princípio é explicar a maior parte da variação total em um subconjunto de variáveis original por poucas variáveis latentes chamadas componentes principais (CP). Isso é alcançado por uma rotação ortogonal do sistema de coordenadas original através de um operador linear. As coordenadas dos pontos originais neste sistema girado são chamadas de *scores* (SHARMA, 1995; MANLY, 2008; HAIR et al., 2009). Assim, a ACP é uma técnica de redução de dimensionalidade, útil para facilitar a visualização e o entendimento de problemas multivariáveis como os de qualidade da água.

Neste trabalho a ACP, em conjunto com a classificação obtida a partir do Filtro de Hampel, foi utilizada para a construção de dois modelos para o reconhecimento e a classificação de eventos críticos de poluição das águas superficiais. No primeiro modelo todo o conjunto de dados foi utilizado, ou seja, amostras classificadas em críticas e não críticas, com o objetivo de verificar a capacidade da ACP em discriminá-las. Uma resposta afirmativa significa que ele pode ser empregado para reconhecer eventos críticos de qualidade da água por comparação com uma referência composta apenas por eventos não críticos.

Para realizar essa análise os dados foram padronizados em escala z (média 0 e desvio padrão 1), evitando a prevalência daqueles parâmetros com maiores desvios-padrões sobre a definição dos eixos (as componentes principais) no novo sistema (rotacionado) de coordenadas (SERGEANT et al., 2016). A avaliação foi realizada utilizando os gráficos de *scores* (*score plots*) considerando todas as combinações de pares entre

os componentes principais.

Dada a capacidade da ACP de identificar amostras críticas e não críticas, um segundo modelo de PCA foi obtido usando apenas amostras não críticas – um modelo de referência. Esse modelo foi, então, alimentado com as amostras críticas. Esses dados foram inicialmente padronizados usando as médias dos parâmetros e desvios padrão obtidos com as amostras não críticas. Os *scores* foram calculados e plotados nos gráficos de *scores* (*score plots*) em relação à região de referência fornecida pelo modelo ACP. A análise visual desses gráficos permitiu a identificação de agrupamentos críticos, ou seja, conjunto de amostras que possuem valores elevados de determinados parâmetros. Com a identificação dos parâmetros de qualidade de água com maior influência sobre um agrupamento em particular, associou-se esse agrupamento a fontes de poluição presentes na área de estudo. O mapa de uso e ocupação do solo da região também foi utilizado para auxiliar as análises. Para a comparação dos valores dos parâmetros de qualidade dos diferentes agrupamentos críticos em relação à região de referência, foi aplicado o teste não paramétrico de Mann-Whitney, ao nível de significância de 5% ($\alpha=5\%$).

Todas as análises estatísticas realizadas foram implementadas em linguagem de programação R versão 4.0.3¹, de licença livre e código aberto. Também foram gerados gráficos no MATLAB R2020a.

RESULTADOS E DISCUSSÃO

Pré-processamento dos dados

A partir da matriz de dados brutos $X_{[n=148, p=12]}$, as análises preliminares dos dados consistiram na elaboração de histogramas e gráficos de dispersão e de séries temporais para conhecer o conjunto de dados. Para exemplificar, a Figura 4 apresenta o histograma para o parâmetro cloreto total e o gráfico de série temporal para sólidos totais. Dados anômalos podem ser vistos em ambos, à direita em (a) e como picos em (b). Este trabalho faz uso deles para primeiro identificar e, então, classificar eventos críticos com relação às fontes de poluição da qualidade da água.

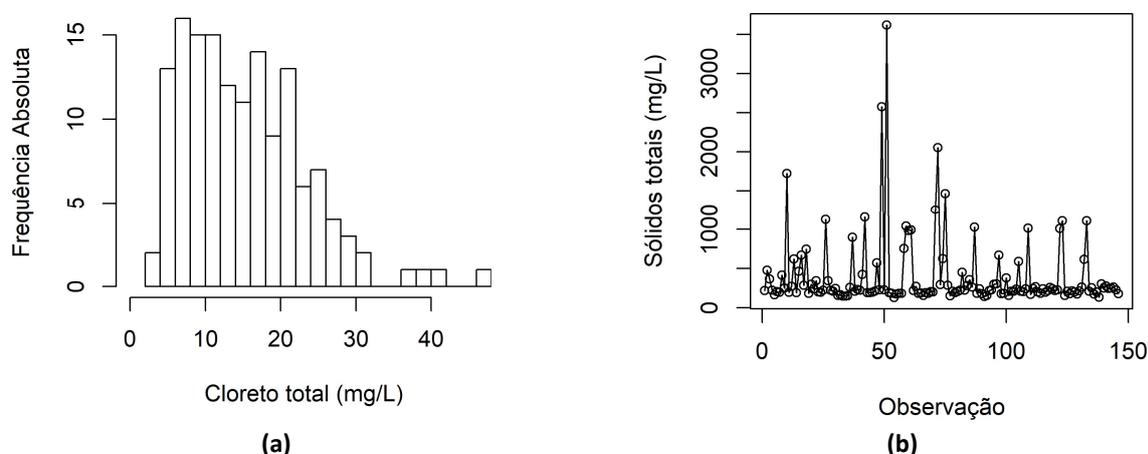


Figura 4: (a) Histograma para o parâmetro cloreto total e (b) gráfico temporal para o parâmetro sólidos totais.

¹ <https://www.r-project.org/>

Duas amostras que continham dados faltantes foram excluídas do banco de dados. Com isso, o conjunto de dados de trabalho passou a conter 146 registros. A Tabela 2 apresenta um resumo da estatística descritiva dos 12 parâmetros de qualidade da água.

Tabela 2: Resumo da estatística descritiva dos parâmetros de qualidade da água, entre 1997 e 2018, para a estação de monitoramento BV137.

Parâmetro	Mínimo	Máximo	Média	Desvio-padrão	Percentil 25	Mediana	Percentil 75	CV ¹ (%)
AS _T	0,001	0,138	0,033	0,021	0,020	0,029	0,040	62,5
Cl ⁻	3,27	46,10	15,28	8,12	8,88	14,05	20,53	53,0
Colif. term.	50	160000	41640	58385	3300	13000	49028	139,7
CE	96	524	255	97	177	238	311	38,1
DBO	2,00	131,00	9,29	12,26	4,93	6,45	9,00	131,5
P _T	0,010	1,610	0,412	0,267	0,230	0,335	0,568	64,7
N-NO ₃ ⁻	0,010	6,660	0,893	0,951	0,235	0,665	1,118	106,1
OD	0,50	8,10	3,38	1,38	2,60	3,55	4,175	40,6
pH	6,00	8,00	7,12	0,37	6,90	7,20	7,375	5,1
S _T	130	3622	395	460	194	226	304	116,1
T _{água}	18,00	31,70	24,80	2,54	22,80	25,00	26,90	10,2
Turb.	4	3100	174	382	17	40	136	218,8

¹CV - coeficiente de variação = desvio padrão/média

Ressaltam-se os elevados valores de coliformes termotolerantes e concentração de DBO, com valores médios de 41.640 NMP/100 mL e 9,29 mg/L, respectivamente; e o parâmetro oxigênio dissolvido com percentil 75 igual a 4,18 mg/L, abaixo do limite legal da DN COPAM/CERH-MG n° 01/2008 (MINAS GERAIS, 2008) para a classe de enquadramento do curso d'água (classe 3).

O coeficiente de variação (CV) é uma medida relativa da variabilidade dos dados (desvio padrão/média). Os menores valores de CV foram observados para os parâmetros pH, T_{água}, CE e OD, indicando menor variabilidade nos dados desses parâmetros. A maioria dos parâmetros apresentaram valores de CV altos indicando grande variabilidade dos dados das amostras, característica intrínseca desse tipo de dado, devido às influências naturais e antrópicas a que os cursos de água estão sujeitos. Comprovado também pela amplitude (máximo-mínimo) elevada dos dados. Resultados semelhantes foram descritos por Toledo et al. (2002) e Follador (2010).

Os testes de normalidade de Shapiro-Wilk foram aplicados para cada parâmetro de qualidade. Nenhum deles apresentou aderência à distribuição normal ao nível de significância (α) de 5%. Resultados já esperados, visto que os dados de qualidade de água geralmente são caracterizados por uma distribuição não normal (HELSEL et al., 2002).

Para a identificação dos *outliers* foi aplicado o Filtro de Hampel. Identificados os valores atípicos, analisou-se cada uma das observações individualmente, para verificar se esses valores podiam ser considerados ou não erros de medição. Com base nessa análise, nenhum dos valores discrepantes foram excluídos. Os dados utilizados nessa pesquisa foram consistidos e confirmados, sendo os *outliers* indicativos de condições críticas de qualidade da água. De acordo com Dressing et al. (2016), em alguns casos, os valores extremos podem ser mais importantes para apontar a necessidade de atenção e/ou intervenção.

Então, de acordo com a metodologia, após aplicação do Filtro de Hampel, as observações foram classificadas em dois grupos: críticas (74 registros) e não críticas (72 registros). A Figura 5(a-b) mostra os

gráficos temporais, com a identificação desses tipos de observações, para coliformes termotolerantes e nitrato, respectivamente.

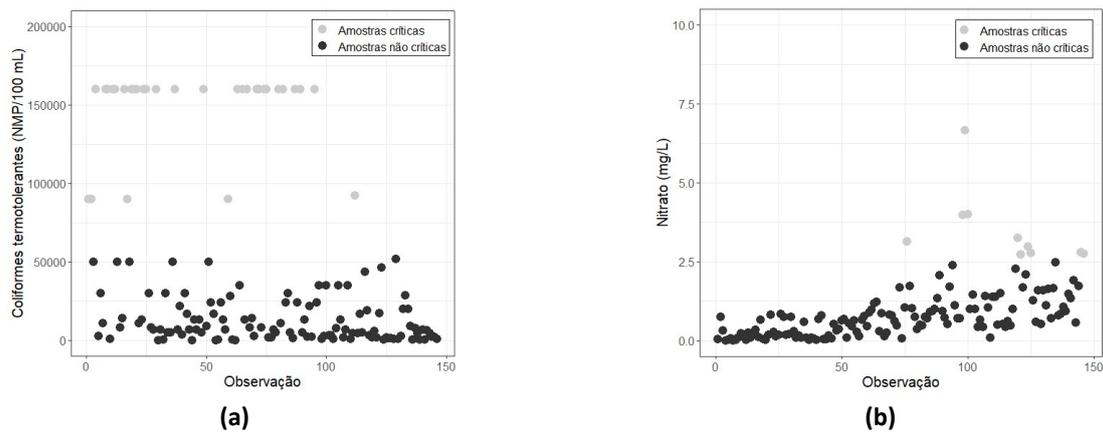


Figura 5: Gráficos com dados discrepantes e não discrepantes identificados pelo Filtro de Hampel para os parâmetros (a) coliformes termotolerantes e (b) nitrato.

Análise por Componentes Principais

O primeiro modelo de ACP foi construído e verificou-se sua capacidade em discriminar as amostras críticas e não críticas fornecidas pelo Filtro de Hampel. Todos os gráficos de *scores* entre os pares de componentes principais, desde CP1-CP2, CP1-CP3, ..., até CP11-CP12 foram analisados.

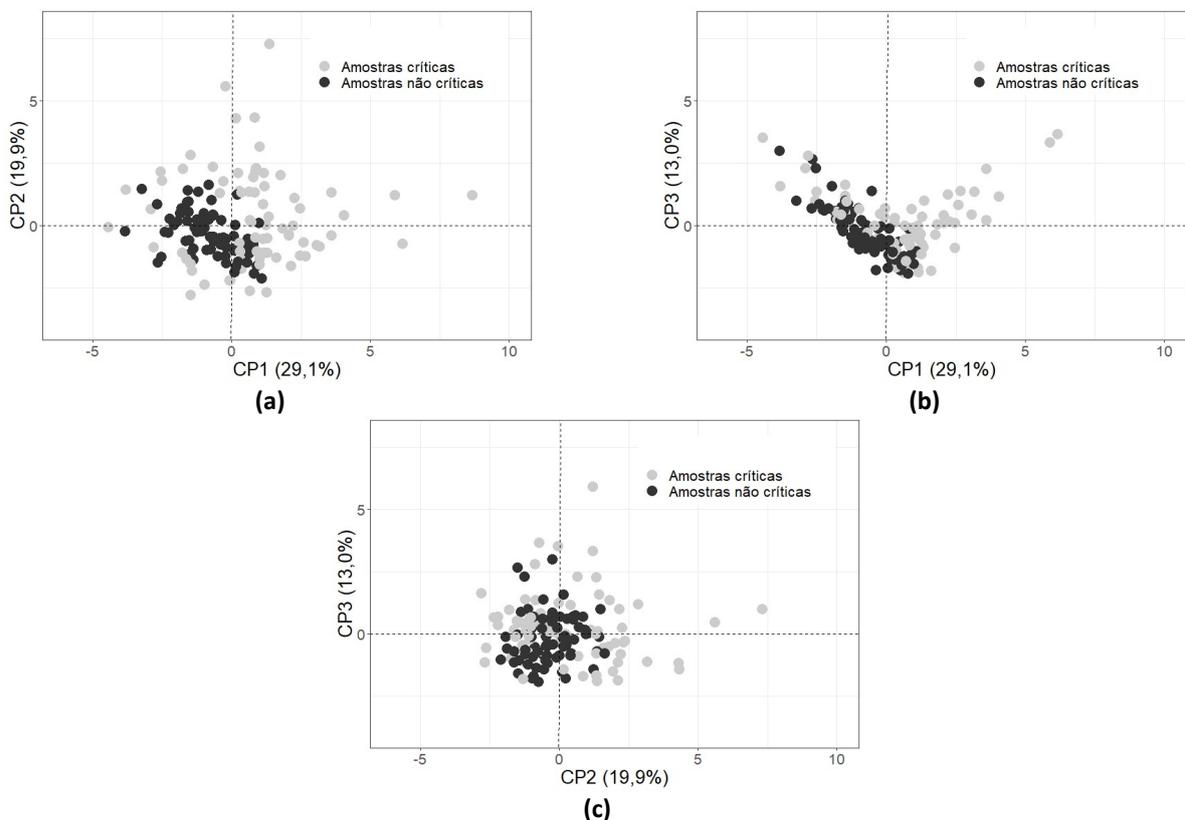


Figura 6: Gráficos de *scores*: (a) CP1-CP2, (b) CP1-CP3, (c) CP2-CP3, (d) CP1-CP8.

A Figura 6 (a-c) mostra gráficos de *scores* para os componentes principais mais relevantes (CP1, CP2, CP3), com as respectivas variâncias explicadas entre parênteses. Os resultados são análogos para todas as

combinações de pares possíveis. Em geral, observa-se os dados não críticos concentrados numa nuvem de pontos (círculos sólidos), e os dados críticos no entorno deles, concentrados ou dispersos (círculos vazados). Portanto, há uma correspondência entre a técnica de ACP e a classificação segundo Hampel. A estratificação dos dados é consequência das diferenças existentes entre os valores das variáveis usadas na ACP. Do ponto de vista prático, a nuvem de pontos resultante dos *scores* não críticos pode ser tomada como uma região de referência.

Portanto, confirmada a capacidade de ACP em discriminar as amostras críticas e não críticas, o segundo modelo de ACP foi construído. Os resultados estão resumidos na Tabela 3. O percentual de variância total dos dados originais explicada pelas três primeiras componentes ($k = 3$) é igual a 59,5%. A utilização desse número de CPs é comum, principalmente, para análise de agrupamento visual, conforme feito neste trabalho, para facilitar a interpretação do problema.

Tabela 3: Pesos (*weights*) do modelo PCA obtido a partir de amostras não críticas.

Parâmetro /Componente Principal (CP)	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12
As _T	-0,292	0,119	0,289	-0,415	-0,231	-0,280	0,497	-0,313	0,309	-0,139	-0,221	0,069
Cl ⁻	-0,443	-0,074	0,119	0,025	0,314	0,300	-0,167	-0,260	0,102	0,237	-0,255	-0,607
Colif. term.	0,287	-0,082	0,432	-0,148	-0,026	-0,111	-0,656	-0,311	0,084	-0,386	-0,079	0,021
CE	-0,471	-0,064	0,159	0,070	0,161	0,197	-0,199	-0,156	-0,028	0,232	0,069	0,745
DBO	-0,008	0,113	0,517	0,558	-0,408	-0,018	0,150	-0,118	-0,423	0,121	-0,046	-0,099
P _T	-0,214	-0,372	0,236	-0,312	0,310	-0,394	0,056	0,244	-0,557	-0,057	0,157	-0,102
N-NO ₃ ⁻	-0,207	0,444	-0,231	-0,262	-0,292	-0,189	-0,293	-0,264	-0,196	0,227	0,495	-0,163
OD	-0,283	0,378	-0,333	0,152	0,064	-0,227	-0,172	0,071	-0,297	-0,371	-0,566	0,082
pH	-0,363	0,030	0,204	0,164	-0,231	-0,231	-0,264	0,630	0,450	-0,018	0,111	-0,101
S _T	-0,061	0,465	0,288	-0,195	0,162	0,554	0,133	0,246	-0,121	-0,419	0,238	-0,032
T _{ÁGUA}	0,132	0,378	0,147	0,340	0,621	-0,421	0,122	-0,149	0,219	0,045	0,230	0,003
Turb.	0,305	0,344	0,239	-0,346	0,066	-0,023	-0,112	0,291	-0,070	0,581	-0,401	0,080
Variância (Componente Principal)	3,5	2,1	1,6	1,0	0,8	0,7	0,6	0,6	0,4	0,3	0,2	0,1
Variância Explicada (%)	29,0	17,5	13,1	8,5	6,6	5,9	5,1	4,9	3,6	2,8	2,1	1,0
Variância Explicada Acumulada (%)	29,0	46,5	59,5	68,0	74,6	80,5	85,7	90,5	94,1	96,9	99,0	100,0

Uma vez que o modelo PCA foi obtido usando apenas amostras não críticas, seus *scores* podem ser usados como região de referência. Assim, calculou-se os *scores* para as observações críticas utilizando esse modelo PCA. Três regiões (Região 1, Região 2 e Região 3) entre o conjunto de dados críticos foram identificados, conforme a descrição a seguir. Como essas três regiões apresentam valores muito elevados de alguns parâmetros de qualidade de água, se destacando das demais amostras, essas regiões foram visíveis em todas as combinações possíveis de CPs.

A região de referência foi envolta por um limite de controle calculado com nível de confiança de 95% (refere-se à elipse nos gráficos de *scores*). Ela é relativa ao conjunto de dados não críticos, utilizado para a construção do modelo de ACP (Tabela 3). Pode-se observar que a nuvem de dados não críticos se localiza sempre dentro do limite de controle, conforme desejado (Figura 7).

A Região 1 (em vermelho) localiza-se no interior do limite de controle no plano CP1-CP2 (Figura 7-a); porém, fora dele nos planos CP1-CP3 (Figura 7-b) e CP2-CP3 (Figura 7-c). Observa-se que não há uma relação significativa com CP2.

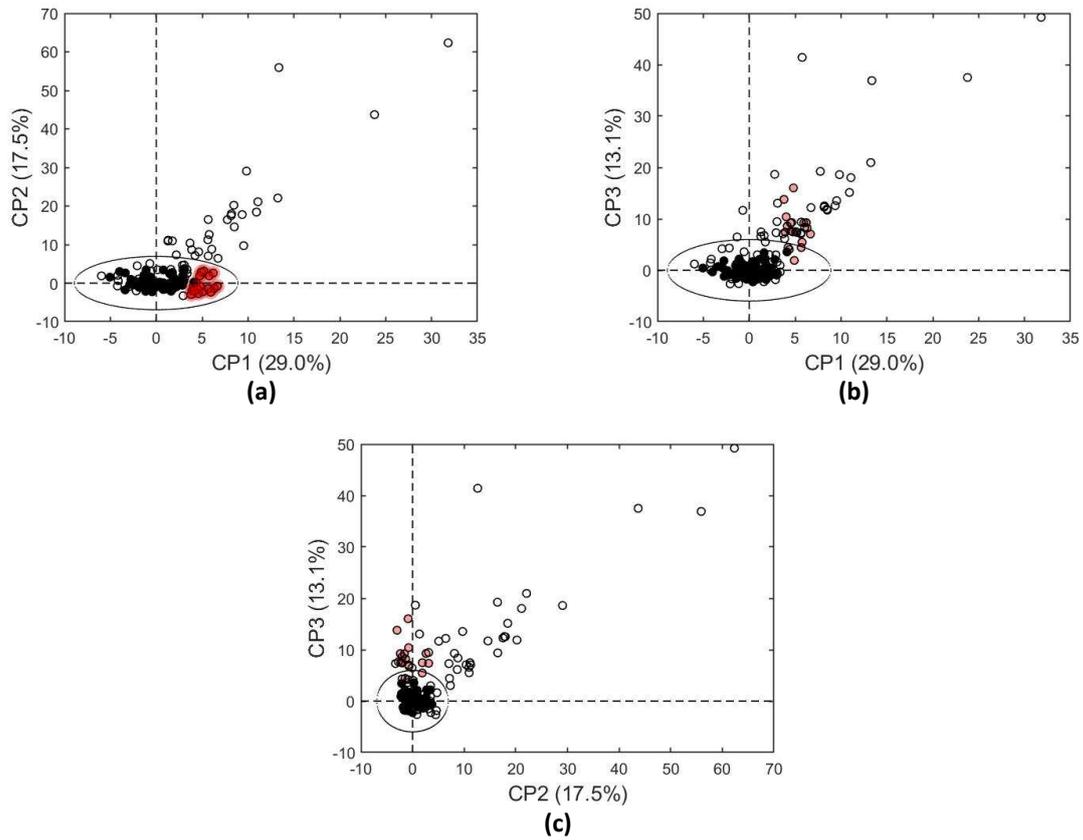


Figura 7: Gráficos de scores com destaque na Região 1: (a) CP₁-CP₂, (b) CP₁-CP₃, e (c) CP₂-CP₃.

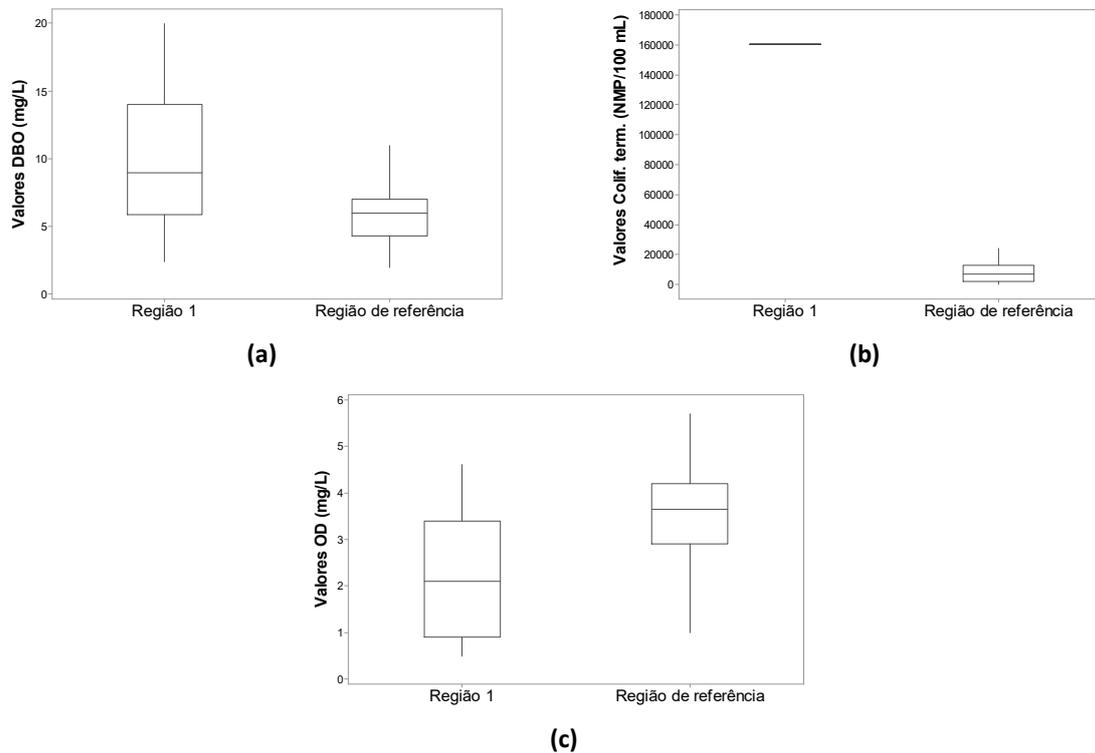


Figura 8: Box-plot para comparação das concentrações entre a Região 1 e a região de referência para os parâmetros: (a) DBO, (b) colif. term., e (c) OD.

Para comparar os valores das concentrações dos parâmetros que se destacaram na Região 1 em relação à região de referência, foi realizado o teste de Mann-Whitney. Na Região 1, as concentrações de colif. term. (Figura 8(a)) e DBO (Figura 8(b)) são significativamente superiores (p -valor $< 0,05$) e as concentrações

de OD (Figura 8(c)) significativamente inferiores (p -valor $< 0,05$) em relação à região de referência.

Essa condição é um indicativo, sobretudo, do lançamento de efluentes domésticos não tratados ou tratados de forma insuficiente. Como a estação BV137 está localizada a jusante de municípios da RMBH, ela recebe influência dos lançamentos de esgotos de grandes centros urbanos como, por exemplo, Sabará, Belo Horizonte, Contagem, Santa Luzia e Lagoa Santa (Figura 2), sendo um trecho crítico em termos de poluição (IGAM, 2010; IGAM, 2018). Verificou-se que todas as amostras da Região 1, à exceção de duas, apresentaram valores de colif. term. superiores a 160.000 NMP/100 mL.

O município imediatamente a montante da estação BV137 é o de Lagoa Santa (que faz parte da RMBH). De acordo com levantamento realizado pela Agência Nacional de Águas e Saneamento Básico - ANA, esse município que possui uma população urbana em torno de 53.000 habitantes (censo demográfico de 2010 do Instituto Brasileiro de Geografia e Estatística), gera uma carga orgânica total estimada associada ao esgoto de cerca de 2.900 kg DBO/dia. Quase metade dessa carga total gerada (42,5%) não é coletada e nem tratada e é disposta diariamente a céu aberto ou encaminhadas para soluções precárias de esgotamento (ANA, 2017). Esse *déficit* de coleta e tratamento de esgotos resulta em uma parcela significativa de carga poluidora que atinge o curso de água, causando implicações negativas aos usos múltiplos dos recursos hídricos, especialmente o abastecimento humano, além de poder impactar na saúde da população (ANA, 2017).

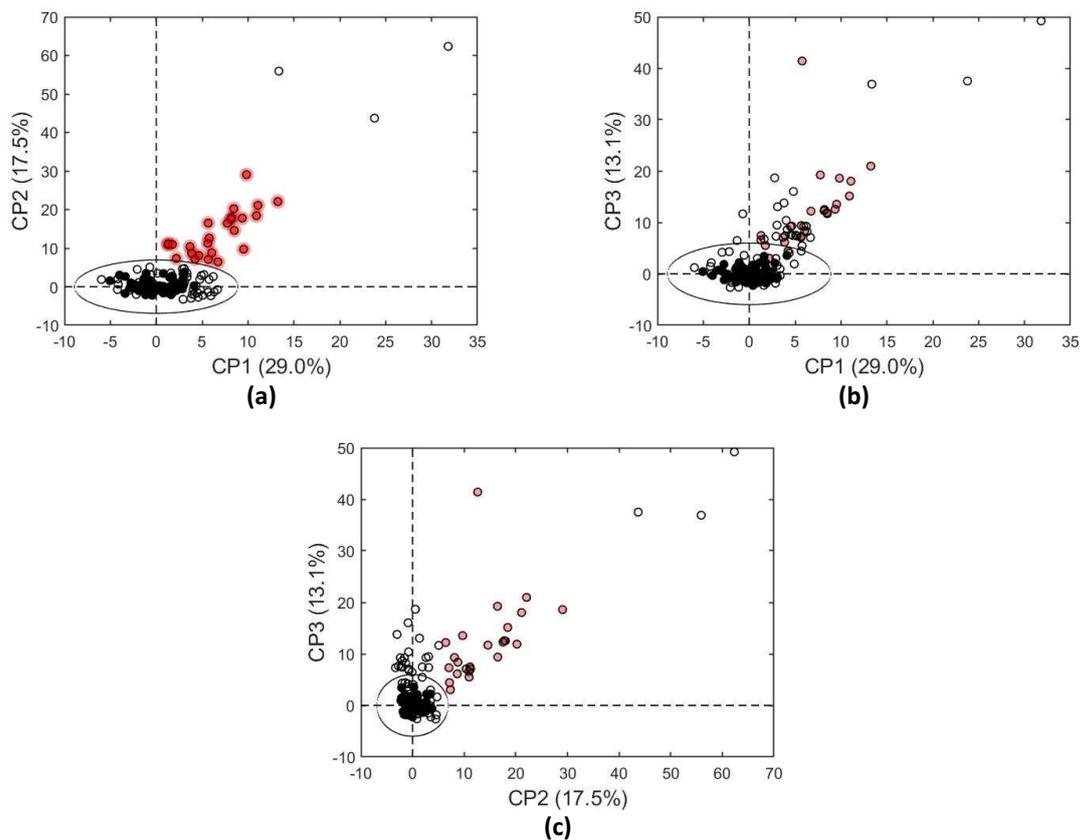


Figura 9: Gráficos de scores: (a) CP₁-CP₂, (b) CP₁-CP₃, e (c) CP₂-CP₃.

Outra fonte de poluição que contribui para essa condição são os lançamentos de efluentes industriais (principalmente de indústrias de papel, têxteis e de abate de animais), que resultam também em grandes quantidades de matéria orgânica no rio. Tanto os lançamentos de esgotos quanto os lançamentos de

efluentes industriais, nesse trecho do rio, comprometem os níveis de oxigênio na água.

Alguns dos municípios com as condições de qualidade mais críticas identificados em um estudo do IGAM na bacia do rio das Velhas (IGAM, 2013) estão localizados a montante da estação BV137 (Figura 2), justificando os resultados de colif. term., DBO e OD encontrados na estação BV137.

A Figura 9 mostra a Região 2 (em vermelho). A facilidade de reconhecimento de um mesmo agrupamento varia entre as combinações de componentes principais. Para essa região, os planos candidatos em uma atividade de monitoramento seriam CP1-CP3 e CP2-CP3.

Na Região 2, os valores de turb. e ST são significativamente superiores (p -valor $< 0,05$) (Figura 10) às concentrações da região de referência. Esses resultados podem estar relacionados, principalmente, às extensas atividades agropecuárias, predominantes nesta área, conforme apresentado no mapa de uso e ocupação do solo do rio das Velhas (Figura 3). Essas atividades contribuem para intensificar os processos erosivos que, em combinação com as ações pluviais regulares, resultam no transporte de componentes expostos do solo para os ambientes aquáticos causando a degradação do solo e assoreamento dos cursos d'água (IGAM, 2010).

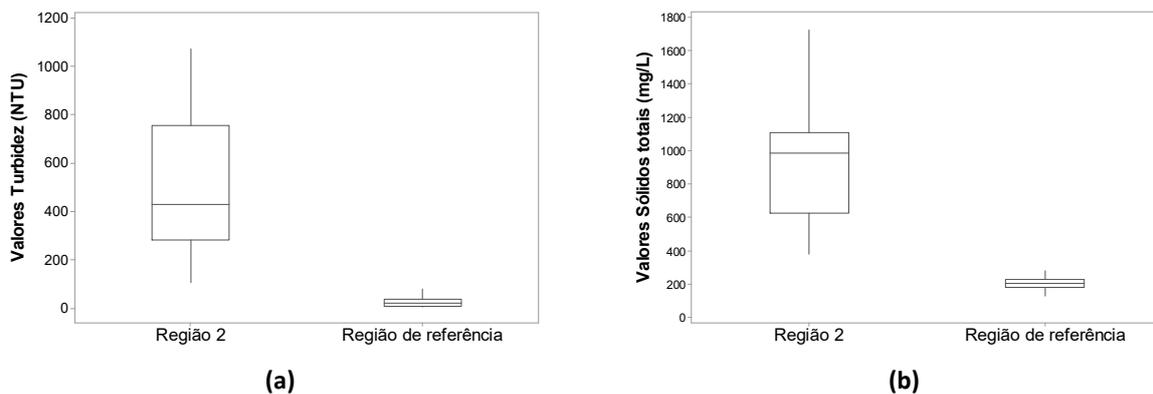


Figura 10: Box-plot para comparação das concentrações entre a região 2 e a região de referência para os parâmetros: (a) turbidez e (b) sólidos totais.

A Figura 11 é um exemplo de uma condição extrema, contendo três observações consideravelmente discrepantes, constituindo a Região 3. Essa última região é característica de concentrações significativamente altas, sobretudo, de arsênio total, sólidos totais e turbidez, em relação à região de referência. Dada a magnitude dessa discrepância, essas amostras críticas foram reconhecíveis em qualquer dos planos de componentes principais.

As altas concentrações de arsênio e sólidos são um indicativo da presença de atividades minerárias e processos erosivos na região a montante da estação BV137 – a região do Alto rio das Velhas.

O Alto rio das Velhas compreende parte da área denominada Quadrilátero Ferrífero (CBH VELHAS, 2015). De acordo com Borba (2002) e Borba et al. (2003), esta área apresenta uma anomalia natural de arsênio que, com a ação do intemperismo das rochas ao longo do tempo, é liberado para sedimentos, águas superficiais e subterrâneas. Além disso, devido às antigas atividades de mineração no passado, as fontes locais e regionais de contaminação por arsênio também estão relacionadas a passivos ambientais (solo contaminado, pilhas de rejeitos antigas, minas antigas) (BORBA, 2002; BORBA et al., 2003).

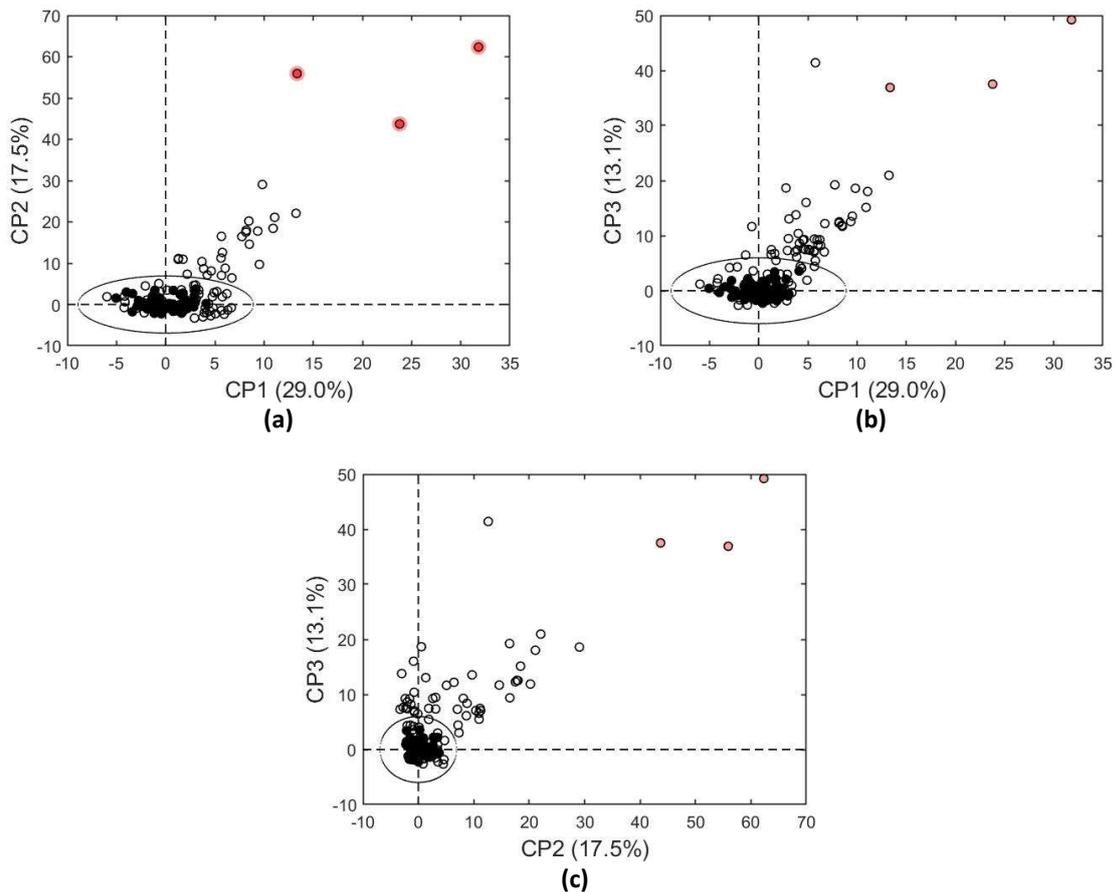


Figura 11: Gráficos de scores: (a) CP₁-CP₂, (b) CP₁-CP₃, e (c) CP₂-CP₃.

Segundo IGAM (2014), as fontes de arsênio na bacia do rio das Velhas se concentram principalmente em seu alto curso, sendo que as ocorrências de arsênio no médio e baixo curso estão associadas a sedimentos, onde parte foi depositada ao longo dos anos. No atlas geoquímico da bacia do rio das Velhas, elaborado por Viglio et al. (2010), também foram relatadas altas concentrações de arsênio em amostras de sedimentos do fundo do rio na área do Quadrilátero Ferrífero.

Em um estudo realizado por Calazans et al. (2018), os autores apontaram a estação BV137 como uma das mais relevantes na calha do rio das Velhas e prioritária para manutenção na rede. Nesse estudo os autores aplicaram análise de cluster para verificar a similaridade entre as estações da calha do rio das Velhas e a estação BV137 não se agrupou a nenhuma outra, apresentando na ACP/AF fatores correlacionados com os parâmetros de qualidade da água característicos de atividades de mineração e matéria orgânica. Resultados que confirmam as análises deste trabalho. Além disso, as fontes de poluição ambiental neste ponto do rio das Velhas corroboram com as caracterizações das Regiões (1 a 3) descritas pelo modelo de ACP, o que é fundamental para a sua validação.

Os resultados demonstraram o uso potencial de ACP para associar condições críticas de qualidade da água a fontes de poluição e como a abordagem local pode ser útil para um estudo detalhado do trecho de um curso de água. A partir das análises realizadas os órgãos gestores e instituições de recursos hídricos podem se concentrar nos parâmetros mais significativos que contribuíram para a poluição do rio.

CONCLUSÕES

Os bancos de dados gerados com os Programas de Monitoramento da Qualidade das Águas (PMQAs) constituem-se numa rica fonte de informação. A desconsideração de um número significativo de parâmetros altamente correlacionados entre si pode comprometer a análise de dados e o alcance de um processo de tomada de decisão mais efetivo. Nesse contexto, a técnica estatística multivariada, denominada Análise por Componentes Principais (PCA), é uma ferramenta útil a ser empregada em sistemas de suporte à decisão ao monitoramento de qualidade de água, e pode ser empregada com diferentes objetivos.

Neste estudo, explorou-se o uso de um modelo baseado em PCA, através de seus *scores*, associado ao filtro de Hampel, para o monitoramento simultâneo de doze parâmetros de qualidade da água de uma estação da calha do rio das Velhas (BV137). Observou-se, ao final, um desempenho satisfatório, com a identificação de *outliers* e a associação de regiões espaciais nos gráficos de *scores* com eventos críticos de poluição em relação à qualidade da água. Essas informações podem servir como sinal de alerta sobre desvios em relação à condição de referência de qualidade de água, e para diagnóstico, ao se estabelecer, *a priori*, associações entre essas regiões espaciais e fontes de poluição. Ademais, a metodologia utilizada pode ser aplicada a outras estações de monitoramento, proporcionando uma avaliação detalhada de cada trecho do curso de água.

REFERÊNCIAS

ANA. Agência Nacional de Águas e Saneamento Básico. **Atlas Esgotos: Despoluição de bacias hidrográficas**. Brasília: ANA, 2017.

BALLABIO, D.. A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. **Chemometrics and Intelligent Laboratory Systems**, v.149, p.1-9, 2015. DOI: <https://doi.org/10.1016/j.chemolab.2015.10.003>

BAPTISTA, M.; PÁDUA, V. L.. **Restauração de Sistemas Fluviais**. Belo Horizonte: Manole, 2016.

BARTRAM, J.; BALLANCE, R.. World health organization & united nations environment programme. **Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programs**. London: E & FN Spon, 1996.

BEHMEL, S.; DAMOUR, M.; LUDWIG, R.; RODRIGUEZ, M. J.. Water quality monitoring strategies - A review and future perspectives. **Science of The Total Environment**, v.571, p.1312-1329, 2016. DOI: <https://doi.org/10.1016/j.scitotenv.2016.06.235>

BENGRÁINE, K.; MARHABA, T. F.. Using principal component analysis to monitor spatial and temporal changes in water quality. **Journal of Hazardous Materials**, v.100, n.1-3, p.179-195, 2003. DOI: [https://doi.org/10.1016/S0304-3894\(03\)00104-3](https://doi.org/10.1016/S0304-3894(03)00104-3)

BODO, B. A.. Robust graphical methods for diagnosing trend in irregularly spaced water quality time series.

Environmental Monitoring and Assessment, v.13, n.2-3, p.407-428, 1989. DOI: https://doi.org/10.1007/978-94-009-1960-0_19

BORBA, R. P.. **Arsênio em Ambiente Superficial: Processos Geoquímicos Naturais e Antropogênicos em uma área de Mineração Aurífera**. Dissertação (Mestrado em Geociências) – Universidade Estadual de Campinas, Campinas, 2002.

BORBA, R. P.; FIGUEIREDO, B. R.; MATSCHULLAT, J.. Geochemical distribution of arsenic in waters, sediments and weathered gold mineralized rocks from Iron Quadrangle, Brazil. **Environmental Geology**, v.44, n.1, p.39-52, 2003. DOI: <https://doi.org/10.1007/s00254-002-0733-6>

BRO, R.; SMILDE, A. K.. Principal component analysis. **Anal. Methods**, v.6, n.9, p.2812–2831, 2014. DOI: <https://doi.org/10.1039/C3AY41907J>

CALAZANS, G. M.; PINTO, C. C.; COSTA, E. P.; PERINI, A. F.; OLIVEIRA, S. C.. Using multivariate techniques as a strategy to guide optimization projects for the surface water quality network monitoring in the Velhas river basin, Brazil. **Environmental Monitoring and Assessment**, v.190, n.12, p.726, 2018. DOI: <https://doi.org/10.1007/s10661-018-7099-z>

CAMARGOS, L. M. M.. **Plano Diretor de Recursos Hídricos da Bacia Hidrográfica do Rio das Velhas: Resumo Executivo**. Dezembro 2004. Belo Horizonte: IGAM/CBH Velhas, 2005.

CHEN, Q.; WU, W.; BLANCKAERT, K.; MA, J.; HUANG, G.. Optimization of water quality monitoring network in a large

river by combining measurements, a numerical model and matter-element analyses. **Journal of Environmental Management**, v.110, p.116-124, 2012. DOI: <https://doi.org/10.1016/j.jenvman.2012.05.024>

CBH VELHAS. **Plano Diretor de Recursos Hídricos da Bacia Hidrográfica do Rio das Velhas 2015**: Resumo Executivo. Belo Horizonte: CBH Velhas, 2015.

COSTA, E. P.; PINTO, C. C.; SOARES, A. L. C.; MELO, L. D. V.; OLIVEIRA, S. M. A. C.. Evaluation of violations in water quality standards in the monitoring network of São Francisco River basin, the third largest in Brazil. **Environmental Monitoring and Assessment**, v.189, n.11, p.590, 2017. DOI: <https://doi.org/10.1007/s10661-017-6266-y>

COTHERN, C. R.; ROSS, N. P.. **Environmental Statistics, Assessment, and Forecasting**. Lewis Publishers, 1993.

COURTEMANCH, D. L.. Bridging the old and new science of biological monitoring. **Journal of the North American Benthological Society**, v.13, n.1, p.117-121, 1994. DOI: <https://doi.org/10.2307/1467272>

CRUZ, M. A. S.; GONÇALVES, A. A.; ARAGÃO, R.; AMORIM, J. R. A.; MOTA, P. V. M.; SRINIVASAN, V. S.; GARCIA, C. A. B.; FIGUEIREDO, E. E.. Spatial and seasonal variability of the water quality characteristics of a river in Northeast Brazil. **Environmental Earth Sciences**, v.78, n.3, p.68, 2019. DOI: <https://doi.org/10.1007/s12665-019-8087-5>

DAVIES, L.; GATHER, U.. The Identification of Multiple Outliers. **Journal of the American Statistical Association**, v.88, n.423, p.782-792, 1993. DOI: <https://doi.org/10.1080/01621459.1993.10476339>

DRESSING, S. A.; MEALS, D. W.; HARCUM, J. B.; SPOONER, J.; STRIBLING, J. B.; RICHARDS, R. P.; MILLARD, C. J.; LANBERG, S. A.; O'DONNELL, J. G.. **Monitoring and Evaluating Nonpoint Source Watershed Projects**. Washington: Tetra Tech (Developed under Contract to U.S. Environmental Protection Agency), 2016.

FOLLADOR, F. A. C.. **Controle estatístico de processo aplicado à qualidade de águas superficiais**. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Paraná, 2010.

GATICA, E. A.; ALMEIDA, C. A.; MALLEA, M. A.; DEL CORIGLIANO, M. C.; GONZÁLEZ, P.. Water quality assessment, by statistical analysis, on rural and urban areas of Chocancharava River (Río Cuarto), Córdoba, Argentina. **Environmental Monitoring and Assessment**, v.184, n.12, p.7257-7274, 2012. DOI: <https://doi.org/10.1007/s10661-011-2495-7>

GE, Z.; SONG, Z.. **Multivariate Statistical Process Control**. London: Springer, 2013.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L.. **Análise Multivariada de dados**. 6 ed. Porto Alegre: Book, 2009.

HELSEL, D. R.; HIRSCH, R. M.. **Statistical methods in water resources. Techniques of Water Resources Investigations**. Book 4, chapter A3. U.S. Geological Survey, 2002.

HOTELLING, H.. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v.24, n.6, p.417-441, 1933. DOI: <https://doi.org/10.1037/h0071325>

IGAM. **Monitoramento da qualidade das águas superficiais na sub-bacia do rio das Velhas em 2009**. Belo Horizonte: IGAM, 2010.

IGAM. **Identificação de municípios com condição crítica para a qualidade de água na bacia do rio das Velhas**. Belo Horizonte: IGAM, 2013.

IGAM. **Avaliação da Qualidade das Águas Superficiais de Minas Gerais em 2017**: Resumo Executivo Anual. Belo Horizonte: IGAM, 2018.

IGAM. **Análise de Tendência da Qualidade das Águas na Bacia do rio das Velhas no período de 2000 a 2013**. Belo Horizonte: IGAM, 2014.

LIN, B.; RECKE, B.; KNUDSEN, J. K. H.; JØRGENSEN, S. B.. A systematic approach for soft sensor development. **Computers & Chemical Engineering**, v.31, n.5-6, p.419-425, 2007. DOI: <https://doi.org/10.1016/j.compchemeng.2006.05.030>

MANLY, B. F. J.. **Métodos Estatísticos Multivariados: uma introdução**. 3 ed. Porto Alegre: Bookman, 2008.

MINAS GERAIS. **Deliberação Normativa n.01 de 05 de maio de 2008**. Dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu enquadramento, bem como estabelece as condições e padrões de lançamento de efluentes, e dá outras providências. Belo Horizonte: DOE, 2008.

MOHAMED, I.; OTHMAN, F.; IBRAHIM, A. I. N.; ALAA-ELDIN, M. E.; YUNUS, R. M.. Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia. **Environmental Monitoring and Assessment**, v.187, n.1, p.1-12, 2015. DOI: <https://doi.org/10.1007/s10661-014-4182-y>

OLSEN, R. L.; CHAPPELL, R. W.; LOFTIS, J. C.. Water quality sample collection, data treatment and results presentation for principal components analysis: literature review and Illinois River watershed case study. **Water Research**, v.46, n.9, p.3110-3122, 2012. DOI: <https://doi.org/10.1016/j.watres.2012.03.028>

PEARSON, K.. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v.2, n.11, p.559-572, 1901. DOI: <https://doi.org/10.1080/14786440109462720>

PHUNG, D.; HUANG, C.; RUTHERFORD, S.; DWIRAHMADI, F.; CHU, C.; WANG, X.; NGUYEN, M.; NGUYEN, N. H.; DO, C. M.; NGUYEN, T. H.; DINH, T. A. D.. Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. **Environmental Monitoring and Assessment**, v.187, n.5, p.229, 2015. DOI: <https://doi.org/10.1007/s10661-015-4474-x>

REIMANN, C.; FILZMOSER, P.; GARRETT, R. G.; DUTTER, R..

Statistical Data Analysis Explained: Applied Environmental Statistics with R. Chichester: John Wiley & Sons, Ltd, 2008.

SABINO, C. V. S.; LAGE, L. V.; ALMEIDA, K. C. B.. Uso de métodos estatísticos robustos na análise ambiental.

Engenharia Sanitária e Ambiental, v.19, p.87-94, 2014. DOI: <http://dx.doi.org/10.1590/S1413-41522014019010000588>

SERGEANT, C. J.; STARKEY, E. N.; BARTZ, K. K.; WILSON, M. H.; MUETER, F. J.. A practitioner's guide for exploring water quality patterns using principal components analysis and Procrustes. **Environmental Monitoring and Assessment**, v.188, n.4, p.249, 2016. DOI: <https://doi.org/10.1007/s10661-016-5253-z>

SHAPIRO, S. S.; WILK, M. B.. An analysis of variance test for normality (complete samples). **Biometrika**, v.52, n.3-4, p.591-611, 1965. DOI: <https://doi.org/10.1093/biomet/52.3-4.591>

SHARMA, S.. **Applied Multivariate Techniques**. New York: John Wiley & Sons, Inc., 1995.

SINGH, K. P.; MALIK, A.; MOHAN, D.; SINHA, S.. 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, n.18, p.3980-3992, 2004. DOI: <https://doi.org/10.1016/j.watres.2004.06.011>

TOLEDO, L. G.; NICOLELLA, G.. Índice de qualidade de água em microbacia sob uso agrícola e urbano. **Scientia Agricola**, v.59, n.1, p.181-186, 2002. DOI: <https://doi.org/10.1590/S0103-90162002000100026>

TRINDADE, A. L. C.; ALMEIDA, K. C. B.; BARBOSA, P. E.; OLIVEIRA, S. M. A. C.. Tendências temporais e espaciais da

qualidade das águas superficiais da sub-bacia do Rio das Velhas, estado de Minas Gerais. **Engenharia Sanitária e Ambiental**, v.22, n.1, p.13-24, 2017. DOI: <https://doi.org/10.1590/s1413-41522016131457>

UNEP; GEMS. **Water Quality for Ecosystem and Human Health Taiwan Review**. 2 ed. Ontario: UNEP/GEMS, 2008.

VEGA, M.; PARDO, R.; BARRADO, E.; DEBÁN, L.. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. **Water Research**, v.32, n.12, p.3581-3592, 1998. DOI: [https://doi.org/10.1016/S0043-1354\(98\)00138-9](https://doi.org/10.1016/S0043-1354(98)00138-9)

VIGLIO, E. P.; CUNHA, F. G.. **Atlas geoquímico da bacia do rio das Velhas**. Belo Horizonte: CPRM, 2010.

VOZA, D.; VUKOVIĆ, M.. The assessment and prediction of temporal variations in surface water quality: a case study. **Environmental Monitoring and Assessment**, v.190, n.7, p.434, 2018. DOI: <https://doi.org/10.1007/s10661-018-6814-0>

ZIMMERMAN, S. M.; DARDEAU, M. R.; CROZIER, G. F.; WAGSTAFF, B.. The second battle of mobile bay - Using SPC to control the quality of water monitoring. **Computers & Industrial Engineering**, v.31, n.1-2, p.257-260, 1996. DOI: [https://doi.org/10.1016/0360-8352\(96\)00125-8](https://doi.org/10.1016/0360-8352(96)00125-8)

ZHANG, Z.; TAO, F.; DU, J.; SHI, P.; YU, D.; MENG, Y.; SUN, Y.. Surface water quality and its control in a river with intensive human impacts: a case study of the Xiangjiang River, China. **Journal of Environmental Management**, v.91, n.12, p.2483-2490, 2010. DOI: <https://doi.org/10.1016/j.jenvman.2010.07.002>

A CBPC – Companhia Brasileira de Produção Científica (CNPJ: 11.221.422/0001-03) detém os direitos materiais desta publicação. Os direitos referem-se à publicação do trabalho em qualquer parte do mundo, incluindo os direitos às renovações, expansões e disseminações da contribuição, bem como outros direitos subsidiários. Todos os trabalhos publicados eletronicamente poderão posteriormente ser publicados em coletâneas impressas sob coordenação da **Sustenere Publishing**, da Companhia Brasileira de Produção Científica e seus parceiros autorizados. Os (as) autores (as) preservam os direitos autorais, mas não têm permissão para a publicação da contribuição em outro meio, impresso ou digital, em português ou em tradução.