

## **IV-136 - APLICAÇÃO DE TÉCNICA MULTIVARIADA CLASSIFICATÓRIAS NA AVALIAÇÃO DE CORPO HÍDRICO IMPACTADO PELO DESASTRE DE FUNDÃO**

### **Grazielle Rocha dos Santos<sup>(1)</sup>**

Doutoranda da Universidade Federal de Ouro Preto na área de Recursos Hídricos/Meio Ambiente pelo Programa de pós-graduação em Engenharia Ambiental. Engenheira pela mesma instituição. Atua na área de monitoramento de bacias hidrográficas com aplicação de estatística multivariada.

### **Júlia Caroline Peixoto**

Graduada em Química Industrial 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 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), mestre em Agroquímica pela Universidade Federal de Viçosa (2003) e doutora em Ciências (área de concentração Química Analítica) pela Universidade Estadual de Campinas (2007). Atualmente é Professora Associada II da Universidade Federal de Ouro Preto. Atua em Química Analítica com ênfase em quimiometria, métodos de separação, preparo de amostras, análises ambientais e de traços e análise de alimentos.

**Endereço<sup>(1)</sup>:** Campus Morro do Cruzeiro, Bauxita, Ouro Preto, MG. CEP: 35400-000. Brasil. Tel.: (31) 999386287. E-mail: grazielle.santos@aluno.ufop.edu.br

## **RESUMO**

O presente estudo teve como objetivo aplicar uma técnica classificatória de estatística multivariada para avaliar a qualidade da água do rio Gualaxo do Norte - MG, afetado pelo desastre do rompimento da barragem de Fundão, ocorrido em novembro de 2015. A área de estudo desse trabalho está inserida no Quadrilátero Ferrífero, importante região de mineração de ferro. Assim sendo essa pesquisa se desenvolveu por meio de amostragens em 27 pontos localizados ao longo da bacia hidrográfica do rio Gualaxo do Norte, abrangendo regiões afetadas e não afetadas pelo desastre, monitorando variáveis químicas, físicas e microbiológicas de qualidade da água, no período de julho de 2016 a junho de 2017. Os dados obtidos foram tratados com uma técnica de análise multivariada com o intuito de obter as informações mais latentes, uma vez que dados ambientais usualmente geram extensas e complexas matrizes para a compreensão univariada. Aplicou-se a análise discriminante para o desenvolvimento dos modelos, isto é, o reconhecimento de padrões supervisionado, que objetivou desenvolver modelos que fossem capazes de prever classes pré-definidas, baseado no desenvolvimento de funções discriminantes, formadas a partir das variáveis responsáveis pela distinção das classes. Dessa forma dois modelos foram construídos na tentativa de identificar os parâmetros responsáveis pela distinção dos pontos em relação aos locais impactados e não impactados pelo aporte de lama, com o intuito de alcançar uma predição fidedigna com o monitoramento realizado.

**PALAVRAS-CHAVE:** Rio Gualaxo do Norte, qualidade da água, classificação multivariada, desastre de Fundão.

## **INTRODUÇÃO**

Os desafios do mundo atual são diversos, tais como as variações relacionadas às mudanças climáticas, o aumento populacional, o aumento das diferenças econômicas e as dificuldades em relação à disponibilidade de alimentos, energia e água. Em relação à água, definir um padrão geral de qualidade é ainda mais desafiador, devido principalmente aos seus múltiplos usos (OBADE e MOORE, 2018).

A qualidade da água dos rios está atrelada a fatores naturais e antropogênicos. Para o primeiro grupo algumas características como topografia, variação da temperatura, índices pluviométricos e propriedades litológicas são quesitos de relevância. Fatores antropogênicos como diferentes atividades industriais, intensa urbanização e expansão de superfícies impermeáveis, e utilização de infraestruturas como barragens e açudes são atividades que influenciam a qualidade da água de bacias hidrográficas (CABRIA, BARQUÍN e PEÑAS, 2016).

A estimativa da qualidade da água de uma bacia hidrográfica é um dos meios mais relevantes para avaliar em qual nível as atividades que se desenvolvem em seu entorno geram impactos quantitativos e qualitativos nos seus recursos. Tal avaliação deve ser intensificada quando as consequências das atividades antropogênicas ultrapassam o previsto, como aconteceu na bacia hidrográfica do rio Doce (BHRD), com o desastre da barragem de Fundão. A bacia hidrográfica do rio Doce é uma das principais bacias presentes em Minas Gerais (MG) e sua importância se acentua devido ao papel econômico que a mesma representa para o estado, uma vez que se trata de uma bacia substancialmente voltada para a extração mineral.

A importância da BHRD é acentuada, a nível regional e global, pela inserção de sua cabeceira no Quadrilátero Ferrífero (QF), polo mundial da exploração de minério de ferro, e que abriga diversos empreendimentos minerários de grande porte. A exploração mineral na BHRD data de mais de 200 anos, ocasionando, desde o início, elevado fluxo de pessoas sem ordenamento territorial adequado, intensificando as atividades antrópicas e concomitantemente seus impactos negativos (NASCIMENTO, 2016).

Nesse contexto, no dia 05 de novembro de 2015, ocorreu o rompimento da barragem de Fundão, localizada no distrito de Bento Rodrigues, município de Mariana - MG, pertencente à empresa Samarco Mineração S.A. O barramento era destinado a receber e armazenar o rejeito gerado pela atividade de beneficiamento de minério de ferro. O desastre causou diversos impactos socioeconômicos e ambientais na bacia hidrográfica do rio Doce, com várias implicações na qualidade da água em toda sua extensão (IGAM, 2017).

Dos 50 milhões de metros cúbicos contidos no barramento, cerca de 34 milhões foram liberados diretamente na parte à jusante da barragem, destruindo o vilarejo de Bento Rodrigues e todo o ecossistema do entorno. O rio Gualaxo do Norte apresenta uma extensão de cerca de 60 km, sendo que desses 48 km foram atingidos pelo resíduo oriundo da barragem, isto é, apenas 12 km não foram impactados pelo desastre. O fato de possuir uma parte afetada e não afetada pelo desastre acaba por tornar a bacia hidrográfica do rio Gualaxo do Norte uma área propícia para estudos.

O rompimento da barragem de Fundão, em relação ao volume de rejeitos que foram dispostos no meio ambiente e à magnitude dos danos para a sociedade, está sendo considerado um dos piores desastres mundiais dos últimos tempos; em termos de distância percorrida por rejeitos advindos de barragens é o maior já registrado na América Latina (FERNANDES et al., 2016). Sendo assim, a região necessita de estudos que tenham como foco pesquisas que possam identificar quais são as demandas dos compartimentos ambientais afetados pelo desastre.

Pesquisas na área ambiental têm como característica a compilação de muitos dados, ainda mais se tratando dos estudos que envolvem a problemática até aqui exposta. Nesse sentido, as técnicas de estatística multivariada são ferramentas interessantes, uma vez que permitem a interpretação dessas matrizes extensas, na busca de informações mais concisas e aplicáveis. Isso ocorre devido ao fato de que a análise multivariada realiza a avaliação conjunta das variáveis consideradas, não desprezando as relações entre as mesmas, podendo reduzir a complexidade dos dados, mas mantendo a informação original dos mesmos (SABINO, LAGE e ALMEIDA, 2014).

Sendo assim, o objetivo dessa pesquisa foi realizar a aplicação da técnica multivariada de reconhecimento de padrões supervisionado designada por análise discriminante (AD), aos dados de monitoramento de qualidade da água. Essa técnica foi utilizada com a finalidade de construir modelos de classificação com capacidade preditiva da condição da água da bacia hidrográfica do rio Gualaxo do Norte, em função dos parâmetros de qualidade investigados no monitoramento e em função da região amostrada, nesse caso, regiões impactadas e não impactadas.

## MATERIAIS E MÉTODOS

### PONTOS DE AMOSTRAGEM

Esta pesquisa ocorreu por meio de 12 amostragens de coleta de água, realizadas mensalmente, no período de julho de 2016 a junho de 2017. Foram estabelecidos 27 pontos ao longo da bacia, dos quais 11 estão dispostos na calha principal e 16 nas proximidades da confluência dos tributários com o rio. Determinou-se tais locais observando-se locais impactados pelo desastre, além de bases da hidrografia e densidades populacionais das regiões. As coordenadas geográficas foram obtidas por meio do Sistema de Posicionamento Global (GPS), com equipamento modelo GPS II, marca Garmim, sistema de coordenadas UTM, Datum SIRGAS 2000, Fuso 23K.

### PARÂMETROS DE QUALIDADE DA ÁGUA

As variáveis investigadas foram alcalinidade, condutividade, coliformes totais (CT), demanda bioquímica de oxigênio (DBO), *Escherichia coli* (E. Coli), fósforo total, cloreto, nitrato, nitrito, porcentagem de saturação do oxigênio dissolvido, potencial de oxi-redução (ORP), potencial hidrogeniônico (pH), sólidos totais (ST), temperatura e turbidez. Dentre estes, a condutividade, o ORP e o pH foram medidos in situ, por meio da utilização de um instrumento multiparâmetro portátil da marca Myron L. Company, Ultrameter II, 6P, e a porcentagem de saturação do oxigênio dissolvido e temperatura também foram obtidas in situ, utilizando-se uma sonda HACH 40 D HQ com eletrodo LDO101. As análises laboratoriais (Tabela 1) foram realizadas no Laboratório de Saneamento Ambiental, Laboratório de Estudos em Quimiometria e no Laboratório de Microbiologia e Bioprospecção Tecnológica, todos localizados na Universidade Federal de Ouro Preto (UFOP).

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

PARÂMETROS	MÉTODOS
Alcalinidade	2320 B (APHA, 2012)
CT e <i>E. Coli</i>	Colilert 9223 B (APHA, 2012)
DBO	5210 B (APHA, 2012)
Fósforo total	4500 B (APHA, 2012)
Íons cloreto	Método Titulométrico
Íons nitrato	Método colorimétrico
Íons nitrito	Método colorimétrico
ST	2540 B (APHA, 2012)
Turbidez	Turbidímetro HACH

### TRATAMENTO MULTIVARIADO DOS DADOS

Dentre as diversas técnicas de reconhecimento de padrões supervisionado, as baseadas em similaridades têm como suposição que amostras posicionadas no espaço próximas entre si provavelmente pertençam à mesma categoria; esta ideia de proximidade implica no conceito de distância, onde a análise discriminante considera para o seu desenvolvimento a distância de Mahalanobis. Em suma o algoritmo é implementado em dois estágios; primeiro um modelo é construído e refinado baseado no conjunto de treinamento (i.e., amostras conhecidas); mais tarde o modelo é usado para fazer previsões de classes de novas amostras (i.e., amostras desconhecidas).

A AD é sensível à magnitude dos dados, isto é, se uma variável possuir um valor muito alto e outra variável um valor muito baixo, as medidas serão tendenciosas; e à quantidade numérica de amostras, uma vez que amostras pequenas podem induzir a classificações incorretas. Para que a diferença entre as unidades não interferisse na análise dos dados, os mesmos foram padronizados de acordo com a eq. 1.

$$Y_{ij} = (x_{ij} - x_{mín}/x_{máx} - x_{mín}) * (y_{máx} - y_{mín}) \quad (\text{eq. 1})$$

Em que  $Y_{ij}$  é o valor padronizado,  $x_{ij}$  é o valor original de cada variável,  $i$  é o índice das linhas existentes,  $j$  o índice de colunas existentes,  $x_{máx}$  e  $x_{mín}$  são os valores máximos e mínimos de cada variável, respectivamente, e  $y_{máx}$  e  $y_{mín}$  são os valores máximos e mínimos adotados na padronização, que nesse caso foram 0 e 100.

A matriz utilizada para a AD foi desenvolvida inserindo-se os dados referentes a dez amostragens, levando em consideração questões temporais, isto é, foram inseridas coletas realizadas nas estações úmida e seca. As linhas representam os valores referentes às análises realizadas em cada ponto e as colunas as variáveis consideradas nesse estudo. Cabe ressaltar que esse banco de dados de 324 linhas (12 campanhas com 27 locais amostrados em cada) foi dividido de forma que 66% foram utilizados para a construção do modelo e 22% foram empregados na validação; os 22% restantes foram usados para avaliar a predição.

Os modelos foram construídos de maneira a descrever as características intrínsecas aos dados. Para tal, a avaliação das variáveis responsáveis pela discriminação dos padrões existentes foi realizada por meio da ferramenta stepwise, onde as variáveis independentes são inseridas de forma automática, ponderando a multicolinearidade, e selecionadas de forma a ter como resposta a melhor combinação possível entre elas (SANTOS, 2018). Para o desenvolvimento do modelo de classificação multivariada utilizou-se a versão livre do software Statistica 10.0.

## RESULTADOS

Dentre as diversas técnicas de análise multivariada existem as que objetivam reconhecer um comportamento e buscar um modelo que traduza os padrões encontrados para os dados avaliados. Dessa forma técnicas multivariadas classificatórias são usualmente empregadas para a construção de modelos de classificação para classes pré-definidas, com a possibilidade de verificar em qual classe novas amostras seriam alocadas. Para o desenvolvimento dos modelos propostos nesse trabalho aplicou-se a análise discriminante, abordagem multivariada que a partir de combinações lineares das variáveis avaliadas constrói funções que classificam amostras, de acordo com classes estabelecidas (HAJIGHOLIZADEH e MELESSE, 2017).

De acordo com o desastre ocorrido na região pesquisada e a quantidade de pontos monitorados ao longo da bacia foram desenvolvidos dois modelos a fim de verificar aquele que melhor descreva o sistema, tendo em vista o desastre. A Tabela 2 apresenta as premissas adotadas para o desenvolvimento dos modelos, nos quais o objetivo foi verificar a construção de uma função que conseguisse discriminar corpos hídricos localizados nas regiões afetadas e não afetada.

**Tabela 2: Definição das classes e variáveis utilizadas para a confecção do modelo.**

MODELO	PONTOS	CLASSES	VARIÁVEIS
Modelo 1	1 - 10	Rios não afetados - 1	Alcalinidade (Alca), Coliformes Totais (CT), Condutividade (Cond), Cloreto (Clor), Demanda Bioquímica de Oxigênio (DBO), Escherichia coli (EC), Fósforo Total (PT), Nitrato (NO <sub>3</sub> ), Nitrito (NO <sub>2</sub> ), Potencial Hidrogeniônico (pH), Potencial de Oxi-redução (ORP), Saturação do Oxigênio Dissolvido (SatOD), Sólidos Totais (ST)
	11 - 27	Rios afetados - 2	
Modelo 2	1, 3, 5, 8, 10	Rio principal região não afetada - 1	
	2, 4, 6, 7, 9	Tributários região não afetada - 2	
	11, 14, 18, 21, 25 27	Rio principal região afetada - 3	
	12, 13, 15, 16, 17, 19, 20, 22, 23, 24	Tributários região afetada - 4	

O desenvolvimento da técnica classificatória se estabelece a partir da determinação das variáveis que mais discriminam as amostras em relação às classes estabelecidas. Esse apontamento dos parâmetros discriminatórios ocorre por meio da definição do lambda de Wilks ( $\Lambda$  Wilks). Hatvani (2011) afirma que  $\Lambda$  Wilks indica as

variáveis que contribuem de forma relevante para a formação das classes, podendo variar de 0 a 1, onde os valores próximos de 0 assinalam grandes dissimilaridades entre as classes. A Tabela 3 indica os valores de  $\Delta$  Wilks e as distribuição estatística F e p, para os dois modelos estabelecidos com as suas devidas classes.

A alocação das amostras nas devidas classes ocorre através da função de classificação, que por sua vez é fundamentada na identificação das variáveis de maior poder discriminatório. As funções discriminantes apresentadas na Tabela 4 representam combinações lineares das variáveis consideradas significantes, sendo análogas a uma reta de regressão. Dessa forma, as funções estimadas representam uma tentativa de elucidar toda a informação possível existente nas classes consideradas.

**Tabela 3: Estatística descritiva para as amostras utilizadas na construção dos modelos.**

	MODELO 1			MODELO 2		
	$\Delta$ Wilks	F	p	$\Delta$ Wilks	F	p
Cond	0,92	17,08	0,00	0,81	15,70	0,00
ORP	1,00	-	-	1,00	-	-
SatOD	0,63	122,38	0,00	0,73	25,94	0,00
pH	1,00	-	-	1,00	-	-
Turb	0,82	46,00	0,00	0,29	171,64	0,00
Ta	0,98	4,73	0,03	0,91	6,53	0,00
NO <sub>3</sub>	1,00	-	-	1,00	-	-
NO <sub>2</sub>	0,94	14,47	0,00	1,00	-	-
Clor	0,92	18,68	0,00	0,87	10,51	0,00
PT	1,00	-	-	1,00	-	-
ST	1,00	-	-	1,00	-	-
Alca	0,94	14,48	0,00	0,81	16,10	0,00
CT	1,00	-	-	1,00	-	-
EC	1,00	-	-	0,95	3,49	0,02
DBO	1,00	-	-	1,00	-	-

Alcalinidade (Alca), Coliformes Totais (CT), Condutividade (Cond), Cloreto (Clor), Demanda Bioquímica de Oxigênio (DBO), Escherichia coli (EC), Fósforo Total (PT), Nitrato (NO<sub>3</sub>), Nitrito (NO<sub>2</sub>), Potencial Hidrogeniônico (pH), Potencial de Oxi-redução (ORP), Saturação do Oxigênio Dissolvido (SatOD), Sólidos Totais (ST), Temperatura da água (Ta), Turbidez (Turb).

**Tabela 4: Função de classificação para as classes analisadas para a construção dos dois modelos.**

MODELO 1		MODELO 2				
Função			Classe 1	Classe 2	Classe 3	Classe 4
Cond	0,0152	Cond	0,109	0,062	0,038	0,029
satOD	-0,0548	satOD	0,182	0,11	0,236	0,286
Turb	-0,0206	Turb	-0,168	0,093	-0,137	-0,176
Ta	-0,008	Ta	0,013	0,023	-0,026	-0,024
NO <sub>2</sub>	0,0156	Clor	0,064	0,05	0,087	0,125
Clor	-0,0161	Alca	0,097	0,019	0,032	0,043
Alca	0,0135	EC	-0,021	-0,016	0,01	-0,018
Constante	3,6322	Constante	-14,219	-11,003	-9,914	-14,943

Legenda: Condutividade (Cond), Saturação do Oxigênio Dissolvido (SatOD), Turbidez (Turb), Temperatura da água (Ta), Nitrito (NO<sub>2</sub>), Cloreto (Clor), Alcalinidade (Alca).

Com os valores dos coeficientes de regressão expressos na Tabela 4 são apresentados os modelos, assinalados pelas equações enumeradas de 2 a 6.

- **Modelo 1**

$$y = 0,0152x_1 - 0,0548x_2 - 0,0206x_3 - 0,0080x_4 + 0,0156x_5 - 0,0161x_6 + 0,0135x_7 + 3,6322 \quad (\text{eq. 2})$$

Considerando  $y$  a resposta,  $x_1$ : condutividade,  $x_2$ : saturação do oxigênio dissolvido,  $x_3$ : turbidez,  $x_4$ : temperatura da água,  $x_5$ : nitrito,  $x_6$ : cloreto e  $x_7$ : alcalinidade.

- **Modelo 2**

$$y_1 = 0,109x_1 + 0,182x_2 - 0,168x_3 + 0,013x_4 + 0,064x_5 - 0,097x_6 - 0,021x_7 - 14,219 \quad (\text{eq 3})$$

$$y_2 = 0,062x_1 + 0,110x_2 + 0,093x_3 + 0,023x_4 + 0,050x_5 + 0,019x_6 - 0,016x_7 - 11,003 \quad (\text{eq 4})$$

$$y_3 = -0,038x_1 + 0,236x_2 - 0,137x_3 - 0,026x_4 + 0,087x_5 + 0,032x_6 + 0,010x_7 - 9,914 \quad (\text{eq 5})$$

$$y_4 = 0,029x_1 + 0,286x_2 - 0,176x_3 - 0,024x_4 + 0,125x_5 + 0,043x_6 - 0,018x_7 - 14,943 \quad (\text{eq 6})$$

Considerando  $y_1$ ,  $y_2$ ,  $y_3$  e  $y_4$  as respostas das funções discriminantes e  $x_1$ : condutividade,  $x_2$ : saturação do oxigênio dissolvido,  $x_3$ : turbidez,  $x_4$ : temperatura da água,  $x_5$ : cloreto,  $x_6$ : alcalinidade e  $x_7$ : Escherichia coli.

Para distinguir a classe de cada amostra para o modelo 1 é necessário verificar os centroides de cada grupo em cada função; a partir dos valores dos centroides calcula-se os escores de corte para cada classe e verifica-se a possível alocação de novas amostras. Assim sendo inserindo-se os valores dos respectivos parâmetros para o modelo 1, os que apresentarem valores maiores que 0,0011 são classificados como amostras da classe 1 e os que apresentarem valores menores, alocados na classe 2. Considerando a análise discriminante múltipla com mais de duas classes, além da classificação utilizando o escore de corte, também é possível empregar as funções de classificação, representadas pelas equações do modelo 2, sendo que desse modo cada caso é computado em cada função e inserida na classe de maior escore.

O estágio final de uma AD envolve a exatidão do processo de classificação, sendo necessário verificar o percentual de acerto das funções em relação às classes estabelecidas e assim garantir que os resultados tenham validade. Com esse objetivo, os dados foram divididos para o desenvolvimento do modelo e para a validação do mesmo. As Tabelas 5 e 6 indicam a porcentagem de acerto para os modelos 1 e 2, respectivamente, ressaltando-se que as classificações corretas estão apresentadas nos valores presentes na diagonal principal da tabela, enquanto a diagonal secundária indica os valores incorretos.

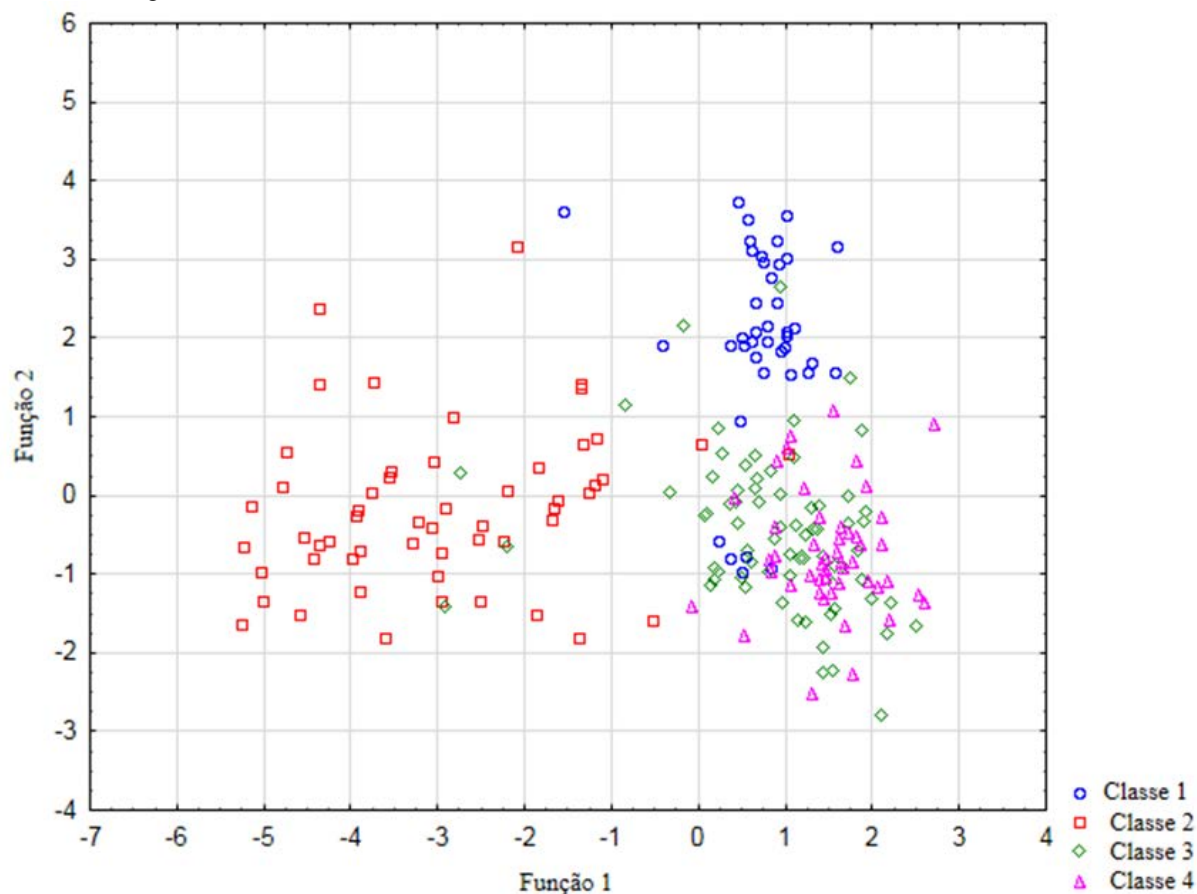
**Tabela 5: Resultado das classificações corretas para o modelo 1.**

CLASSIFICAÇÃO			
Modelo			
	% Correta	Classe 1	Classe 2
Classe 1	90,00	72,00	8,00
Classe 2	94,85	7,00	129,00
Total	93,06	79,00	137,00
Validação			
	% Correta	Classe 1	Classe 2
Classe 1	100,00	17,00	0,00
Classe 2	82,35	6,00	28,00
Total	88,24	23,00	28,00

**Tabela 6: Resultado das classificações corretas para o modelo 2.**

CLASSIFICAÇÃO					
Modelo					
	% Correta	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	85,00	34,00	0,00	6,00	0,00
Classe 2	83,93	2,00	47,00	7,00	0,00
Classe 3	66,67	3,00	3,00	48,00	18,00
Classe 4	54,17	1,00	0,00	21,00	26,00
Total	71,76	40,00	50,00	82,00	44,00
Validação					
	% Correta	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	100,00	9,00	0,00	0,00	0,00
Classe 2	76,92	1,00	10,00	2,00	0,00
Classe 3	64,71	2,00	0,00	11,00	4,00
Classe 4	25,00	0,00	1,00	8,00	3,00
Total	64,71	12,00	11,00	21,00	7,00

Para o modelo 2, como trata-se de uma análise discriminante múltipla, é possível observar as amostras em um gráfico, conhecido como mapa territorial de distribuição das amostras segundo as classes estabelecidas, indicado na Figura 1.



**Figura 1: Mapa territorial da análise discriminante múltipla para o modelo 2.**

Avaliando-se os modelos algumas considerações precisam ser realizadas. O modelo 1 apresentou valores da distribuição lambda de Wilks muito próximos aos valores máximos, o que pode levar a classificações incorretas visto que as variáveis discriminam pouco as classes estabelecidas. Outro ponto relevante são as variáveis saturação de oxigênio dissolvido e a temperatura da água, tidas como algumas das variáveis responsáveis pela dissimilaridade entre os grupos. Tendo em vista que o foco principal é a classificação em função do desastre ocorrido na região, esses parâmetros não são condizentes. A saturação do oxigênio dissolvido, nesse caso, é discriminatória devido às características geográficas da bacia, que possui elevada altitude, fazendo com que sua foz tenha, por natureza, maiores índices de reoxigenação. A temperatura da água também não pode ser considerada como fator discriminatório uma vez que as amostragens não ocorreram na mesma hora nos pontos avaliados.

O modelo 2, embora tenha também apresentado a saturação do oxigênio dissolvido e a temperatura como variáveis discriminatórias, apresentou como o principal parâmetro responsável pela discrepância entre as classes a turbidez, o que em termos do desastre é coerente com o ocorrido na região, uma vez que esse foi o parâmetro mais afetado com o rompimento da barragem de Fundão. Além disso variáveis como alcalinidade, cloreto e *Escherichia coli* representam as ações antrópicas que já impactavam as águas da referida bacia hidrográfica, como o descarte incorreto de efluentes domésticos e industriais.

Dessa forma, o modelo 2 é o mais indicado para ser aplicado para a previsão de futuras amostras, uma vez que é o mais fidedigno aos processos desenvolvidos na bacia avaliada. Ainda ressalta-se que dos 15 parâmetros avaliados, o modelo tido como o melhor necessita de sete variáveis, isto é, ocorreu a diminuição de mais de 50% das determinações a serem realizadas.

Salienta-se ainda que o referido modelo pode atuar na predição de amostras novas, ou seja, não utilizadas para a construção do modelo, podendo ser utilizado para a determinação da condição da qualidade da água, explicitando se a amostra está com característica de região afetada e quando estaria mais próxima da condição de não impactada.

## CONCLUSÃO

O monitoramento da qualidade da água durante um ciclo hidrológico realizado na bacia do rio Gualaxo do Norte - MG, abrangendo regiões a montante e a jusante do rompimento da barragem de Fundão, foi importante para a compreensão dos impactos antrópicos que ocorriam na região antes e após o desastre ocorrido em novembro de 2015.

Tendo em vista os resultados obtidos, a abordagem de estatística multivariada utilizada foi útil para auxiliar a interpretação das matrizes de dados obtidas com o ciclo hidrológico monitorado, fornecendo subsídio para a compreensão das diferentes regiões investigadas, de acordo com as características da bacia hidrográfica avaliada. Os modelos de classificação que foram desenvolvidos são capazes de prever a adesão à classe mais provável para novas amostras, isto é, para novas campanhas de monitoramento da região, destacando que, no caso de pesquisas que tenham como intuito verificar o diagnóstico entre área afetada e não afetada, houve a diminuição potencial dos parâmetros que devem ser determinados.

## AGRADECIMENTOS

Os autores agradecem à CAPES, CNPQ, FAPEMIG, Fundação Gorceix, Proamb e UFOP pelo auxílio financeiro concedido ao trabalho.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. APHA. *Standard Methods for examination of water and wastewater*. Washington: American Water Work Association. Water Environmental Federation. 2012
2. CABRIA, M. A., BARQUÍN, J., PEÑAS, F. J. *Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors*. *Journal Science of the Total Environment*, vol. 545, n. 546, p. 152 – 162, 2016.

3. 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 JÚNIOR, P. M.; MILLAN, V. E. G.; MIELKE, C.; RAMIREZ, J. L.; NEVES, A.; ROGASS, C.; RIBEIRO, S. P. R.; FILHO, B. S. *Deep into the mud: ecological and socio-economic impacts of the dam breach in Mariana, Brazil. Natureza e Conservação*, vol. 2, n. 2, p. 35 – 45, 2016.
4. HATVANI, I. G.; JÓZSEF K.; ILONA S. K.; PÁL J.; KORPONAI, J. *Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. Ecological Engineering*, [s.l.], v. 37, n. 4, p.629-635, abr. 2011.
5. HAJIGHOLIZADEH, M.; MELESSE, A. M. *Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. Catena*, [s.l.], v. 151, p.247-258, abr. 2017.
6. IGAM, Instituto de Gestão de Águas Mineiras. Acompanhamento da Qualidade das Águas do Rio Doce Após o Rompimento da Barragem da Samarco no distrito de Bento Rodrigues – Mariana/MG (17/05/17). Disponível em [www.igam.mg.gov.br](http://www.igam.mg.gov.br), (acessado em 15 de maio de 2017), 2017.
7. NASCIMENTO, L P. Análise Ambiental da Bacia Hidrográfica do Rio do Peixe, Quadrilátero Ferrífero, Minas Gerais. 170 f. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Ambiental, Universidade Federal de Ouro Preto, Ouro Preto, 2016.
8. OBADE, V. P., MOORE, R. *Synthesizing water quality indicators from standardized geospatial information to remedy water security challenges: A review. Journal Environment International*, vol.119, p. 220 – 331, 2018.
9. SABINO, C. V. S.; LAGE, L. V.; ALMEIDA, K. C. B. Uso de métodos estatísticos robustos na análise ambiental. *Engenharia Sanitária e Ambiental*, [s.l.], v. 19, n. 1, p.87-94, 2014.
10. SANTOS, G. R. Técnicas de reconhecimento de padrões para a avaliação de corpos hídricos afetados por rejeitos do rompimento da barragem de Fundão. Dissertação (Mestrado). Programa de Pós-graduação em Engenharia Ambiental. Universidade Federal de Ouro Preto. Ouro Preto, 2018.