

IV-364 - CLASSIFICAÇÃO MULTIVARIADA DO RIO GUALAXO DO NORTE PÓS DESASTRE DE FUNDÃO

Grazielle Rocha dos Santos⁽¹⁾

Doutora em Engenharia Ambiental na área de Recursos Hídricos/ Meio Ambiente pelo PROAMB/UFOP. Engenheira Ambiental pela mesma instituição. Atuando na área de monitoramento de bacias hidrográficas com aplicação de estatística multivariada.

Luisa Cardoso Maia

Engenheira Ambiental pela Universidade Federal de Itajubá (UNIFEI). Mestre e Doutora em Engenharia Ambiental pelo Programa de Pós-Graduação em Engenharia Ambiental (ProAmb) da Universidade Federal de Ouro Preto (UFOP).

Paloma Capato

Graduanda em Biologia pela Universidade Federal de Ouro Preto.

Aníbal da Fonseca Santiago

Professor Adjunto do Departamento de Engenharia Civil da Escola de Minas da Universidade Federal de Ouro Preto (UFOP). Engenheiro Ambiental pela Universidade Federal de Viçosa (2006), Mestre em Engenharia Hidráulica e Saneamento pela Universidade de São Paulo (2008) e Doutor em Engenharia Civil pela Universidade Federal de Viçosa (2013). Tem experiência na área de tratamento simplificado de águas residuárias e produção de biomassa algal.

Gilmare Antônia da Silva

Graduada em Química (Licenciatura) pela Universidade Federal de São João del-Rei (2001), mestrado em Agroquímica pela Universidade Federal de Viçosa (2003) e doutorado em Ciências (área de concentração Química Analítica) pela Universidade Estadual de Campinas (2007). Atualmente é Professora Associada II da Universidade Federal de Viçosa/MG, tendo iniciado a carreira docente na Universidade Federal de Ouro Preto/MG, pelo período de 2008 a 2020. Atua em Química Analítica com ênfase em quimiometria, métodos de separação, preparo de amostras, análises ambientais e de traços e análise de alimentos.

Endereço ⁽¹⁾: Campus Morro do Cruzeiro, Bauxita, Ouro Preto, MG. CEP:35400-000. Brasil. Tel (31) 999386287. Email: grazielle.santos@aluno.ufop.edu.br

RESUMO

O estudo teve como principal objetivo realizar a construção de modelos de classificação multivariada a partir dos dados de monitoramento da qualidade da água da bacia hidrográfica do rio Gualaxo do Norte, Minas Gerais/Brasil, que em 2015 foi atingida pelo rompimento da barragem de Fundão da mineradora Samarco S.A., que liberou no rio uma quantidade considerável de rejeitos. Foram avaliados 27 pontos, abrangendo áreas afetadas e não afetadas, monitorando variáveis químicas, físicas e microbiológicas, no período de julho de 2016 a junho de 2017. Aos dados obtidos foi aplicada técnica multivariada classificatória que objetivou desenvolver um modelo capaz de prever as classes pré-definidas, formadas a partir das variáveis responsáveis pela distinção das mesmas. O modelo de classificação utilizando PLS-DA alcançou a predição de três classes, rio principal não afetado, rio principal afetado e tributários. Foi possível inferir que o modelo desenvolvido permitiu claramente a diferenciação das três classes estabelecidas para os dados avaliados, alcançando médias relacionadas a 90% de acerto.

PALAVRAS-CHAVE: Rio Gualaxo do Norte, Classificação multivariada, qualidade da água, Desastre de Fundão.

INTRODUÇÃO

A água é um recurso finito e indispensável para a manutenção da vida no planeta. Sua existência em boa qualidade é definida como fator essencial para a manutenção da saúde humana e ambiental (Oliveira-Filho e Parron, 2007). Sendo a água essencial, a contaminação dos corpos hídricos devido às atividades produtivas humanas se tornou uma grande preocupação nesta década. A gestão imprópria do recurso, o elevado crescimento demográfico e a expansão industrial acelerada nos últimos anos têm contribuído de forma direta

para a degradação dos rios. Esta, por sua vez, provoca alterações qualitativas e quantitativas na qualidade das águas, evidenciando a necessidade de uma política de gestão eficiente (Menezes et al., 2016).

A Política Nacional de Recursos Hídricos (PNRH) instituída pela lei nº 9.433 de 8 de janeiro de 1997, conhecida como Lei das Águas, define em uma de suas diretrizes gerais que a articulação da gestão deste bem com a do uso do solo é fundamental, reconhecendo a bacia hidrográfica como unidade territorial para sua implementação. Diante disso se destaca a importância de estudos, análises e diagnósticos ambientais que tenham como objetivo apontar a situação atual das bacias hidrográficas do país, bem como a evolução de atividades antrópicas na modificação dos padrões de uso do solo, sendo este um dos principais fatores de influência na qualidade da água de ambientes fluviais.

Dentre a vasta rede hídrica brasileira se encontra a bacia hidrográfica do rio Doce (BHRD). Localizada nos estados de Minas Gerais e Espírito Santo possui um importante papel no desenvolvimento e ocupação da região, uma vez que em seu território estão presentes atividades econômicas como a extração do ouro e ferro, além de intensas atividades de agricultura e pecuária. Assim o sistema de drenagem da BHRD é fundamental para a economia da região, fornecendo água para os mais diversos usos (CBH DOCE, 2021).

Com sua porção territorial contida majoritariamente em Minas Gerais (86%), o dado faz jus ao cognome associado ao estado de "caixa d'água do Brasil". Tal associação se deve a expressiva representatividade na disponibilidade hídrica nacional, 3,5% do total, tendo o rio Doce como um de seus cursos mais importantes. Devido a essa grandiosidade, a BHRD é subdividida em seis Unidades de Planejamento e Gestão dos Recursos Hídricos (UPGRH), visando facilitar a tomada de decisão no gerenciamento deste bem e atentando-se sempre às características naturais, sociais e econômicas de cada localidade (SEMAD, 2012).

O rio Doce se forma na UPGRH DO1, chamada de bacia hidrográfica do rio Piranga (BHRP) e localizada no Alto Rio Doce, região marcada por atividades intensas do setor industrial, agropecuário e minerário. De acordo com o Instituto Mineiro de Gestão das Águas (IGAM), a BHRP tem problemáticas evidentes relacionadas ao uso inadequado dos recursos naturais e somado a isso foi impactada pelo rompimento da barragem de Fundão, de responsabilidade da mineradora Samarco, ocorrido em novembro de 2015.

O barramento se localizava no distrito de Bento Rodrigues, município de Mariana/MG, e seu colapso liberou aproximadamente 34 milhões de metros cúbicos de rejeito de minério de ferro nos corpos hídricos de toda a bacia, que percorreram todo o rio Doce até chegarem ao oceano Atlântico, 16 dias depois do rompimento (CPRM, 2015). O desastre de Fundão foi o maior desastre ambiental do país, elevando ainda mais a preocupação relacionada à qualidade da água que, mesmo após cinco anos, é visivelmente prejudicada pelos rejeitos que ainda se encontram nas encostas dos rios afetados.

O rio Gualaxo do Norte, primeiro rio assim impactado, apresenta uma extensão de cerca de 60 km e, destes, 48 km foram atingidos pelo resíduo oriundo da barragem. O fato de possuir uma parte afetada e não afetada consente a compreensão da magnitude dos impactos ambientais decorrentes do rompimento da barragem. Esse entendimento é alcançado pela possibilidade de se avaliar os distintos compartimentos do mesmo rio e inferir como o aporte de rejeito perturbou a qualidade da água, permitindo ainda inferências sobre possíveis impactos nos demais rios impactados pelo rejeito.

Diante do exposto é notável a relevante demanda de pesquisas na região citada e, quando se trata de estudos ambientais, é conhecido que muitos dados são compilados. Nesse contexto, se destaca o potencial das ferramentas de estatística multivariada, uma vez que permitem a interpretação dessas matrizes extensas, na busca de informações mais concisas, confiáveis e aplicáveis. Isso ocorre devido ao fato de que a análise multivariada realiza a avaliação conjunta das variáveis consideradas, podendo reduzir a complexidade dos dados, mas mantendo a informação original (Sabino, Lage e Almeida, 2014).

Estudos recentes (Chounlamany, Tanchuling and Inoue, 2017; Hajjgholizadeh and Melesse, 2017; Kara et al., 2017; Carvalho et al., 2015; Kovács et al., 2014) demonstram que a quimiometria tem sido cada vez mais aplicável para resolução de problemas ambientais. Inúmeras têm sido as pesquisas que utilizam técnicas multivariadas para os mais diversificados fins e aqui se destaca a possibilidade de construir modelos de classificação baseados em padrões de comportamento (Shafii et al., 2019). Nesse contexto o rio Gualaxo do Norte, no viés de afetado e não afetado pelo desastre minerário se mostra, a princípio, adequado para aplicação de técnicas que busquem avaliar, por meio de modelos matemáticos, o comportamento de amostras em ambas as regiões.

Neste contexto o objetivo desta pesquisa foi aplicação de ferramentas quimiométricas capazes de classificar amostras do rio Gualaxo do Norte quanto a natureza de afetado ou não pelo rompimento da barragem de Fundão. Especificamente para essas considerações foi aplicada a análise hierárquica de cluster (HCA), na etapa exploratória, e a partial least square-discriminant analysis (PLS-DA) para obter resultados que identifiquem semelhanças e dissimilaridades entre os locais amostrados e classifiquem amostras indicando características de afetado e não afetado.

MATERIAIS E MÉTODOS

ÁREA DE ESTUDO

Para tal foram confeccionados mapas de localização dos pontos coletados e mapas do uso do solo da região investigada. A utilização da cartografia para o entendimento da dinâmica de bacias hidrográficas tem apresentado extrema relevância e o destaque se dá pelo fato de admitir uma visão ampliada das atividades, permitindo uma integração de informações, o que possibilita uma melhor tomada de decisão.

Os mapas foram confeccionados utilizando o software ESRI® ArcGis 10.2 utilizando para tal dados vetoriais e a imagem raster obtida no portal Earth Explorer, da United States Geological Survey (USGS); as informações disponíveis no site da Agência Nacional das Águas (ANA), do Departamento Nacional de Produção Mineral (DNPM); e o portal do Departamento Nacional de Infraestrutura e Transporte (DNIT). Os dados referentes às unidades da federação, municípios e distritos foram obtidos a partir das informações disponíveis no portal do Instituto Brasileiro de Geografia e Estatística (IBGE).

DADOS DE QUALIDADE DA ÁGUA

Esta pesquisa se desenvolveu a partir da realização de uma campanha de monitoramento de qualidade da água, que se caracterizou por meio da execução de 12 amostragens de água, efetuadas mensalmente de julho de 2016 a junho de 2017, em 27 pontos de coleta ao longo da bacia (Figura 1). Ressalta-se que os pontos foram divididos de forma que 11 foram localizados na calha do rio principal e 16 em tributários. No tocante à região afetada e à não afetada pelo aporte de lama tem-se que 10 pontos foram fixados a montante da barragem de Fundão, isto é, região não afetada.

A Tabela 1 mostra as variáveis avaliadas em cada uma das duas campanhas. Condutividade, potencial de oxidação-redução (ORP) e pH foram medidos in situ, por meio da utilização de um instrumento multiparâmetro portátil da marca Myron L. Company, Ultrameter (USA) II, 6P. A porcentagem de saturação do oxigênio dissolvido e a temperatura foram obtidas in situ, utilizando uma sonda HACH40 D HQ com eletrodo LDO101 (Hach™, Alemanha). As demais análises foram realizadas seguindo as recomendações do Standard Methods for the Examination of Water and Wastewater da American Public Health Association (APHA, 2012).

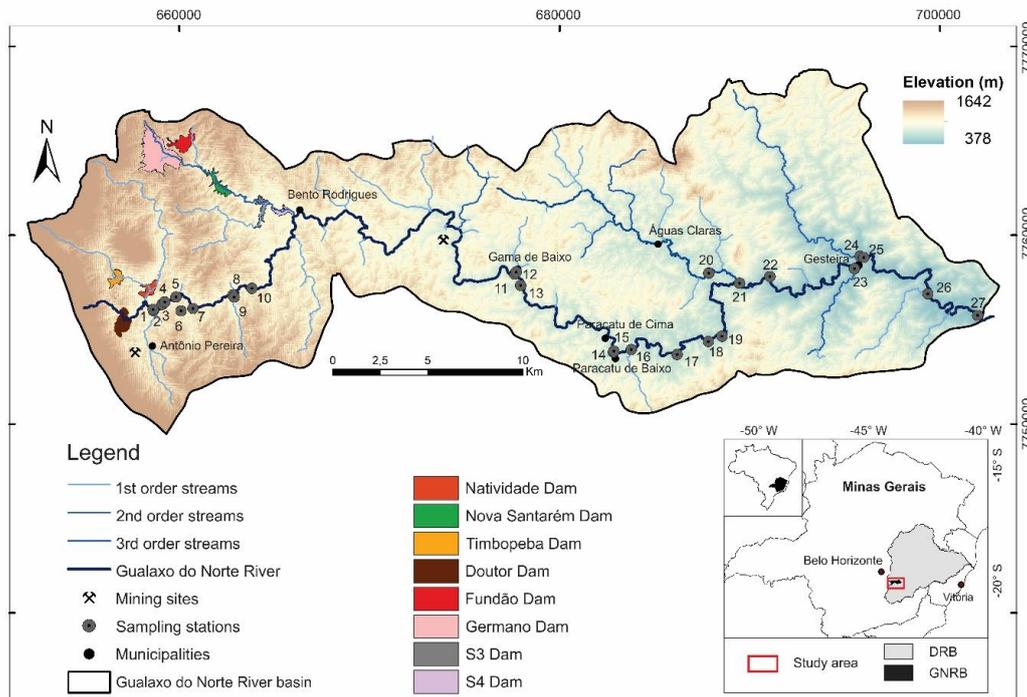


Figura 1: Localização dos pontos amostrais na bacia hidrográfica do rio Gualaxo do Norte, Minas Gerais.

Tabela 1 – Métodos analíticos utilizados para as medidas realizadas em laboratório.

PARÂMETROS	MÉTODOS
Alcalinidade	2320 B (APHA, 2012)
Coliformes totais e <i>Escherichia coli</i>	Colilert 9223 B (APHA, 2012)
Condutividade (in situ)	Multiparâmetro
Cloreto	Método titulométrico
Demanda bioquímica de oxigênio	5210 B (APHA, 2012)
Fósforo total	4500 B (APHA, 2012)
Nitrato	HACH 8171 (APHA, 2012)
Nítrito	HACH 8173 (APHA, 2012)
ORP (in situ)	Multiparâmetro
pH (in situ)	Multiparâmetro
% de saturação oxigênio dissolvido	Sonda HACH 40 D HQ
Sólidos totais	2540 B (APHA, 2012)
Temperatura (in situ)	Sonda HACH 40 D HQ
Turbidez	Turbidímetro HACH 2100

TRATAMENTO QUIMIOMÉTRICO DOS DADOS

Para identificar os possíveis padrões existentes no conjunto foi aplicada a técnica exploratória multivariada análise hierárquica de cluster (HCA). Para tal função a HCA utiliza toda a variância contida na matriz avaliada com o intuito de encontrar similaridades entre as amostras, se baseando também nas dissimilaridades entre as mesmas, formando os chamados clusters. Neste trabalho foi utilizado o método Ward, selecionado de forma a alcançar clusters mais significativos, utilizando o quadrado das distâncias euclidianas como fator de semelhança. O output da técnica é um dendograma no qual cada observação começa em um grupo separado e une-se aos outros à medida que a distância de ligação cresce, até restar apenas um grupo (Barakat et al., 2016; Gazzaz et al., 2012; Novák et al., 2017; Saha and Rahman, 2018; Vega et al., 1998).

Após a definição das classes inicia-se a etapa desenvolvimento de um modelo de classificação, para o qual existe uma ampla gama de algoritmos capazes de estruturar regras de classificação. Nesta pesquisa foi aplicada a técnica PLS-DA, comumente utilizada para construção de modelos matemáticos capazes de alocar uma amostra de acordo com a sua classe pré-estabelecida (Jímenez-Carvelo et al., 2021). Para o desenvolvimento do modelo o banco de dados, com 327 observações, foi dividido de forma que 78% foram utilizados na calibração e 22% para a predição das amostras desconhecidas ao conjunto de treinamento.

O método PLS-DA cria um modelo matemático capaz de alocar amostras de acordo com a classe informada. Aqui se esclarece que a técnica não assimila classes 1, 2 ou 3, e sim valores binários, isto é, 0 ou 1; assim sendo, foram necessárias três colunas para representar as classes 1, 2 e 3 na matriz, respectivamente. A técnica exige um pré-processamento dos dados quando inseridos no software, e nesse caso foi utilizado o auto-escalamento, que faz com que as diferentes magnitudes das variáveis não interfiram de forma tendenciosa nos resultados finais.

A escolha das variáveis latentes (VL) ocorreu selecionando a que possui menor valor de RMSECV (Root Mean Square Error of Cross Validation) e, nesse caso, foram selecionadas as quatro primeiras VL. Ainda em relação a matriz, ressalta-se que para a aplicação da técnica é necessário a inserção de dois conjuntos de dados, o primeiro com as amostras e as variáveis, denominado X, e o segundo, com os vetores representando as classes, designado Y. O desempenho da PLS-DA foi avaliado por meio da tabela de contingência e, conseqüentemente, pelas figuras de mérito, nesse caso a sensibilidade (classificação das amostras verdadeiramente positivas) e a especificidade (classificação das amostras verdadeiramente negativas). As análises multivariadas foram desenvolvidas utilizando os softwares MatLab 9.1 (The MathWorks, Co., USA), PLS Toolbox 8.2 (Eigenvector Research, Inc., USA) e Statistic Free, versão 10.

RESULTADOS

ÁREA DE ESTUDO

A Figura 2 apresenta o mapa de uso do solo da bacia hidrográfica do rio Gualaxo do Norte.

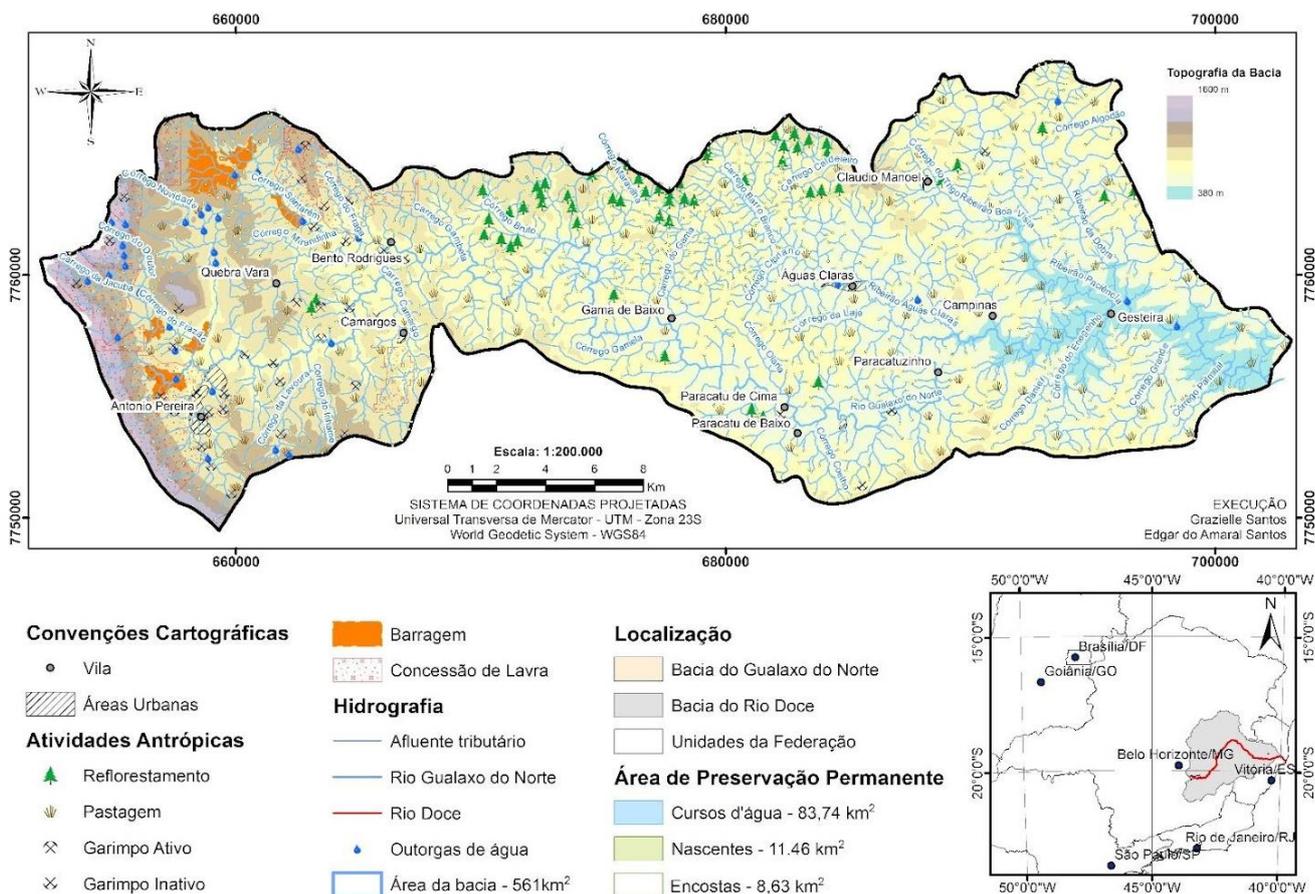


Figura 2: Mapa de uso do solo da bacia hidrográfica do rio Gualaxo do Norte

TRATAMENTO QUIMIOMÉTRICO ANÁLISE HIERARQUICA DE CLUSTER

A Figura 3 explicita o dendograma que apresenta o comportamento dos 27 pontos monitorados nas 12 campanhas, identificando dois grupos principais (1 e 2) e seis subgrupos (3, 4, 5, 6, 7 e 8).

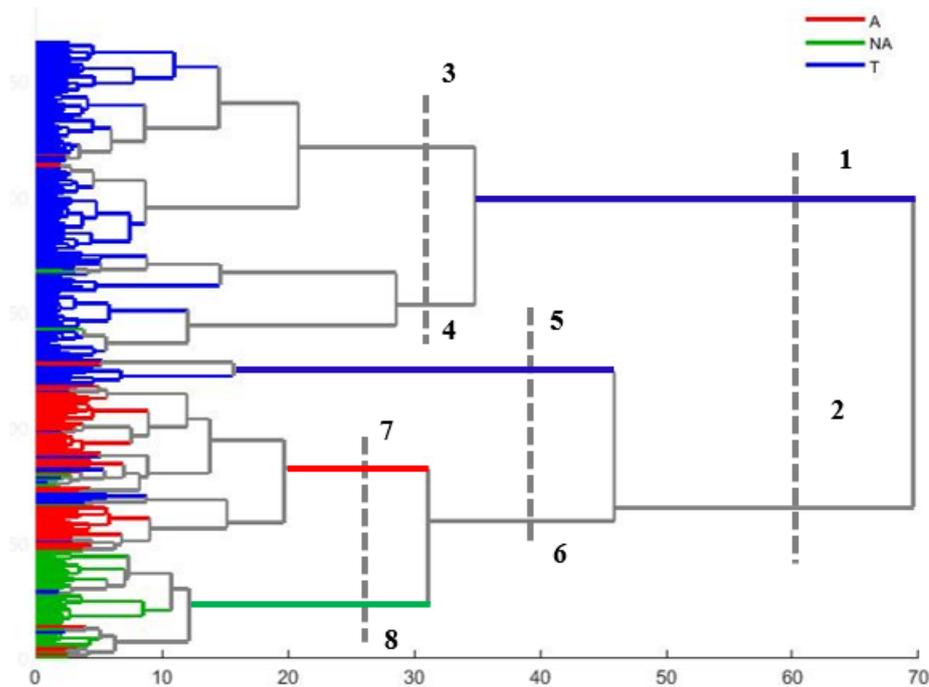


Figura 3: Dendograma obtido a partir da aplicação da análise hierárquica de cluster aplicada aos dados de monitoramento do rio Gualaxo do Norte, Minas Gerais.

Legenda: A: rio principal afetado, NA: rio principal não afetado e T: tributário.

ANÁLISE DISCRIMINANTE POR QUADRADOS MÍNIMOS PARCIAIS (PLS-DA)

De acordo com os padrões observados na avaliação exploratória foram estabelecidas três classes para discriminar os 27 pontos de coleta no rio Gualaxo do Norte: classe 1 - rio principal não afetado, classe 2 - rio principal afetado e classe 3 - tributários. Baseando-se nessa premissa foi aplicada a PLS-DA aos dados com o intuito de construir um modelo de classificação capaz de discriminar os pontos, de acordo com as classes pré-estabelecidas.

Para a construção dos modelos de classificação foram selecionadas quatro variáveis latentes que para a matriz X explicaram 46% dos dados e para o vetor y cerca de 50%. O gráfico biplot (Figura 4) mostra que as duas primeiras variáveis latentes explicaram juntas cerca de 30% da informação dos dados. A disposição das classes permite inferir sobre as variáveis responsáveis pela distinção de cada uma. Assim, a classe 1 (rio principal não afetado) foi influenciada pela alcalinidade e condutividade; a classe 2 (rio principal afetado) foi induzida pelo fósforo total, sólidos totais e turbidez; e a classe 3 (tributários), por sua vez, foi influenciada pelas variáveis cloreto, DBO, *Escherichia coli*, nitrato e pH.

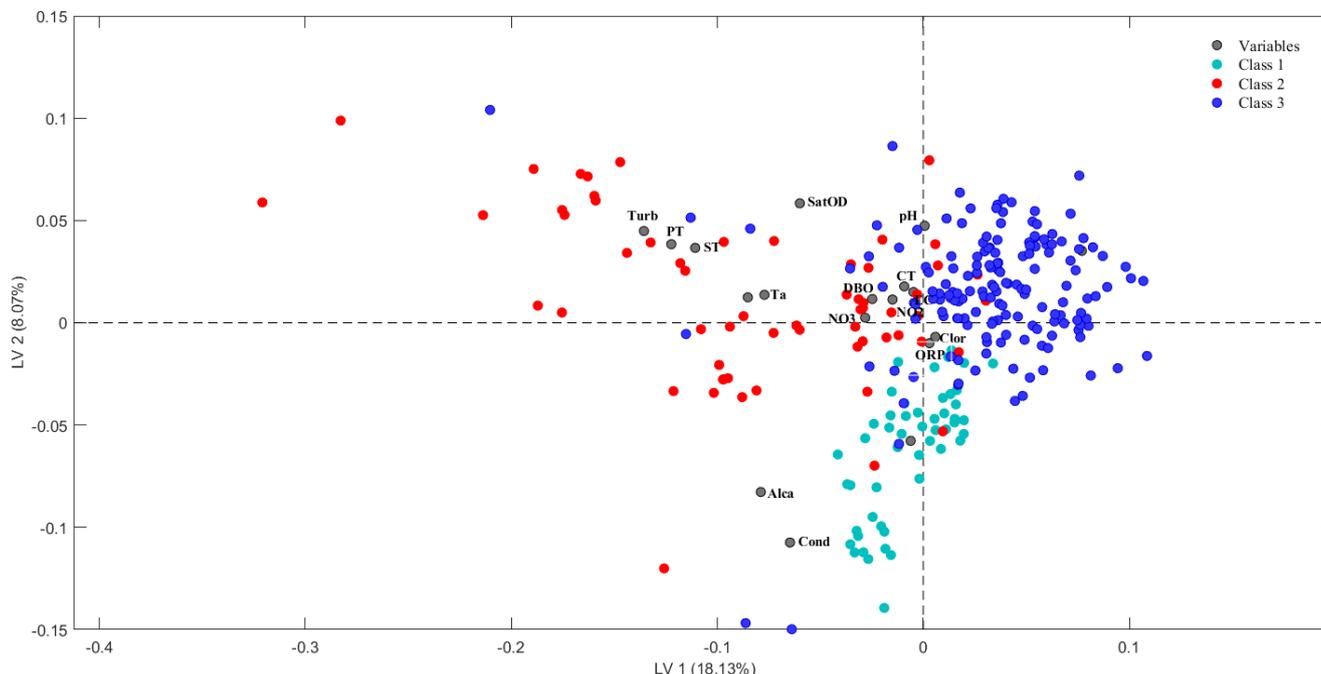


Figura 4 – Identificação das classes existentes no conjunto de dados avaliados por meio da aplicação da PLS-DA.

Legenda: Class 1: rio principal não afetado, Class 2: rio principal afetado e Class 3: tributário. Alca: alcalinidade (mg/L), Cond: condutividade ($\mu\text{S}/\text{cm}$), Clor: cloreto (mg/L), CT: coliformes totais (NMP/100 mL), DBO: demanda bioquímica de oxigênio (mg/L), EC: *E. coli* (NMP/100 mL), %sat: porcentagem de saturação do oxigênio dissolvido (%), NO₃: nitrato (mg/L), NO₂: nitrito (mg/L), ORP: potencial de oxidação-redução (mV), pH: potencial hidrogeniônico, PT: fósforo total (mg/L), ST: sólidos totais (mg/L), Ta: temperatura da água ($^{\circ}\text{C}$), Turb: turbidez (NTU).

A exatidão do modelo discriminatório foi avaliada por duas figuras de mérito: a sensibilidade e a especificidade. A Tabela 2 mostra a porcentagem de classificações corretas para os conjuntos de construção do modelo, o que inclui a validação, e para o conjunto de previsão.

Tabela 1: Precisão do modelo da análise discriminante por mínimos quadrados parciais para as três classes estabelecidas.

Classificação	TREINAMENTO		VALIDAÇÃO		PREVISÃO	
	SEN (%)	SPC (%)	SEN (%)	SPC (%)	SEN (%)	SPC (%)
Classe 1	92	91	92	90	90	97
Classe 2	83	97	86	95	83	100
Classe 3	93	88	95	90	96	86

Legenda: SEN: sensibilidade e SPC: especificidade

DISCUSSÃO

O rio Gualaxo do Norte (Figura 2) possui área de aproximadamente 254 km² e está localizado no estado de Minas Gerais, especificamente entre o Quadrilátero Ferrífero e a Zona da Mata mineira, sendo um afluente do rio Doce. Perpassa três municípios, Ouro Preto, Mariana e Barra Longa, abrangendo respectivamente

nascente, médio curso e foz. Segundo a classificação Köpen apresenta clima de verões quentes e invernos secos, característica das zonas de Cwa; a pluviosidade média anual pode variar de 1200 a 1400 mm (INMET, 2016). As regiões do alto e médio Gualaxo são marcadas por atividades de exploração mineral, enquanto a porção baixa se caracteriza pela presença de agropecuária, plantações de eucalipto, com vistas à produção do carvão siderúrgico, e predominância da pecuária leiteira extensiva.

A principal estrutura geológica da região é o Quadrilátero Ferrífero e aqui se ressalta o intemperismo das rochas localizadas no alto Gualaxo como responsável pela elevação de íons bicarbonato na água, o que consequentemente altera os valores de alcalinidade e condutividade encontrados na bacia. Essa unidade geológica representa a principal fonte econômica de Ouro Preto e Mariana que é a exploração mineral; especificamente na referida bacia coexistem as minas de exploração de minério de ferro de Timbopeba e Alegria e o complexo minerário de Germano, pertencentes à Vale e à Samarco Mineração S.A., respectivamente.

A barragem de Fundão representava uma importante estrutura do complexo de Germano. O sistema de contenção de rejeitos de Fundão foi inaugurado em 2008, projetado para conter um volume de 79,6 milhões de m³ de rejeitos finos e 32 milhões de m³ de rejeitos arenosos, tendo como previsão uma vida útil de 25 anos (SUPRAM, 2008). Em novembro de 2015, após oito anos de operação, ocorreu o colapso da barragem, quando o sistema continha em torno de 56 milhões de m³. Deste total, 76% foram extravasados atingindo o rio Gualaxo do Norte, destruindo 80% das edificações do distrito de Bento Rodrigues, causando a morte de 19 pessoas, desabrigando mais de 600 famílias, além de causar danos, de ordem ainda não calculada, em diversos corpos d'água da bacia hidrográfica do rio Doce e a todos os ecossistemas associados (Fernandes et al., 2016; Sánchez et al., 2018).

ANÁLISE HIERÁRQUICA DE CLUSTER

O cluster 1 (Figura 3) se caracteriza por pontos coletados nos tributários enquanto o cluster 2, majoritariamente, é constituído por sítios localizados no rio principal (afetado e não afetado). Os clusters 3, 4 e 5 são formados pelas amostras monitoradas nos tributários, o que a nível de bacia hidrográfica está sob uma lógica coerente. Cada afluente está relacionado a diferentes tipos de atividades de uso do solo, consequentemente, a distintas formas de poluição, assim como diferentes características como volume de água e largura do rio, o que explica a diferença dos clusters. O cluster 6, por sua vez, são os pontos monitorados no rio principal e, como já mencionado, o rio Gualaxo do Norte pôde ser dividido em dois compartimentos distintos, um impactado pelo aporte de lama (cluster 7) e outro não impactado (cluster 8), justificando a separação dos grupos.

Utilizando a HCA para avaliar a qualidade da água do rio Doce Passos et al. (2021), encontraram cinco clusters, sendo os três principais relacionados a sazonalidade, isto é, nítida separação entre campanhas do período seco e chuvoso e os subgrupos relacionados às características qualitativas dos pontos avaliados. Soares et al. (2021) também buscaram avaliar as similaridades de afluentes que abastecem um reservatório, localizado no estado de Minas Gerais (Brasil) aplicando HCA. Os autores encontraram quatro clusters e relataram a estreita relação desse comportamento com os níveis de poluição, constatada na avaliação dos parâmetros de qualidade mensurados.

O uso da HCA para avaliações de qualidade da água tem apresentado excelentes resultados como reportado por inúmeros trabalhos (Passos et al., 2021; Soares et al., 2021; Subba Rao and Chaudhary, 2019; Bilgin and Konanç, 2016; Muangthong and Shrestha, 2015; Wang et al., 2013; Varol et al., 2012). Nesta pesquisa, a técnica se mostrou útil na identificação dos grupos com características similares, assim como na identificação de comportamentos coerentes com o compartimento ambiental avaliado. Atrelado a isso, tem-se ainda a potencialidade de utilizar essa ferramenta para identificar padrões que podem ser utilizados na construção de modelos de classificação multivariada. Para este trabalho, baseado nos clusters específicos identificados (Figura 3), foram notados três grandes padrões de comportamento das amostras relacionados aos grupos (1) dos tributários, (2) rio principal afetado e (3) rio principal não afetado.

Diversas são as técnicas de classificação multivariada e, em suma, elas utilizam informações prévias dos dados para construir modelos de classificação para grupos específicos. Além de alocarem os dados nas chamadas classes, é possível ainda encontrar a classe de novas amostras. Dessa forma utilizou-se as técnicas análise

discriminante por quadrados mínimos parciais (PLS-DA) para o desenvolvimento de um modelo capaz de alocar as amostras em classes específicas de acordo com suas respectivas características qualitativas.

ANÁLISE DISCRIMINANTE POR QUADRADOS MÍNIMOS PARCIAIS (PLS-DA)

O PLS-DA é um método muito utilizado para o reconhecimento supervisionado de padrões, aplicado com o intuito de criar modelos discriminatórios a partir de dados com altas dimensões (Yang et al., 2021). Métodos de classificação supervisionada são empregados quando existem informações prévias sobre o comportamento do sistema avaliado, isto é, é de conhecimento o padrão que será utilizado para a alocação das amostras (Santana et al., 2020).

Dessa forma, técnicas de classificação multivariada estabelecem uma relação matemática entre uma matriz de dados descritivos e uma variável qualitativa, isto é, os padrões observados. No PLS-DA a variância original dos dados é descrita pelas chamadas variáveis latentes, que advêm de uma combinação linear das variáveis originais, permitindo a interpretação dos dados por meio de gráficos, que conseqüentemente descrevem os diferentes padrões modelados (Ballabio e Consonni, 2013).

De forma geral, ao se avaliar as variáveis relacionadas às classes (Figura 4), é possível associar as características da região, como litologia, falta de estruturas de saneamento básico e o desastre com a barragem de Fundão. Nesse sentido, os valores de condutividade e alcalinidade para a região não afetada se relacionam com as rochas do Quadrilátero Ferrífero, na medida em que o processo de intemperismo libera no ambiente fluvial íons bicarbonato (Medeiros et al., 2020). Para os pontos que representam a área impactada, os parâmetros relacionados à presença de particulados no rio são totalmente coerentes com o aporte de rejeito ao qual a região foi submetida. A classe que representa os tributários foi descrita por variáveis relacionadas ao descarte incorreto de efluentes domésticos e, segundo Fraga et al. (2021) uma das principais fontes de poluição na bacia hidrográfica do rio Doce, onde a bacia hidrográfica do rio Gualaxo do Norte está inserida, é o lançamento de água residuárias sem tratamento adequado.

A exatidão do modelo discriminatório foi avaliada por duas figuras de mérito: a sensibilidade e a especificidade. A Tabela 2 mostra a porcentagem de classificações corretas para os conjuntos de construção do modelo, o que inclui a validação, e para o conjunto de previsão. Dos valores mostrados na Tabela 2 é possível inferir que o modelo desenvolvido permitiu claramente a diferenciação das três classes estabelecidas para os dados avaliados. Para as amostras referentes à classe 1, tanto para o treinamento, validação e previsão, os acertos se apresentaram acima de 90%.

Em relação à classe 2 (Tabela 2), ressaltam-se os acertos com porcentagens elevadas para a especificidade; destaque para o valor de 100% na previsão, fato de suma importância, pois trata sobre a inexistência de amostras que não foram afetadas pelo desastre serem classificadas como afetadas. Nos valores referentes aos acertos da classe 3 (Tabela 2), a sensibilidade apresentou melhores resultados que a especificidade nas três etapas do modelo.

Os valores obtidos na Tabela 2 evidenciam a satisfatória capacidade do modelo desenvolvido com o PLS-DA. O desempenho na classificação das amostras de previsão, principalmente para as da classe 2 (impactadas), demonstra a potencialidade do modelo para avaliações contínuas, uma vez que um dos principais questionamentos em relação ao impacto gerado pelo aporte de rejeitos é quando os compartimentos ambientais retornariam às características não relacionadas diretamente ao desastre.

CONCLUSÕES

O monitoramento da qualidade da água em diferentes regiões de uma bacia hidrográfica é uma importante ferramenta para a compreensão dos fatores naturais e antrópicos responsáveis pela degradação deste recurso e auxilia na gestão e tomada de decisão que envolvem este bem. A presente pesquisa permitiu aferir de forma concisa que as atividades que se desenvolvem ao entorno de um ambiente fluvial, bem como as estações sazonais, afetam diretamente a variação dos parâmetros avaliados.

Nesse contexto, a estatística multivariada auxiliou na obtenção de respostas coerentes e estatisticamente consistentes, mesmo em um sistema dinâmico e impactado como o investigado. O emprego da HCA como uma ferramenta da análise exploratória foi de relevância notável, uma vez que indicou os possíveis padrões existentes nos dados, principalmente no tocante às características peculiares das regiões envolvidas no estudo, áreas afetada e não afetada pelo desastre com a barragem de Fundão. O método de reconhecimento de padrões supervisionado PLS-DA foi importante na confirmação do comportamento dos dados explicitados pela aplicação da HCA, além de apontarem as variáveis de maior poder discriminatório entre as classes.

REFERÊNCIAS BIBLIOGRÁFICAS

1. APHA, 2012 APHA. Standard Methods for examination of water and wastewater. Washington: American Water Work Association. Water Environmental Federation. 2012
2. Barakat, A., Baghdadi, M. E., Rais, J., Aghezzaf, B., Slassi, M., 2016. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. **Int. Soil Water Conserv. Res.**, 4, 284-292. doi.org/10.1016/j.iswcr.2016.11.002.
3. Bilgin, A., Konanç, M.U., 2016. Evaluation of surface water quality and heavy metal pollution of Coruh River Basin (Turkey) by multivariate statistical methods. **Environ.**, 75, 1-18. doi.org/10.1007/s12665-016-5821-0.
4. Carvalho, N. B., Minim V. P. R., Nascimento, M., Vidigal, M. C. T. R., Ferreira, M. A. M., Gonçalves, A. C. A., Minim, L. A., 2015. A discriminant function for validation of the cluster analysis and behavioral prediction of the coffee market. **Food Res. Int.**, 77, 400-407. doi.org/10.1016/j.foodres.2015.10.013.
5. CBH - DOCE - Comitê da bacia hidrográfica do rio Doce (2008). Disponível em: <http://www.cbhdoce.org.br/institucional/a-bacia.>>. Acesso em: 12 de fevereiro. 2021.
6. Chounlamany, V., Tanchuling, M. A., Inoue, T., 2017. Spatial and temporal variation of water quality of a segment of Marikina River using multivariate statistical methods. **Water Sci. Technol.**, 76, 1510-1522. doi.org/10.2166/wst.2017.279.
7. CPRM – Companhia de Pesquisa de Recursos Minerais. Monitoramento Especial da Bacia do Rio Doce. Relatório I e II. In: CPRM, 2015.
8. Fernandes, G. W., Goulart, F. F., Ranieri, B. D., Coelho, M. S., Boesche, K. D. N., Bustamante, M., Carvalho, F. A., Carvalho, D. C., Dirzo, R., Fernandes, S., Galetti Jr, P. M., Millan, V. E. G., Mielke, C., Ramirez, J. L., Neves, A., Rogass, C., Ribeiro, S. P. R., Filho, B. S., 2016. Deep into the mud: ecological and socio-economic impacts of the dam breach in Mariana, Brazil. **Nat. Conserv.**, 2, 35 – 45. doi.org/10.1016/j.ncon.2016.10.003
9. Gazzaz, N. M., Yusoff, M. K., Ramli, F., Aris, A. Z., Juahir, H., 2012. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. **Mar. Pollut.**, 64, 688 – 698. doi.org/10.1016/j.marpolbul.2012.01.032

10. Hajjgholizadeh, M., Melesse, A. M., 2017. Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. **Catena**, 151, 247-258. doi.org/10.1016/j.catena.2016.12.018.
11. INMET, Instituto Nacional de Meteorologia. Ministério da Agricultura, pecuária e abastecimento. Acesso em 10 de agosto de 2021, 2016.
12. Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., Cuadros-Rodríguez, L., 2019. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review. **Food Res. Int.**, 122, 25-39. doi.org/10.1016/j.foodres.2019.03.063.
13. Kara, G. T., Kara, M., Bayram, A., Gündüz, O., 2017. Assessment of seasonal and spatial variations of physicochemical parameters and trace elements along a heavily polluted effluent-dominated stream. **Environ. Monit. Assess.**, 189, 585-598. doi.org/10.1007/s10661-017-6309-4.
14. Kovács, J., Kovács, S., Magyar, N., Tanos P., Hatvani, I. G., Anda, A., 2014. Classification into homogeneous groups using combined cluster and discriminant analysis. **Environ. Model Softw.**, 57, 52-59. doi.org/10.1016/j.envsoft.2014.01.010.
15. Menezes *et al.*, 2016 MENEZES, João Paulo Cunha *et al.* Relação entre padrões de uso e ocupação do solo e qualidade da água em uma bacia hidrográfica urbana. *Eng. Sanit. Ambient.*, Rio de Janeiro, v. 21, n. 3, p. 519-534, Sept. 2016.
16. Muangthong, S., Shrestha, S., 2015. Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. **Environ. Monit. Assess.**, 187, 1-12. doi.org/10.1007/s10661-015-4774-1.
17. Novák, M., Palya, D., Bodai, Z., Nyiri, Z., Magyar, N., Kovács, J., Eke, Z., 2017. Combined cluster and discriminant analysis: An efficient chemometric approach in diesel fuel characterization. **Forensic Sci. Int.**, 270, 61-69. doi.org/10.1016/j.forsciint.2016.11.025.
18. Novák, M., Palya, D., Bodai, Z., Nyiri, Z., Magyar, N., Kovács, J., Eke, Z., 2017. Combined cluster and discriminant analysis: An efficient chemometric approach in diesel fuel characterization. **Forensic Sci. Int.**, 270, 61-69. doi.org/10.1016/j.forsciint.2016.11.025.
19. OLIVEIRA-FILHO, EC. e PARRON, LM. Avaliação de qualidade das águas no Brasil: o Rio Preto examinado. *Espaço & ambiente; Geografia*, vol. 10, p. 279-294, 2007.
20. Passos, J. B. C., Teixeira, D. B. D., Campos, J. A., Lima, R. P. C., Fernandes-Filho, I. E., Silva, D. D., 2021. Multivariate statistics for spatial and seasonal quality assessment of water in the Doce River basin, Southeastern Brazil. *Environ. Monit. Assess.*, 193, 1 - 16. doi.org/10.1007/s10661-021-08918-1.
21. Saha, N., Rahman, M. S., 2018. Multivariate statistical analysis of metal contamination in surface water around Dhaka export processing industrial zone, Bangladesh. **Environ. Nanotechnol. Monit. Manag.**, 10, 206-211. doi.org/10.1016/j.enmm.2018.07.007.
22. Sánchez, L. E., Alger, K., Alonso, L., Barbosa, F. A. R., Brito, M. C. W., LAUREANO, F. V., May, P., Roeser, H., Kakabadse, Y., 2018. Os impactos do rompimento da Barragem de Fundação O caminho para uma mitigação sustentável e resiliente. doi.org/10.2305/IUCN.CH.2018.18.p
23. Santana, F. B., Souza, A. M., Almeida, M. R., Breikreitz, M. C., Filgueiras, P. R., Sena, M. M., Poppi, R. J., 2020. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte V. **Quim. Nova**, 43, 371-381. doi.org/10.21577/0100-4042.20170480.

24. Santos, O. S. H., Avellar, F. C., Alves, M., Trindade, R. C., Menezes, M. B., Ferreira, M. C., França, G. J., Cordeiro, J., Sobreira, F. G., Yoshida, I. M., Moura, P. M., Baptista, M. B., Scotti, M. R., 2019. Understanding the environmental of a mine dam rupture in Brazil: prospects for remediation. **J. Environ.**, 48, 439–449. doi.org/10.2134/jeq2018.04.0168.
25. SEMAD - Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável, 2012. A importância de Minas no cenário nacional das águas. Disponível em: Secretaria de Estado de Meio-Ambiente e Desenv. Sustentável - SEMAD - Secretaria de Meio Ambiente lembra a importância de Minas no cenário nacional das águas. Acessado em 15 de janeiro de 2021.
26. Shafii, N. Z., Saudi, A. S. M., Pang, J. C., Abu, I. F., Sapawe, N., Kamarudin, M. K. A., Saudi, H. F. M., 2019. Application of chemometrics techniques to solve environmental issues in Malaysia. **Heliyon**, 5, 25-34. doi.org/10.1016/j.heliyon.2019.e02534.
27. Subba, R. N., Chaudhary, M., 2019. Hydrogeochemical processes regulating the spatial distribution of groundwater contamination, using pollution index of groundwater (PIG) and hierarchical cluster analysis (HCA): A case study. **Groundw for Sustain Dev.**, 9. doi.org/10.1016/j.gsd.2019.100238.
28. Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. **Water Res.**, 32, 3581-3592. doi.org/10.1016/S0043-1354(98)00138-9.
29. Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., Mustavich, L. F., Li, B., 2013. Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. **J Hydro-Environ Res.**, 7, 30 - 40. doi.org/10.1016/j.jher.2012.10.003.
30. Yang, J., Yin, C., Miao, X., Meng, X., Liu, Z., Hu, L., 2021. Rapid discrimination of adulteration in Radix Astragali combining diffuse reflectance mid-infrared Fourier transform spectroscopy with chemometrics. **Spectrochim Acta A Mol. Biomol.**, 248, 1 – 16. doi.org/10.1016/j.saa.2020.119251.