

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA**

CAROLINA CRISTIANE PINTO

**MONITORAMENTO DE QUALIDADE DE ÁGUA BASEADO EM ANÁLISE POR
COMPONENTES PRINCIPAIS: UMA ABORDAGEM LOCAL**

**BELO HORIZONTE - MG
2021**

CAROLINA CRISTIANE PINTO

**MONITORAMENTO DE QUALIDADE DE ÁGUA BASEADO EM ANÁLISE POR
COMPONENTES PRINCIPAIS: UMA ABORDAGEM LOCAL**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do Grau de Doutora em Engenharia Química.

Linha de Pesquisa: Engenharia de Sistemas em Processos.

Orientador: Dr. Gustavo Matheus de Almeida.

Coorientadora: Dra. Sílvia Maria Alves Corrêa Oliveira.

P659m

Pinto, Carolina Cristiane.

Monitoramento de qualidade de água baseado em análise por componentes principais [recurso eletrônico] : uma abordagem local / Carolina Cristiane Pinto. – 2021.

1 recurso online (xv, 151 f. : il., color.) : pdf.

Orientador: Gustavo Matheus de Almeida.

Coorientadora: Sílvia Maria Alves Corrêa Oliveira.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Engenharia.

Anexos: f. 150-151.

Bibliografia: f. 133-149.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia química - Teses. 2. Água - Qualidade - Teses. 3. Água - Análise - Teses. 4. Visualização de dados - Teses. 5. Monitoramento ambiental - Teses. I. Almeida, Gustavo Matheus de. II. Oliveira, Sílvia Maria Alves Corrêa. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 66.0(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

FOLHA DE APROVAÇÃO

"MONITORAMENTO DE QUALIDADE DE ÁGUA BASEADO EM ANÁLISE POR COMPONENTES PRINCIPAIS: UMA ABORDAGEM LOCAL"

Carolina Cristiane Pinto

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Química da Escola de Engenharia da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de **DOCTORA EM ENGENHARIA QUÍMICA**.

55ª TESE APROVADA EM 12 DE MARÇO DE 2021 POR:



Documento assinado eletronicamente por **Gustavo Matheus de Almeida, Professor do Magistério Superior**, em 12/03/2021, às 17:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Silvia Maria Alves Correa Oliveira, Professora do Magistério Superior**, em 12/03/2021, às 17:27, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Roberto da Costa Quinino, Coordenador(a) de curso**, em 12/03/2021, às 17:27, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Coutinho de Paula, Professor do Magistério Superior**, em 12/03/2021, às 17:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Patricia da Luz Mesquita, Usuário Externo**, em 12/03/2021, às 17:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Wanderlene Ferreira Nacif, Usuário Externo**, em 12/03/2021, às 17:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0600437** e o código CRC **09A01641**.

AGRADECIMENTOS

Meus sinceros agradecimentos,

A Deus, por sempre iluminar meus caminhos e guiar meus passos.

Aos meus pais, José e Sandra, pelo amor incondicional, incentivo, exemplo e suporte em todos os momentos. Obrigada pelos valores que me ensinaram.

Aos meus irmãos, Camila, Cristiano e Clarissa, por estarem sempre ao meu lado e torcendo pelo meu sucesso. Vocês iluminam meus dias!

Ao Felipe, obrigada pela paciência e pelas palavras de motivação e incentivo. Obrigada por compreender minha ausência em diversos momentos e dividir comigo os sonhos e os dias.

Ao meu orientador, Gustavo, pela confiança e orientação, contribuindo para meu amadurecimento profissional.

À minha coorientadora Sílvia, por todo apoio e ensinamentos. Obrigada por acreditar em mim, e, principalmente, por me fazer acreditar que eu seria capaz! Você é meu exemplo a ser seguido!

Aos professores do Programa, que contribuíram para minha formação.

Aos membros da banca examinadora, que gentilmente aceitaram avaliar esta Tese e pelas contribuições realizadas.

Aos membros dos grupos de pesquisa dos quais faço parte: Data Analysis and Visualization Group (DAG) e Grupo de Estudos para Tratamento Estatístico de Dados Ambientais (GETEDA). Obrigada pelos saberes compartilhados.

Ao Fábio Palmer, pelo auxílio, que foi muito importante para o início da realização dessa pesquisa.

À Livia e à Eliz, obrigada pela amizade, ajuda e generosidade.

À Ana, obrigada pela amizade, por toda ajuda e infinitas discussões, científicas ou não! Com certeza a caminhada ficou mais leve.

À Regis, pelas palavras sempre de carinho e atenção. Obrigada por todos os conselhos e por sua alegria contagiante!

Ao Instituto Mineiro de Gestão das Águas (Igam) e à Agência Nacional de Águas e Saneamento Básico (ANA), pela disponibilização dos dados, sem os quais não seria possível realizar esta pesquisa.

A todos que, de alguma forma, fizeram parte dessa conquista!

Muito obrigada!!!

É Preciso Também Não Ter Filosofia Nenhuma

Não basta abrir a janela para ver os campos e o rio.

Não é bastante não ser cego para ver as árvores e as flores.

É preciso também não ter filosofia nenhuma. Com filosofia não há árvores: há ideias apenas.

Há só cada um de nós, como uma cave.

Há só uma janela fechada, e todo o mundo lá fora;

E um sonho do que se poderia ver se a janela se abrisse,

Que nunca é o que se vê quando se abre a janela.

***Alberto Caeiro,
in “Poemas Inconjuntos”
(Heterônimo de Fernando Pessoa)***

RESUMO

O monitoramento da qualidade da água superficial ganhou mais atenção nos últimos anos devido à crescente degradação dos ambientes aquáticos. Os dados obtidos a partir desse monitoramento são essenciais no gerenciamento de recursos hídricos, e o desafio é transformá-los em informação relevante. Muitas vezes, a dificuldade existe não pela quantidade de dados, mas pelas interações complexas existentes nos cursos de água e do caráter multivariável dos dados. Nesse sentido, o objetivo deste trabalho foi aplicar a técnica multivariada de Análise por Componentes Principais (ACP), com foco em três aplicações distintas. As metodologias foram propostas para serem aplicadas a uma estação de monitoramento individualmente (abordagem local), possibilitando uma descrição detalhada das condições de qualidade de água de uma região específica. Como estudo de caso, foram utilizados dados secundários de 12 parâmetros físico-químicos e microbiológicos, coletados durante 21 anos (1997-2018), em uma estação de amostragem (BV137), identificada, em estudos anteriores, como prioritária para o monitoramento e controle da poluição do rio das Velhas. Na primeira aplicação, o modelo de ACP construído foi capaz de identificar amostras críticas em termos de poluição, e por meio de gráficos de *scores*, foi possível associar regiões espaciais aos fatores de pressão ambiental característicos da área em estudo (lançamentos de efluentes domésticos e industriais e atividades minerárias). Na segunda aplicação, as cartas de controle multivariadas para as estatísticas T^2 e Q foram capazes de identificar 79,73% e 72,97% dos dados extremos, respectivamente, mostrando-se promissoras no monitoramento para a detecção das condições críticas de qualidade de água. Na terceira aplicação, foram realizadas análises envolvendo ACP e testes não paramétricos para avaliação da variabilidade temporal dos dados. Os resultados indicaram que três componentes principais foram responsáveis por explicar 62,47% da variabilidade total da qualidade da água e que os meses de janeiro, julho e outubro foram os que mais contribuíram para a explicação da variabilidade dos dados. As análises realizadas podem ser utilizadas como suporte à tomada de decisão em atividades de monitoramento e gerenciamento de qualidade de água, fornecendo meios para direcionar e avaliar a eficácia das políticas de controle e prevenção da poluição.

Palavras-chave: Monitoramento; Qualidade de água; Análise por Componentes Principais; Visualização de dados.

ABSTRACT

Surface water quality monitoring has gained more attention in recent years due to the increasing degradation of aquatic environments. The data obtained from this monitoring is essential for the management of water resources, whose challenge lies in transforming them into relevant information. Often, the difficulty exists not because of the amount of data, but due to the complex existing interactions in watercourses and to the multivariate nature of the data. In this sense, the objective of this work was to apply the statistical multivariate technique, called Principal Component Analysis (PCA), with a focus on three different applications. The methodologies were proposed to be applied to a monitoring station individually (local approach), allowing a detailed description of the water quality conditions in a specific region. As a case study, secondary data from 12 physical-chemical and microbiological parameters were used, collected over 21 years (1997-2018) in a sampling station (BV137) identified in previous studies as a priority for pollution monitoring and control in the Velhas River. In the first application, the PCA model was able to identify critical samples in terms of pollution, and through score plots, it was possible to associate spatial regions in them with environmental pressure factors characteristic of the study area (discharge of domestic and industrial effluents and mining activities). In the second application, the multivariate control charts for the T^2 and Q statistics were able to identify 79.73% and 72.97% of the extreme data, respectively, showing promise in monitoring for the detection of critical water quality conditions. In the third application, analyzes involving PCA and non-parametric tests were performed to assess the temporal data variability. The results indicated that three main principal components were responsible for explaining 62.47% of the total variability in water quality, and that January, July and October were the months that most contributed to explain the variability of the data. The analyzes performed can be used to support decision making in water quality monitoring and management activities providing means to handle and evaluate the effectiveness of pollution control and prevention policies.

Keywords: Monitoring; Water quality; Principal Component Analysis; Data visualization.

LISTA DE FIGURAS

Figura 1.1 - Visão geral do projeto de uma rede de monitoramento	3
Figura 2.1 - Representação geométrica das componentes principais.....	15
Figura 2.2 - Redução de dimensionalidade em um conjunto de dados após aplicação de ACP.....	16
Figura 2.3 - Modelo de ACP do conjunto de dados tridimensionais com destaque para valores <i>outliers</i> das estatísticas T^2 e Q	21
Figura 3.1 - Localização geográfica da bacia hidrográfica do rio das Velhas e seus municípios	27
Figura 3.2 - Pontos de monitoramento de qualidade da água superficial da rede básica do Programa Águas de Minas em operação no ano de 2018	28
Figura 3.3 - Localização da estação de monitoramento de qualidade da água superficial BV137	30
Figura 3.4 - Registro fotográfico das regiões (a) a montante e (b) a jusante da estação de monitoramento BV137.....	31
Figura 3.5 - Mapa de uso e ocupação do solo da bacia do rio das Velhas, com foco na bacia de drenagem da estação de monitoramento BV137	33
Figura 3.6 - Distribuição percentual das categorias de uso e cobertura do solo na região do Médio Alto rio das Velhas	34
Figura 3.7 - Mapa das Áreas Prioritárias para Conservação na bacia hidrográfica do rio das Velhas.....	36
Figura 3.8 - Localização geográfica das estações de monitoramento pluviométrico, fluviométrico e de qualidade da água superficial, utilizadas neste estudo.....	39
Figura 3.9 - Histograma dos parâmetros de qualidade de água analisados no período de 1997 a 2018	44
Figura 3.10 - Gráfico de séries temporais dos parâmetros de qualidade de água analisados no período de 1997 a 2018	45
Figura 3.11 - Percentual de violação aos limites legais, considerando os dados no período de 1997 a 2018	48
Figura 3.12 - Representação gráfica da matriz de correlação de Spearman relacionando todas as variáveis de qualidade de água selecionadas	49

Figura 3.13 - Gráfico temporal com dados discrepantes e não discrepantes identificados pelo filtro de Hampel para os parâmetros (a) arsênio total, (b) DBO, (c) fósforo total e (d) turbidez.....	51
Figura 4.1 - Fluxograma da metodologia adotada para o mapeamento e a classificação das fontes de poluição.....	59
Figura 4.2 - Gráficos de <i>scores</i> para as combinações das componentes principais: (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3.....	63
Figura 4.3 - Histogramas para as componentes (a) CP1, (b) CP2 e (c) CP3	64
Figura 4.4 - Gráficos de <i>scores</i> com destaque na região 1 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3.....	66
Figura 4.5 - <i>Box-plot</i> para comparação das concentrações entre a região 1 e a região de referência para os parâmetros (a) DBO, (b) Colif. term. e (c) OD	70
Figura 4.6 - Gráficos de <i>scores</i> com destaque na região 2 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3.....	71
Figura 4.7 - <i>Box-plot</i> para comparação das concentrações entre a região 2 e a região de referência para os parâmetros (a) turbidez e (b) sólidos totais	73
Figura 4.8 - Gráficos de <i>scores</i> com destaque na região 3 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3.....	74
Figura 5.1 - Visão geral de um gráfico de controle univariado	82
Figura 5.2 - Comparação entre as abordagens de cartas de controle univariada (à esquerda) e multivariada (à direita).....	86
Figura 5.3 - Metodologia utilizada para o desenvolvimento da abordagem de CEP Multivariado baseado em ACP para o monitoramento de qualidade da água.....	89
Figura 5.4 - Fluxograma da metodologia utilizada para a aplicação de CEPM-ACP no monitoramento de qualidade de água	93
Figura 5.5 - Divisão do banco de dados para aplicação de CEPM-ACP	95
Figura 5.6 - Gráfico <i>scree</i> com os autovalores e a variabilidade acumulada das componentes principais obtidas a partir do modelo de ACP	96
Figura 5.7 - Cartas de controle multivariadas para as estatísticas (a) T^2 e (b) Q , calculadas a partir do conjunto de dados de teste	97
Figura 5.8 - Gráfico de influência dos dados de qualidade da água calculado a partir de um modelo de cinco componentes principais.....	99

Figura 5.9 - Percentual de contribuição dos parâmetros de qualidade de água na estatística Q , considerando as amostras críticas detectadas	100
Figura 6.1 - Fluxograma da metodologia adotada para a avaliação das variações temporais da qualidade da água superficial	112
Figura 6.2 - Precipitação média mensal (em mm), na bacia de drenagem da estação BV137, de 1984 a 2018 (1° semestre)	115
Figura 6.3 - <i>Box-plot</i> da precipitação acumulada, considerando o conjunto de estações pluviométricas analisadas, no período de 1984 a 2018	116
Figura 6.4 - <i>Box-plot</i> da precipitação acumulada, considerando o conjunto de estações pluviométricas analisadas, no período de 1984 a 2018 (1° semestre), por período climático	116
Figura 6.5 - <i>Box-plot</i> para os parâmetros (a) cloreto total, (b) condutividade elétrica, (c) turbidez, (d) sólidos totais, (e) fósforo total, (f) temperatura da água e (g) vazão, considerando os períodos chuvoso e seco	120
Figura 6.6 - Gráfico <i>scree</i> com os autovalores e a variabilidade acumulada das componentes principais obtidos na ACP	123
Figura 6.7 - Variabilidade temporal das componentes principais, no período de monitoramento de 1997 a 2018, considerando o percentual de contribuição mensal	125
Figura 6.8 - <i>Biplot</i> para as componentes CP1 e CP2 considerando o período climático seco e chuvoso	127

LISTA DE TABELAS

Tabela 3.1 - Áreas Prioritárias para a conservação na bacia hidrográfica do rio das Velhas	35
Tabela 3.2 - Descrição das estações pluviométricas localizadas na bacia de drenagem da estação BV137	38
Tabela 3.3 - Parâmetros físico-químicos e microbiológicos da qualidade da água com respectiva abreviação, unidade e técnica analítica	43
Tabela 3.4 - Estatística descritiva do banco de dados de monitoramento da estação BV137, no período de 1997 a 2018.....	46
Tabela 4.1 - Resultados do modelo ACP obtido a partir de observações não críticas	65
Tabela 4.2 - Resumo das estatísticas básicas dos parâmetros de qualidade da água para as observações da região 1, entre 1997 e 2018	68
Tabela 5.1 - Modelo de ACP obtido a partir de observações não críticas.....	95
Tabela 5.2 - Resumo dos resultados apresentados pelas cartas de controle T^2 de Hotelling e de resíduos Q	99
Tabela 6.1 - Estatística descritiva dos dados amostrados na estação BV137 nos períodos chuvoso e seco.....	118
Tabela 6.2 - Resultado da ACP com as cargas obtidas para cada componente principal	124

LISTA DE QUADROS

Quadro 1.1 - Estudos que utilizaram dados de monitoramento da qualidade da água do Programa Águas de Minas do Igam	8
Quadro 2.1 - Aspectos característicos dos conjuntos de dados ambientais e conceitos necessários para a preparação do banco de dados antes da aplicação de técnicas estatísticas	19
Quadro 3.1 - Parâmetros de qualidade da água analisados nas estações de amostragem do Igam, no Programa Águas de Minas.	29
Quadro 3.2 - Regiões da bacia hidrográfica do rio das Velhas e principais características	32
Quadro 4.1 - Informações sobre estudos nos quais a ACP foi aplicada a dados de qualidade de água superficial.....	55

LISTA DE SIGLAS

AC - Análise de Cluster
ACC – Análise de Correlação Canônica
ACP - Análise por Componentes Principais
AD - Análise Discriminante
ADFLU - Área da bacia de drenagem da estação fluviométrica
ADQA - Área da bacia de drenagem da estação de qualidade da água
AF - Análise Fatorial
ANA - Agência Nacional de Águas e Saneamento Básico
ANOVA - Análise de Variância
CBH - Comitê de Bacia Hidrográfica
CBHSF - Comitê de Bacia Hidrográfica do rio São Francisco
CE - Condutividade elétrica
CEP - Controle Estatístico de Processo
CEPM - Controle Estatístico de Processo Multivariado
CEPM-ACP - Controle Estatístico de Processo Multivariado e Análise por Componentes Principais
CETESB - Companhia Ambiental do Estado de São Paulo
CH - Circunscrição hidrográfica
COPAM - Conselho Estadual de Política Ambiental
CP - Componente principal
CUSUM - Gráfico de Controle da Soma Cumulativa
CV - Coeficiente de variação
DBO - Demanda bioquímica de oxigênio
DN - Deliberação Normativa
Igam - Instituto Mineiro de Gestão das Águas
IQA - Índice de Qualidade de Água
IWM - Integrated Watershed Management
IWRM - Integrated Water Resources Management
LC - Linha central
LSC - Limite superior de controle
LIC - Limite inferior de controle

LD - Limite de detecção
MAD - Median absolute deviation
MMEP - Média móvel exponencialmente ponderada
NMP - Número mais provável
OD - Oxigênio dissolvido
PCA - Principal Components Analysis
PMQA - Programas de monitoramento da qualidade da água
PNQA - Programa Nacional de Avaliação da Qualidade das Águas
RMBH - Região metropolitana de Belo Horizonte
RHN - Rede Hidrometeorológica Nacional
SNIRH - Sistema Nacional de Informações sobre Recursos Hídricos
ST - Sólidos totais
UC - Unidade de conservação
UNCED - Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento
UNT - Unidade nefelométrica de turbidez

APRESENTAÇÃO

Para melhor organizar o trabalho desenvolvido, esta Tese foi dividida em sete capítulos denominados: Capítulo I - Apresentação do tema; Capítulo II - Revisão teórica da Análise por Componentes Principais (ACP); Capítulo III – Definição do estudo de caso e pré-processamento dos dados de monitoramento; Capítulo IV - Utilização de modelo baseado em Análise por Componentes Principais para identificação de condições críticas de qualidade de água superficial; Capítulo V - Monitoramento de qualidade de água superficial por meio de Controle Estatístico de Processo Multivariado; Capítulo VI - Monitoramento e avaliação estatística das variações temporais na qualidade da água: estudo de caso do rio das Velhas e Capítulo VII – Considerações finais.

Dentro dessa organização, o Capítulo I visa apresentar o tema e contextualizar a tese, apresentando o problema da poluição dos cursos de água superficiais, a importância dos programas de monitoramento e os diferentes métodos/técnicas utilizados para análise dos dados. Além disso, apresenta os objetivos traçados e demonstra a relevância do trabalho desenvolvido. O Capítulo II tem como objetivo apresentar detalhes matemáticos das abordagens analíticas e geométricas da Análise por Componentes Principais. Os Capítulos III, IV, V e VI referem-se a cada objetivo específico apresentado na tese. O Capítulo III aborda a definição do estudo de caso, com a descrição da localização da bacia hidrográfica, principais características, estações de monitoramento, detalhamento dos bancos de dados que foram utilizados na pesquisa e todas as etapas de pré-processamento dos dados. No Capítulo IV, foi utilizado um modelo de ACP para a identificação de padrões críticos de comportamento e classificação de fontes de poluição da água. O Capítulo V apresenta uma abordagem combinada, integrando Controle Estatístico de Processo Multivariado e Análise por Componentes Principais (CEPM-ACP), que foi utilizada para detecção de condições críticas da qualidade da água e geração de sinais de alerta, de acordo com os dados históricos de observação. O Capítulo VI, por sua vez, apresenta a aplicação de ACP para a detecção e avaliação de variações temporais em dados de qualidade da água. Finalmente, no Capítulo VII são apresentadas as considerações finais do estudo realizado.

SUMÁRIO

CAPÍTULO 1: APRESENTAÇÃO DO TEMA	1
1.1 CONTEXTUALIZAÇÃO	2
1.2 JUSTIFICATIVA	9
1.3 OBJETIVOS	11
1.3.1 Objetivo Geral	11
1.3.2 Objetivos Específicos	11
CAPÍTULO 2: REVISÃO TEÓRICA DA ANÁLISE POR COMPONENTES PRINCIPAIS	12
CAPÍTULO 3: DEFINIÇÃO DO ESTUDO DE CASO E PRÉ-PROCESSAMENTO DOS DADOS DE MONITORAMENTO	24
3.1 ÁREA DE ESTUDO	25
3.1.1 Localização e características gerais.....	25
3.1.2 Monitoramento da qualidade das águas superficiais em Minas Gerais.....	28
3.1.2.1 Monitoramento da qualidade das águas superficiais na bacia do rio das Velhas	29
3.1.3 Uso e ocupação do solo	32
3.1.4 Áreas prioritárias para a conservação.....	34
3.1.5 Monitoramento pluviométrico e fluviométrico	37
3.2 DEFINIÇÃO DO BANCO DE DADOS DE MONITORAMENTO DE QUALIDADE DA ÁGUA	40
3.3 PRÉ-PROCESSAMENTO DOS DADOS	41
3.4 RESULTADOS E DISCUSSÃO	43
3.4.1 Definição do banco de dados de monitoramento de qualidade da água.....	43
3.4.2 Pré-processamento dos dados.....	44
CAPÍTULO 4: UTILIZAÇÃO DE MODELO BASEADO EM ANÁLISE POR COMPONENTES PRINCIPAIS PARA IDENTIFICAÇÃO DE CONDIÇÕES CRÍTICAS DE QUALIDADE DE ÁGUA SUPERFICIAL	52
4.1 INTRODUÇÃO	54
4.2 METODOLOGIA.....	59
4.2.1 Definição do problema.....	59
4.2.2 Pré-processamento dos dados.....	60

4.2.3 Definição de modelo baseado em ACP para o monitoramento de condições críticas de qualidade da água.....	60
4.2.3.1 Reconhecimento de observações críticas	60
4.2.3.2 Mapeamento e classificação das fontes de poluição.....	61
4.3 RESULTADOS E DISCUSSÃO	62
4.3.1 Definição do banco de dados de monitoramento de qualidade da água e pré-processamento dos dados	62
4.3.2 Definição de modelo baseado em ACP para o monitoramento de condições críticas de qualidade da água.....	62
4.3.2.1 Reconhecimento de observações críticas	62
4.3.2.2 Mapeamento e classificação das fontes de poluição.....	64
4.4 CONCLUSÕES	76
CAPÍTULO 5: MONITORAMENTO DE QUALIDADE DE ÁGUA SUPERFICIAL POR MEIO DE CONTROLE ESTATÍSTICO DE PROCESSO MULTIVARIADO	78
5.1 INTRODUÇÃO	80
5.2 METODOLOGIA.....	88
5.2.1 Área de estudo	88
5.2.2 Banco de dados.....	88
5.2.3 Pré-processamento dos dados.....	88
5.2.4 Monitoramento de qualidade da água com Controle Estatístico de Processo Multivariado baseado em ACP (CEPM-ACP).....	88
5.2.4.1 Etapa 1 – Avaliação preliminar do banco de dados.....	89
5.2.4.2 Etapa 2 – Construção das cartas de controle multivariadas.....	89
5.2.4.3 Etapa 3 – Avaliação das cartas de controle	91
5.3 RESULTADOS E DISCUSSÃO	94
5.3.1 Banco de dados e pré-processamento dos dados	94
5.3.1 Monitoramento de qualidade da água com Controle Estatístico de Processo Multivariado baseado em ACP (CEPM-ACP).....	94
5.4 CONCLUSÃO.....	103
CAPÍTULO 6: MONITORAMENTO E AVALIAÇÃO ESTATÍSTICA DAS VARIAÇÕES TEMPORAIS NA QUALIDADE DA ÁGUA SUPERFICIAL: ESTUDO DE CASO DO RIO DAS VELHAS	105
6.1 INTRODUÇÃO	107

6.2 METODOLOGIA.....	109
6.2.1 Área de estudo	109
6.2.2 Banco de dados.....	109
6.2.2.1 Dados de monitoramento de qualidade da água	109
6.2.2.2 Dados de monitoramento pluviométrico e fluviométrico	109
6.2.3 Dados de vazão obtidos a partir do monitoramento fluviométrico	109
6.2.4 Avaliação dos dados de monitoramento pluviométrico.....	111
6.2.5 Avaliação das variações temporais da qualidade da água	113
6.2.5.1 Análises estatísticas preliminares do banco de dados	113
6.2.5.2 Análise por Componentes Principais.....	113
6.3 RESULTADOS E DISCUSSÃO	115
6.3.1 Avaliação dos dados de monitoramento pluviométrico.....	115
6.3.2 Avaliação das variações temporais da qualidade da água.....	117
6.3.2.1 Análises estatísticas preliminares do banco de dados	117
6.3.2.2 Análise por Componentes Principais.....	123
6.4 CONCLUSÃO.....	128
CAPÍTULO 7: CONCLUSÕES GERAIS	130
REFERÊNCIAS BIBLIOGRÁFICAS	133
ANEXO 1	150

CAPÍTULO 1: APRESENTAÇÃO DO TEMA

1.1 CONTEXTUALIZAÇÃO

A degradação da qualidade da água de rios, lagos e águas subterrâneas tornou-se um problema ambiental de abrangência mundial (BEHMEL et al., 2016). Nas últimas décadas, os ecossistemas aquáticos foram afetados pela elevada carga de poluentes, principalmente, devido às atividades antrópicas (como os lançamentos de águas residuárias domésticas, lixiviados de resíduos sólidos, efluentes industriais e escoamento de atividades agrícolas) (KAZI et al., 2009; TIAN et al., 2019; ZHANG et al., 2010).

Nesse cenário, a qualidade da água dos rios tem se tornado um fator limitante para o desenvolvimento social e econômico (LI; ZOU; AN, 2016), pois a água de baixa qualidade põe em risco a saúde humana e os ecossistemas e reduz a disponibilidade de água potável e de recursos hídricos próprios para outras finalidades (ANA, 2011).

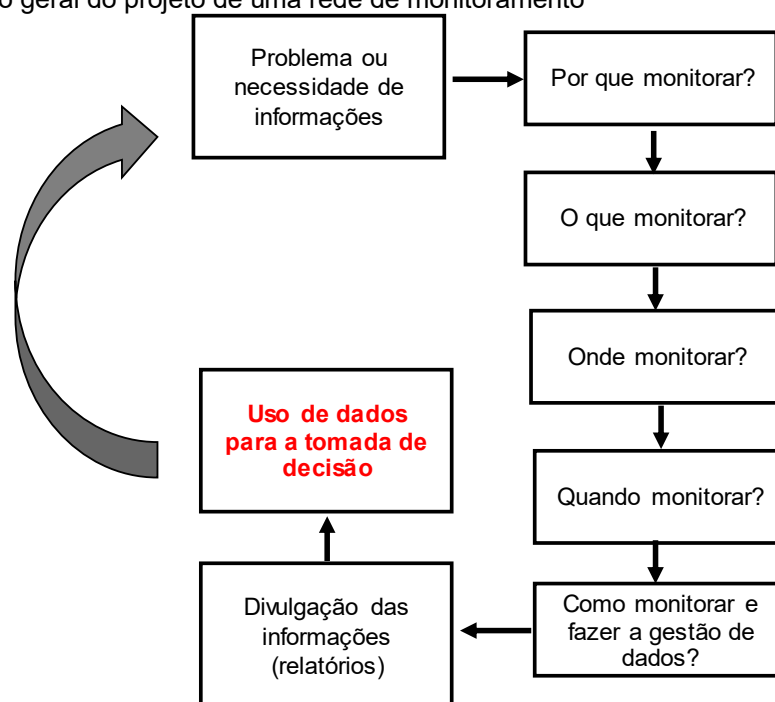
Em consequência dessa crescente pressão sobre os recursos hídricos e visando uma reforma na governança da água, vem sendo adotado universalmente o “desenvolvimento e o manejo integrado dos recursos hídricos”, proposto na Agenda 21, em 1992, na UNCED (Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento), com o objetivo de prevenir e gerenciar a poluição da água (BEHMEL et al., 2016; UNEP, 2012). A partir desse gerenciamento integrado das bacias hidrográficas (*Integrated Watershed Management – IWM*), referido globalmente como Gestão Integrada de Recursos Hídricos (*Integrated Water Resources Management - IWRM*), todas as atividades humanas e recursos naturais utilizados em uma determinada bacia devem ocorrer de forma coordenada e sustentável (BEHMEL et al., 2016; CONSERVATION ONTARIO, 2010). De acordo com Behmel et al. (2016), um dos principais desafios da gestão integrada de bacias hidrográficas é obter uma avaliação confiável da qualidade da água superficial de lagos e rios.

Na Agenda 21, foi proposto, em seu capítulo 18, um programa de “Avaliação dos Recursos Hídricos”, sendo o estabelecimento de bancos de dados nacionais vital para essa avaliação e para a mitigação dos efeitos não só da poluição como também de enchentes, secas e desertificação. Esses programas devem compreender a determinação contínua de fontes potenciais de água, extensão, confiabilidade e qualidade dos recursos hídricos e das atividades humanas que os afetam (MMA, 2019a).

Como a qualidade da água é afetada por variações espaço-temporais, o estabelecimento de programas de monitoramento da qualidade da água (PMQA), regulares e contínuos, é necessário para que forneçam dados representativos e confiáveis (SHRESTHA; KAZAMA, 2007; SIMEONOV et al., 2003; SINGH; MALIK; SINHA, 2005; VOZA; VUKOVIĆ, 2018).

Etapas essenciais estão envolvidas no estabelecimento dos PMQA: coleta de amostras realizada sistematicamente em diversas estações (medições *in loco*), com frequência definida (para reunir dados suficientes para avaliar as variações espaciais e/ou temporais na qualidade da água), e análise das amostras para determinação da concentração de diversos parâmetros, seguida do estudo e da avaliação dos resultados analíticos e do relato dos resultados (BARTRAM et al., 1996). As questões que devem fazer parte da estrutura de base dos programas de monitoramento são: o porquê, o quê, onde, quando e como amostrar (Figura 1.1) (GOVERNMENT OF WESTERN AUSTRALIA, 2009).

Figura 1.1 - Visão geral do projeto de uma rede de monitoramento



Fonte: Adaptado de Government of Western Australia (2009).

O estudo de Behmel et al. (2016), apresenta uma revisão de literatura para relatar as questões que um gestor de bacia hidrográfica deve abordar ao planejar ou otimizar um PMQA: por exemplo, otimizar o número e a distribuição de pontos de

amostragem; avaliar a representatividade de uma rede de amostragem para a qualidade da água de uma bacia hidrográfica; avaliar a representatividade do tipo e número de parâmetros de qualidade da água, explorar relações entre parâmetros de qualidade da água e avaliar frequência de amostragem. Os autores propuseram um inventário das informações, abordagens e ferramentas colocadas à disposição dos gestores das bacias hidrográficas com o objetivo de iniciar uma discussão sobre como as informações, abordagens e ferramentas disponíveis poderiam ser integradas em uma solução mais holística e evolutiva em relação àquelas disponíveis atualmente.

Nas últimas décadas, houve uma demanda crescente por PMQA para rios de diferentes localidades, por meio de medições regulares de vários parâmetros de qualidade da água (MONICA; CHOI, 2016). Com base nos dados dos PMQA, os corpos de água de todo o mundo têm sido objeto de numerosos estudos com o objetivo de identificar as fontes de poluição, avaliar as variações espaciais e temporais e prever as tendências da qualidade da água (VOZA; VUKOVIĆ, 2018).

No Brasil, em todo o território, é realizado o monitoramento hidrológico para gerar informações, ao longo do tempo, sobre a quantidade e a qualidade dos recursos hídricos superficiais e subterrâneos (ANA, 2020a). Em 2010, a Agência Nacional de Águas e Saneamento Básico (ANA) lançou o Programa Nacional de Avaliação da Qualidade das Águas – PNQA, cujo objetivo é ampliar o conhecimento sobre a qualidade das águas superficiais, de forma a orientar a elaboração de políticas públicas para a recuperação da qualidade ambiental em corpos de água interiores como rios e reservatórios, contribuindo assim com a gestão sustentável dos recursos hídricos (ANA, 2010).

No âmbito estadual, o Instituto Mineiro de Gestão das Águas (Igam) é o órgão responsável pelo monitoramento da qualidade das águas superficiais e subterrâneas em Minas Gerais, do programa “Águas de Minas”, em execução desde 1997. Em vigência há vinte e quatro anos, o monitoramento é de extrema importância, especialmente, para o fornecimento de informações básicas necessárias para a definição de estratégias e a própria avaliação da efetividade do sistema de controle ambiental e para o planejamento e gestão integrada dos recursos hídricos (IGAM, 2019a).

Experiências do estado de Minas Gerais relacionadas a acidentes e desastres, nas últimas décadas, colocam ainda mais em evidência o quanto o monitoramento é

primordial para o acompanhamento da situação dos corpos de água, trazendo aos órgãos estaduais gestores das águas e do meio ambiente grandes desafios técnicos e institucionais para recuperação das áreas impactadas (IGAM, 2019b).

O monitoramento da qualidade das águas em situações emergenciais, como por exemplo, acidentes ambientais, possibilita a definição de estratégias que busquem a conservação, a recuperação e o uso racional dos recursos hídricos e a avaliação dos impactos ambientais, aferindo o grau de contaminação do recurso afetado e das fontes de poluição, assegurando o uso, reduzindo os conflitos e direcionando ações (IGAM, 2019b).

Os resultados gerados nos programas de monitoramento ambiental contêm informações importantes sobre as regiões de interesse, mas essas informações acabam sendo pouco exploradas devido às dificuldades encontradas em sua análise e interpretação estatística (CARVALHO, 2003). 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 quantidades de dados, o que aumenta a dificuldade. 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 apropriados para fins de monitoramento e diagnóstico (SONG et al., 2011; ZAVAREH; MAGGIONI, 2018).

Além disso, o problema ao se trabalhar com dados de qualidade de água é a dificuldade de reduzi-los a formas utilizáveis para a tomada de decisão (COURTEMANCH, 1994; ZIMMERMAN et al., 1996), consequência de características intrínsecas que impedem o uso de métodos estatísticos padrão. De maneira geral, o processo ambiental pode ser caracterizado como complexo, multifacetado, caótico e dinâmico (COTHERN; ROSS, 1993). De acordo com Bodo (1989) e Helsel e Hirsch (2002), os dados de recursos hídricos apresentam, em geral, as seguintes características: presença de *outliers*, assimetria positiva, distribuição não normal dos dados, dados censurados, padrões sazonais, autocorrelação e dependência de outras variáveis não controláveis.

A interpretação e análise dos dados de qualidade de água não são, portanto, simples e requerem 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.

Com uma quantidade excessiva de dados, é fácil perder informações e padrões ocultos (ZAVAREH; MAGGIONI, 2018). Dessa forma, várias abordagens têm sido empregadas em avaliações de qualidade de água para determinar a estrutura, classificar e modelar os conjuntos de dados (SIMEONOV et al., 2001) e para evitar interpretações errôneas (SIMEONOV et al., 2002). As abordagens mais utilizadas e referenciadas na bibliografia para avaliação de dados de qualidade de água são os métodos de modelagem numérica e as abordagens baseadas em dados (*data-driven approaches*) (DU et al., 2017).

Em relação aos métodos baseados em dados, os estudos para analisar dados de qualidade da água incluem: rede neural artificial (CHANG et al., 2015; DIAMANTOPOULOU; ANTONOPOULOS; PAPAMICHAIL, 2007; HE et al., 2011; PALANI; LIONG; TKALICH, 2008; SINGH et al., 2009; YAN; ZOU; WANG, 2010), lógica fuzzy (CHANG; CHEN; NING, 2001; ICAGA, 2007; LERMONTOV et al., 2009; OCAMPO-DUQUE et al., 2006; OCAMPO-DUQUE; SCHUHMACHER; DOMINGO, 2007), rede Bayesiana (FORIO et al., 2015; GRONEWOLD; BORSUK, 2010; MA et al., 2014; QIAN; MILTNER, 2015), algoritmos genéticos (PARK et al., 2006; SOTOMAYOR; HAMPEL; VÁZQUEZ, 2018; ZHANG; XIN, 2017); controle estatístico de processo univariado (ARCISZEWSKI et al., 2018; CRUZ; OLIVEIRA; CUNHA FILHO, 2016; FOLLADOR et al., 2012; IGLESIAS et al., 2016; LEE et al., 2013; SANCHO et al., 2016), regressão linear multivariada (SONG et al., 2011) e análise de múltiplos critérios (SCHUWIRTH et al., 2018).

Além desses diferentes métodos, destacam-se as técnicas multivariadas - Análise de Cluster Hierárquica Aglomerativa, Análise por Componentes Principais, Análise Fatorial e a Análise Discriminante - que também têm sido usadas em numerosos estudos sobre a qualidade da água (BARAKAT et al., 2016; KAZI et al., 2009; LI; XU; LI, 2009; MONICA; CHOI, 2016; OLSEN; CHAPPELL; LOFTIS, 2012; PATI et al., 2014; PINTO; CALAZANS; OLIVEIRA, 2019; SHRESTHA; KAZAMA, 2007; SINGH et al., 2004; VEGA et al., 1998; WUNDERLIN et al., 2001; ZHANG et al., 2011; ZHOU; LIU; GUO, 2007). De acordo com Jabbar e Grote (2019), a análise estatística multivariada tornou-se uma ferramenta poderosa para investigar e interpretar os resultados dos dados de qualidade de água.

Os estudos de monitoramento da qualidade da água e a utilização desses diferentes métodos também são de interesse da comunidade científica brasileira com diversos estudos nacionais publicados na literatura (OLIVEIRA; PINTO; SOARES, 2018). Em nível local, muitos estudos relacionados aos cursos de água do estado de Minas Gerais utilizam os dados disponibilizados pelo órgão gestor – Igam, como exemplificado no Quadro 1.1.

Os resultados desses estudos ratificam a importância do tema abordado e permitem, por exemplo, que os órgãos gestores de água imponham regulamentos apropriados sobre esse assunto e que a indústria aplique métodos mais eficientes de controle e tratamento de cargas residuárias.

Verifica-se, portanto, que o acesso por si só a uma grande quantidade de dados não garante a obtenção de informações relevantes para a tomada de decisão. O desafio atual consiste justamente em tratar esses dados para transformá-los em informações úteis às organizações, garantindo a obtenção de avaliações confiáveis da qualidade da água que irão refletir no estabelecimento de projetos eficazes de manejo e proteção ambiental para a prevenção e o controle da poluição da água.

Entre as abordagens de análise multivariada, a Análise por Componentes Principais (ACP) é provavelmente a mais popular e é amplamente utilizada em muitas disciplinas (ABDI; WILLIAMS, 2010), inclusive na área de qualidade de água (OLSEN; CHAPPELL; LOFTIS, 2012). Sua ideia central é reduzir a dimensionalidade de um conjunto de dados que possui muitas variáveis inter-relacionadas, preservando o máximo possível da variabilidade (ou seja, informações estatísticas) presente no conjunto de dados. Mediante a rotação do sistema de coordenadas, a ACP encontra novas variáveis, as componentes principais (CPs), que são funções lineares das variáveis originais. A composição exata das CPs forma o modelo baseado em dados, explicando o novo sistema de coordenadas (JOLLIFFE, 2002; JOLLIFFE; CADIMA, 2016). A ACP é utilizada com diferentes objetivos, por exemplo: redução de dimensionalidade, agrupamento de informações e visualização de estruturas e relações ocultas (BRO; SMILDE, 2014).

Quadro 1.1 - Estudos que utilizaram dados de monitoramento da qualidade da água do Programa Águas de Minas do Igam

Referência	Região de estudo	Objetivo do estudo
Dantas et al. (2020)	Bacia do rio São Francisco	Analisar os dados relacionados ao indicador de contaminação fecal para a porção mineira da bacia do rio São Francisco.
Soares; Pinto; Oliveira (2020)	Bacias dos rios Piracicaba e Paraopeba	Identificar os corpos de água mais impactados das bacias hidrográficas analisadas e calcular o risco relativo de violação aos limites preconizados pela legislação ambiental.
Magalhães Júnior; Lopes; Macedo (2019)	Bacia do rio Jequitinhonha	Relacionar a evolução do uso e ocupação da terra na bacia com os parâmetros de qualidade da água que compõem o Índice de Qualidade da Água (IQA), identificando as principais fontes de pressão sobre os recursos hídricos.
Oliveira et al. (2019)	Bacia do rio Pará	Avaliar e comparar a qualidade das águas superficiais das dez sub-bacias do rio Pará, além de verificar o impacto da sazonalidade e o atendimento aos limites da legislação estadual vigente.
Calazans et al. (2018a)	Bacia do rio Paraopeba	Avaliar e propor ajustes na rede de monitoramento de qualidade da água na bacia do rio Paraopeba, utilizando métodos estatísticos multivariados.
Costa et al. (2017)	Bacia do rio São Francisco	Avaliar a tendência do percentual de violação dos parâmetros de qualidade de água de todas as estações de monitoramento da bacia em um período de 14 anos, permitindo a identificação da variação da qualidade da água ao longo dos anos.
Oliveira; Maillard; Pinto (2016)	Bacia do rio das Velhas	Avaliar como o uso da terra e a cobertura do solo afetam as mudanças nas cargas de nitrogênio amoniacal total, nitrato, nitrito, fósforo total e oxigênio dissolvido no rio das Velhas.
Oliveira et al. (2017)	Médio São Francisco	Investigar o impacto da irrigação na qualidade das águas superficiais da região, utilizando fontes de dados secundários de monitoramento.
Pinto et al. (2017)	Bacia do rio Verde	Calcular as tendências temporais das cargas e concentrações de 11 parâmetros físicos, químicos e biológicos, em nove estações de monitoramento de qualidade da água, no período de 2008 a 2014.
Sabino; Lage; De Noronha (2017)	Baixo rio Grande	Avaliar estatisticamente a variação sazonal e temporal da qualidade das águas do córrego Gameleiras no ponto de monitoramento BG057, fazendo associação entre usos e ocupações do solo na região.
Silva; Faria; Moura (2017)	Bacia do rio Piracicaba	Aplicar um modelo de qualidade da água, o Sistema de Apoio à Decisão/Instituto de Pesquisas Hidráulicas (SAD-IPH).
Trindade et al. (2017)	Bacia do rio das Velhas	Análise da tendência temporal e espacial da qualidade das águas superficiais da sub-bacia, sendo analisados 16.625 dados coletados no período de 2002 a 2011.
Rocha (2016)	Bacia do rio Piranga	Verificar se a atividade suinícola, presente na Unidade de Planejamento e Gestão de Recursos Hídricos - UPGRH DO1, está alterando a qualidade das águas superficiais das sub-bacias hidrográficas que a compõem.
Jardim et al. (2014)	Bacia do rio Doce	Associar estatisticamente parâmetros ambientais e de qualidade da água do rio, para encontrar e explicar as possíveis correlações existentes nas sucessivas florações de cianobactérias.
Maillard; Santos (2008)	Bacia do rio das Velhas	Descrever e testar uma metodologia para avaliar o efeito do uso/cobertura do solo na qualidade da água de um córrego próximo a um ambiente semiárido.

Fonte: Próprio autor.

A maioria dos estudos, relacionado à qualidade de água, foi dedicada a explorar a aplicabilidade de ACP com o foco, principalmente, em *loads*, considerando análises de conjuntos de estações de monitoramento. Os objetivos são, geralmente, de agrupar estações de monitoramento; associar as componentes principais a fontes de contaminação, tais como fator poluição orgânica, salinidade, agricultura, lixiviação do solo, entre outros; e identificar parâmetros com variações espaço-temporal (BENGRÁINE; MARHABA, 2003; BU et al., 2010; PEREIRA et al., 2019; PERONA; BONILLA; MATEO, 1999; RAZMKHAH; ABRISHAMCHI; TORKIAN, 2010; SIMEONOV et al., 2003).

Neste trabalho, a ideia central é explorar as aplicações de ACP considerando três objetivos diferentes: identificação de condições críticas de qualidade da água via *scores*, utilização das estatísticas T^2 e Q para monitoramento da qualidade, e avaliação de variações temporais dos dados. A partir das revisões bibliográficas realizadas, verificou-se que a ACP é aplicada para interpretação de bancos de dados considerando as informações de várias estações de monitoramento simultaneamente, localizadas em uma região, sub-bacia ou curso de água. Neste estudo, as metodologias desenvolvidas foram aplicadas aos dados de uma estação de monitoramento, dando ênfase a uma abordagem local.

1.2 JUSTIFICATIVA

Os impactos causados na qualidade da água, principalmente, pelos lançamentos de efluentes domésticos e industriais e pelas mudanças no uso e ocupação do solo, aumentam as incertezas sobre a disponibilidade de água com qualidade adequada no futuro.

Nas últimas décadas, o desenvolvimento sustentável e a busca pela preservação do meio ambiente tornaram-se objeto de preocupação. Tendo em vista que a deterioração da qualidade da água dos rios pode ser considerada um fator importante para impedir o desenvolvimento social e econômico, avaliar a qualidade de água com precisão é fundamental tanto para a sociedade quanto para a economia. Na área de qualidade de água, muitos estudos ambientais vêm sendo realizados a partir de dados secundários gerados em programas de monitoramento, dando maior destaque e relevância a essas informações.

O monitoramento da qualidade das águas superficiais permite o conhecimento da situação dos corpos hídricos, as tendências de curto e longo prazos e o comprometimento da qualidade pela presença de poluentes. Possibilita, também, a verificação de riscos à saúde humana e animal, bem como do comprometimento dos usos múltiplos.

Os resultados do monitoramento correspondem à mensuração de vários parâmetros de qualidade da água, monitorados em diversas estações de amostragem por vários anos, e consistem em um extenso e complexo banco de dados, de difícil interpretação. Informações importantes podem não ser percebidas caso não se efetue uma adequada avaliação e tratamento desse conjunto de dados.

Ao analisar-se cada variável individualmente, geralmente, a identificação de padrões ambientais é mais complicada e não leva em consideração a estrutura de correlação existente entre as variáveis. Além disso, reconhece-se a necessidade de avaliações pontuais, ou seja, metodologias que possam ser aplicadas por estação de monitoramento. As avaliações considerando diversas estações de monitoramento de uma bacia hidrográfica são importantes, mas, às vezes, não são suficientes para descrever com detalhes as condições de qualidade de água de uma região específica. As análises locais permitem uma compreensão mais detalhada da qualidade da água e das fontes de poluição. A associação da abordagem local e da avaliação global podem ser mais eficazes para subsidiar a elaboração de políticas públicas voltadas para a revitalização e preservação dos cursos d'água e saneamento ambiental.

Nesta tese, a estação de monitoramento utilizada como estudo de caso – estação BV137 localizada na calha do rio das Velhas, a jusante do município de Lagoa Santa – foi identificada em estudos anteriores (CALAZANS et al., 2018b; CORDOVA, 2020) que envolveram uma avaliação global da bacia do rio das Velhas. A partir dos estudos desses autores, a estação BV137 foi identificada como uma das estações mais impactadas e prioritária para manutenção na rede de monitoramento.

Este trabalho justifica-se, portanto, permitindo um melhor entendimento da aplicação de ACP, uma das técnicas estatísticas multivariadas, a dados de qualidade de água, considerando diferentes objetivos e uma abordagem local. Ainda, espera-se contribuir com o conhecimento sobre fontes de poluição, variações temporais e monitoramento das condições de qualidade da água, fornecendo aos órgãos gestores ferramentas que possam ser utilizadas para avaliações e tomadas de decisão.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Aplicar a técnica multivariada de Análise por Componentes Principais considerando três abordagens distintas, para avaliar dados de monitoramento de qualidade de águas superficiais, por meio de uma abordagem local, para suporte à tomada de decisão.

1.3.2 Objetivos Específicos

- Realizar o pré-processamento dos dados;
- Desenvolver um modelo baseado em Análise por Componentes Principais para mapeamento de condições críticas de qualidade da água e identificação de fontes de poluição;
 - Aplicar Controle Estatístico de Processo Multivariado baseado em Análise por Componentes Principais para o monitoramento da qualidade da água, identificando condições críticas de degradação ambiental;
 - Aplicar Análise por Componentes Principais para avaliar variações temporais na qualidade da água superficial, em uma estação de monitoramento do rio das Velhas.

CAPÍTULO 2:
REVISÃO TEÓRICA DA ANÁLISE
POR COMPONENTES PRINCIPAIS

Na literatura existem várias referências sobre Análise por Componentes Principais (*Principal Components Analysis* - PCA) com detalhes matemáticos das abordagens analíticas e geométricas (JACKSON, 2003; JOHNSON; WICHERN, 2007; JOLLIFFE, 2002; LEGENDRE; LEGENDRE, 2012; SHARMA, 1996), além de referências com abordagens mais simples e práticas (DAVIS, 2002; MANLY, 2008; SHAW, 2003). Neste capítulo da tese, é apresentado um resumo dos principais aspectos e pontos teóricos para a compreensão da técnica.

A ACP é uma das estratégias quimiométricas mais comuns para análise exploratória não supervisionada de dados (BALLABIO, 2015), cujos objetivos são: (1) extrair as informações mais importantes do conjunto de dados; (2) reduzir o tamanho do conjunto de dados mantendo apenas essas informações importantes; (3) simplificar a descrição do conjunto de dados; e (4) analisar a estrutura das observações e das variáveis (ABDI; WILLIAMS, 2010).

A ACP é uma técnica multivariada, introduzida por Pearson (1901) e desenvolvida por Hotelling (1933), com o objetivo de reduzir a dimensionalidade do conjunto de dados original, preservando a maior quantidade de informação (variância) possível (LATTIN; CARROLL; GREEN, 2011). Resumidamente, a ACP transforma um conjunto de dados contendo p variáveis, inter-relacionadas ou correlacionadas entre si, em um novo conjunto de dados contendo p novas variáveis ortogonais não correlacionadas, denominadas componentes principais (CPs) (JOHNSON; WICHERN, 2007).

Os dados multivariados podem ser organizados em uma matriz $X_{[n,p]}$, em que n é o número de observações por variável e p é o número de variáveis (Equação 2.1).

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.1)$$

Então, as p variáveis podem ser expressas na forma vetorial, $\mathbf{X} = [X_1 X_2 \dots X_p]'$, com vetor de médias $\boldsymbol{\mu} = [\mu_1 \mu_2 \dots \mu_p]'$. Antes da modelagem ACP, as variáveis são normalmente padronizadas (Equação 2.2), pois as variâncias diferem consideravelmente em termos absolutos.

$$z_{ij} = (x_{ij} - \mu_j) / \sigma_j, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (2.2)$$

Na Equação 2.2, x_{ij} é a observação da i -ésima amostra da j -ésima variável; μ_j e σ_j são a média e o desvio-padrão da j -ésima variável, respectivamente; e z_{ij} é o valor padronizado correspondente a x_{ij} . Essa transformação reduz todas as variáveis à média zero e variância unitária. A matriz resultante $Z_{[n,p]}$ é a versão padronizada da matriz $X_{[n,p]}$.

A matriz de variância e covariância $\Sigma_{p \times p}$ é apresentada na Equação 2.3. A covariância é a medida do grau de relacionamento linear entre duas variáveis aleatórias.

$$\Sigma_{p \times p} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \quad (2.3)$$

Sendo a matriz de covariância uma matriz quadrada, podem-se encontrar os autovetores e autovalores associados a ela (Equação 2.4).

$$W' \Sigma W = \Lambda \quad (2.4)$$

As colunas da matriz $W_{[p,p]}$ são os autovetores de Σ , e a matriz diagonal $\Lambda_{[p,p]}$, contém os autovalores de Σ na diagonal principal.

Cada autovalor $(\lambda_1, \lambda_2, \dots, \lambda_p)$ está associado a um autovetor (w_1, w_2, \dots, w_p) . Podem-se, então, ordenar os autovetores por ordem de importância, ou seja, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_p$. O autovetor com o maior autovalor contém a maior quantidade de informação, em relação à variância total dos dados originais. Assim, a i -ésima componente principal é definida na Equação 2.5.

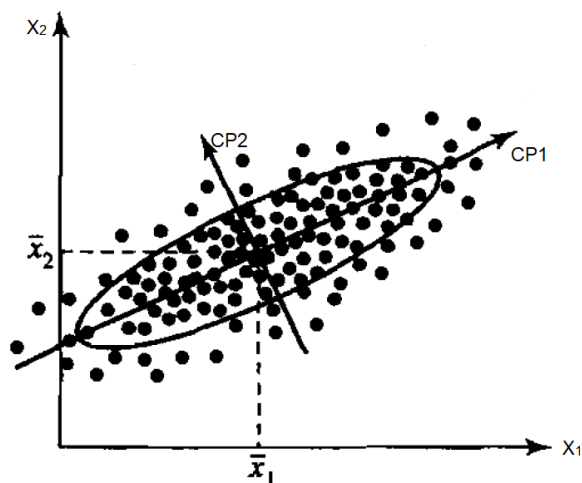
$$CP_i = \sum_{j=1}^p w_{ij} z_{ij} \quad (2.5)$$

$$i = 1, 2, \dots, p; j = 1, 2, \dots, p$$

As componentes principais (CP_i) são combinações lineares das variáveis originais, conforme a Equação 2.5, em que w_{ij} é o peso associado à j -ésima variável na i -ésima componente principal. Pode-se considerar o peso como uma medida da influência da variável na componente principal. Para distinguir entre as variáveis transformadas e a observação transformada, as variáveis transformadas são chamadas de componentes principais e as observações transformadas individuais são chamadas de *scores*. Em outras palavras, os *scores* são as coordenadas dos pontos originais no novo sistema (rotacionado) de coordenadas (JOHNSON; WICHERN, 2007; LATTIN; CARROLL; GREEN, 2011; SHARMA, 1996).

Geometricamente, as componentes principais são os eixos do novo sistema de coordenadas obtidos pela rotação do sistema original (Figura 2.1). Os novos eixos representam as direções de variabilidade máxima (JOHNSON; WICHERN, 2007).

Figura 2.1 - Representação geométrica das componentes principais



Fonte: Adaptado de Johnson e Wichern (2007).

A ACP fornece tantas componentes quanto for o número de variáveis, ou seja, se existem p variáveis, então existirão p CPs.

O traço de uma matriz quadrada qualquer de ordem " n " é definido como a soma dos termos da diagonal principal. Dessa forma, o traço da matriz de covariância Σ , apresentada nas Equações 2.6 e 2.7, é:

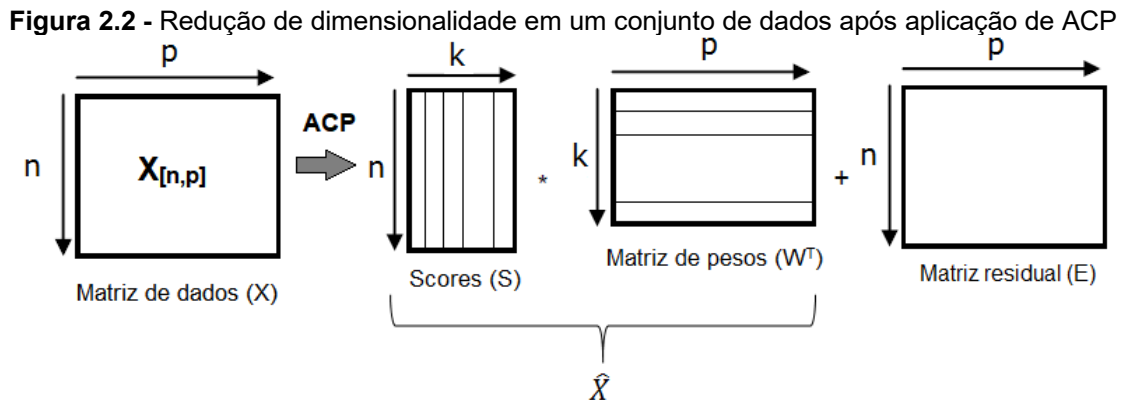
$$\text{Traço}(\Sigma) = \text{traço}(W \Lambda W') = \text{traço}(\Lambda W'W) = \text{traço}(\Lambda I) = \text{traço}(\Lambda) = \sum_{i=1}^p \lambda_i \quad (2.6)$$

E,

$$\text{Traço}(\Sigma) = \sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i \quad (2.7)$$

Portanto, a variabilidade total contida nas variáveis originais é igual a variabilidade total contida nas componentes principais (JOHNSON; WICHERN, 2007).

A partir das equações apresentadas, observa-se que após aplicação da ACP, a matriz de dados X é dividida em duas partes (Figura 2.2): uma descreve a variação do sistema (parte sistemática que é conformada por *scores* (S) e *loadings* (W) - o modelo do processo ($S_k W_k^T$); e a outra captura o ruído ou a informação não modelada (variância residual E) (KOURTI, 2002).



Fonte: Adaptado de Bohm; Smidt; Tintner (2013).

As observações da matriz X podem ser projetadas em um espaço dimensional menor. Essas projeções (*scores*) estão contidas na matriz S . A matriz projetada S - ou vetores de *scores* (*score vectors*) - no novo espaço é definida na Equação 2.8.

$$S_{[n,k]} = X_{[n,p]} W_{[p,k]} \quad (2.8)$$

E a projeção de S de volta ao espaço de observação p -dimensional (Equação 2.9) é:

$$\hat{X}_{[n,p]} = S_{[n,k]} W_{[k,p]}^T \quad (2.9)$$

Onde a diferença entre X e \hat{X} é a matriz residual E (Equações 2.10 e 2.11):

$$E_{[n,p]} = X_{[n,p]} - \hat{X}_{[n,p]} \quad (2.10)$$

$$X_{[n,p]} = \hat{X}_{[n,p]} + E_{[n,p]} = S_{[n,k]} W_{[k,p]}^T + E_{[n,p]} \quad (2.11)$$

Desse modo, dada a matriz de dados originais, $X_{[n,p]}$, busca-se um novo sistema de coordenadas onde k dimensões (componentes principais, ortogonais entre si) expliquem a máxima informação (variância) das p dimensões (variáveis) originais, com $k \ll p$ (MINGOTI, 2005). Isso significa que a matriz de autovetores ($W_{[p,p]}$) pode ter sua dimensão reduzida para uma quantidade k de colunas ($W_{[p,k]}$). Essa redução de dimensões torna a visualização dos dados mais direta e facilita a sua análise.

Diferentes métodos podem ser utilizados para calcular o número de CPs, sendo que a maioria deles utiliza índices monotonicamente crescentes ou decrescentes. Dessa forma, a decisão de escolher o número de componentes principais é muito subjetiva (VALLE; LI; QIN, 1999). Nesta revisão, são apresentados três métodos: critério baseado no percentual de variância explicada, critério de Kaiser e critério do teste “Scree” (HAIR et al., 2009; LATTIN; CARROLL; GREEN, 2011; MINGOTI, 2005).

Os autovalores associados a cada autovetor ou componente principal expressam quanta informação (variação) cada CP explica (Equação 2.12). Então, o critério baseado no percentual de variância explicada consiste em observar a variância acumulada percentual capturada pelas primeiras CPs e escolher um número de CPs que respondem por uma determinada fração especificada da variância total dos dados.

$$\gamma \text{ (percentual da variância explicada) } = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \quad (2.12)$$

Já o critério de Kaiser (ou critério da raiz latente) descreve que o número de componentes retidas deve ser igual ao número de autovalores que excedam a unidade, considerando que as variáveis da análise estão padronizadas.

Outro critério é a abordagem do teste “Scree” que compreende o desenho de um gráfico da variância explicada (autovalor) em relação a cada componente principal, na ordem do maior autovalor para o menor. A partir do gráfico é escolhido um ponto após o qual os autovalores remanescentes declinam de modo aproximadamente

linear, e retêm-se somente aquelas componentes que estão acima desse ponto (LATTIN; CARROLL; GREEN, 2011).

Para extrair informações efetivamente relevantes, antes da aplicação da ACP ao conjunto de dados, geralmente, é necessário fazer um pré-tratamento dos dados (RUSSELL; CHIANG; BRAATZ, 2000). O Quadro 2.1 apresenta os principais aspectos característicos dos conjuntos de dados ambientais e conceitos necessários para o pré-tratamento do banco de dados.

Quadro 2.1 - Aspectos característicos dos conjuntos de dados ambientais e conceitos necessários para a preparação do banco de dados antes da aplicação de técnicas estatísticas

Análises	Conceito	Descrição	Tratamento/Cálculos
Dados faltantes	Valores não computados quando uma coleta é realizada.	Os dados faltantes podem levar à eliminação de outras observações, e uma amostra que era adequada fica reduzida a uma amostra imprópria para a aplicação dos testes estatísticos (HAIR et al., 2009).	- Eliminar variáveis que possuam grande quantidade de dados faltantes; - Eliminar amostras que possuam dados faltantes;
Dados censurados	Valores acima ou abaixo do Limite de Detecção (LD) do método analítico.	A presença de dados censurados interfere em diversas análises estatísticas (por exemplo, parâmetros estatísticos básicos, testes de hipóteses, correlações e análises de tendências) (CHRISTOFARO; LEÃO, 2014).	A prática mais comumente adotada em estudos ambientais é o da substituição destes dados por um valor correspondente a uma fração do limite de detecção (por exemplo, metade do valor do LD ou o próprio valor do LD) (HELSEL, 2011).
<i>Outliers</i>	Observações atípicas que possuem características notavelmente diferentes das demais.	As observações atípicas devem ser vistas no contexto da análise e avaliadas pela informação que possam oferecer (HAIR et al., 2009). Dessa forma, após sua detecção o analista de dados é o responsável por decidir como os <i>outliers</i> devem ser tratados (BRO; SMILDE, 2014).	Existem diversos métodos estatísticos para detecção de <i>outliers</i> , por exemplo, teste de valor extremo, teste de discordância, teste de Rosner e teste de Walsh (EPA, 2006).
Distribuição de frequência dos dados	Existem basicamente dois tipos de distribuição de frequência: simétrica e assimétrica.	Antes da escolha do teste estatístico, é necessário examinar a distribuição dos dados. Para dados que apresentam uma distribuição assimétrica, deve-se escolher um teste não paramétrico ou transformar os dados para que eles possam apresentar uma distribuição normal (SHEATS; PANKRATZ, 2002).	Testes de normalidade de Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, etc (GILBERT, 1987).
Padronização dos dados	Transformação do conjunto original dos dados de modo que cada variável apresente média zero e variância igual a um.	As diferenças de escalas entre as variáveis podem influenciar as análises. Assim, a padronização garante que os dados sejam expressos em unidades comparáveis (LATTIN; CARROLL; GREEN, 2011).	Transformação das variáveis em escala Z, em que $Z=(X - \mu)/\sigma$ (MCBRIDE, 2005).

A partir do Quadro 2.1, observa-se que o pré-tratamento dos dados pode ser composto por diferentes análises e o procedimento apropriado depende normalmente da natureza dos dados investigados (BRO; SMILDE, 2014).

O estudo de Bro e Smilde (2014) apresenta uma descrição de como entender, usar e interpretar a ACP. Os autores destacam, sobretudo, a versatilidade dessa ferramenta, podendo ser aplicada para diferentes objetivos, como: identificação de relações entre variáveis e relações entre amostras (por exemplo, *clustering*), detecção de *outliers*, detecção e quantificação de padrões, modelagem, etapas de pré-processamento, entre outros.

Em relação à aplicação de ACP na área de modelagem, o modelo de ACP de uma matriz de dados inclui: vetores de média e variância, autovalores, *loadings* e limites estatísticos para os *scores* (T^2 e Q) (GALLAGHER et al., 1997).

A estatística T^2 de Hotelling (Equação 2.13) mede a variação dentro do modelo de ACP e é calculada utilizando os *scores* relativos às k primeiras componentes principais.

$$T_i^2 = \sum_{l=1}^k \frac{s_{il}^2}{\lambda_l} \quad (2.13)$$

Onde T_i^2 e s_{il} são os valores de T^2 de Hotelling e os valores dos *scores* na i -ésima amostra na l -ésima componente principal, respectivamente.

Já a estatística Q (Equação 2.14) mede a variação fora do modelo de ACP, ou seja, ela fornece informações adicionais que não estão incluídas no gráfico de *scores* (RIUS; CALLAO; RIUS, 1997; WISE et al., 1999).

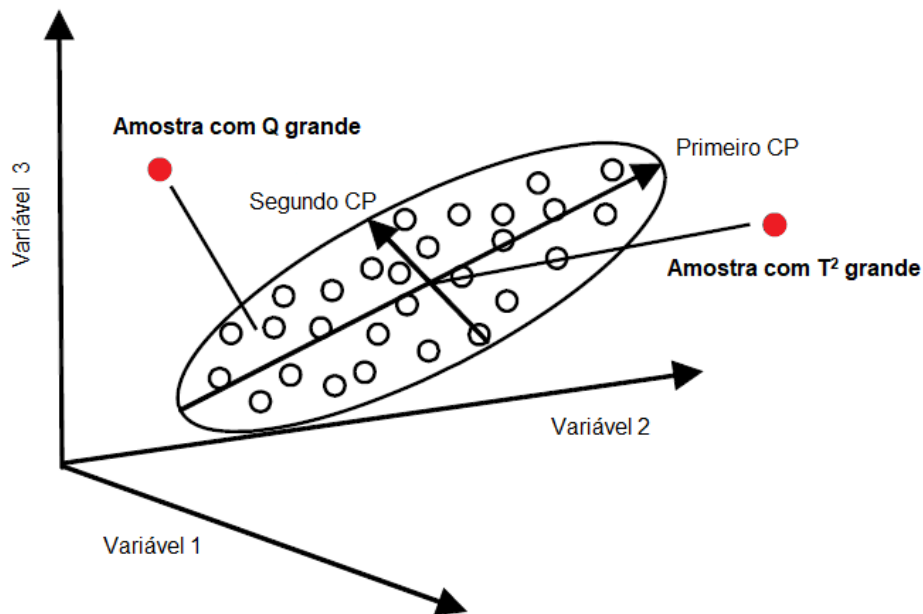
$$Q_i = e_i e_i^T = x_i (I - W_{[p \times k]} W_{[k \times p]}^T) x_i^T \quad (2.14)$$

Onde e_i é a i -ésima amostra da matriz residual E , I é a matriz identidade, W é a matriz de autovetores e x_i é a i -ésima amostra da matriz de dados X .

As medições de três variáveis de um processo e um modelo descrito por duas CPs, são mostrados na Figura 2.3. Observam-se duas amostras em destaque: uma com um valor Q incomumente grande e a outra com um alto valor de T^2 . A estatística Q é uma medida da distância do ponto de operação ao plano formado pelas duas

primeiras CPs, sendo \sqrt{Q} a distância euclidiana. O limite Q define uma distância do plano que é considerada incomum para condições operacionais normais. Por outro lado, a estatística T^2 é uma medida da distância da projeção do ponto de operação no plano definido pelas duas CPs até a origem desse plano. O limite T^2 define uma elipse no plano dentro do qual o ponto operacional normalmente se projeta (WISE et al., 1999).

Figura 2.3 - Modelo de ACP do conjunto de dados tridimensionais com destaque para valores *outliers* das estatísticas T^2 e Q



Fonte: Wise et al. (1999).

Ambas as estatísticas descritas (T^2 e Q) são monitoradas quando utiliza-se Controle Estatístico de Processo Multivariado baseado em ACP (CEPM-ACP) (FERRER, 2007; GE; SONG, 2013).

De maneira geral, a partir dos dados de calibração, as estatísticas T^2 e Q são calculadas e os limites de controle, em um determinado nível de confiança, podem ser estabelecidos nos gráficos. Posteriormente, novos dados são monitorados usando esses limites e anomalias são detectadas quando os limites são excedidos (FERRER, 2007). Os gráficos de T^2 e Q são um teste de hipótese que sinalizam claramente qualquer amostra fora de controle, enquanto a inspeção do gráfico de *scores* é uma ferramenta qualitativa (RIUS; CALLAO; RIUS, 1997).

Além disso, devido à natureza dos modelos de variáveis latentes, a contribuição das variáveis originais para a estatística calculada, em uma dada amostra, pode ser

investigada a partir dos gráficos de contribuição (*contribution plots*) (ALCALA; QIN, 2011; KOURTI; MACGREGOR, 1996; NOMIKOS, 1996; QIN, 2003; WESTERHUIS; GURDEN; SMILDE, 2000).

Um gráfico de contribuição indica quais variáveis fazem a amostra selecionada apresentar um valor significativo para as estatísticas T^2 de Hotelling ou Q . Uma contribuição alta de uma variável geralmente indica um problema com essa variável específica (WESTERHUIS; GURDEN; SMILDE, 2000). Assim, é possível identificar quais variáveis são as mais relevantes quando uma anomalia/falha for detectada.

Como existem diferentes maneiras de decompor as estatísticas calculadas, vários autores propuseram fórmulas diferentes para o cálculo de contribuição (VAN DEN KERKHOF et al., 2013). Nesta tese, são apresentadas as equações dos estudos de Ballabio (2015) e Westerhuis, Gurden e Smilde (2000) para calcular as contribuições individuais das variáveis originais para as estatísticas T^2 e Q , respectivamente.

Em relação à estatística Q , o cálculo das contribuições das variáveis é realizado com base no espaço residual. Em um gráfico da estatística Q , uma amostra acima dos limites representa uma perturbação/falha que não foi descrita pelo modelo de processo. Dessa forma, a Equação 2.15 pode ser utilizada para calcular a contribuição da j -ésima variável no resíduo Q da i -ésima observação.

$$C_{[i,j]}^Q = \frac{(X_{p[i,j]} - X_{p[i,j]}^*)^2}{Q_i} \quad (2.15)$$

Onde $X_{p[i,j]}$ é o valor padronizado da variável j na observação i , $X_{p[i,j]}^*$ é o valor estimado pelo modelo de ACP, e Q_i é o valor da estatística Q correspondente à i -ésima observação (WESTERHUIS; GURDEN; SMILDE, 2000). Quanto maior o valor de $C_{[i,j]}^Q$, maior é a importância relativa daquela variável j sobre o desvio da observação i .

Assim como na estatística Q , as contribuições de cada variável para um valor significativo da estatística T^2 também podem ser calculadas (Equação 2.16).

$$C_{[i,j]}^{T^2} = \sum_{l=1}^k \frac{s_{i,l} \cdot w_{j,l}}{\sqrt{\lambda_l}} \quad (2.16)$$

$$j = 1, 2, \dots, p; i = 1, 2, \dots, n; l = 1, 2, \dots, k$$

Onde $C_{[i,j]}^{T^2}$ é a contribuição da j -ésima variável na estatística T^2 da i -ésima amostra, $s_{i,l}$ é o valor do score da i -ésima amostra na l -ésima componente, $w_{j,l}$ é o peso da j -ésima variável na l -ésima componente, e λ_l é o autovalor associado à l -ésima componente (BALLABIO, 2015).

A partir da revisão apresentada, é possível verificar como a ACP é uma excelente técnica exploratória de dados multivariados (MINGOTI, 2005), sendo um método poderoso e versátil capaz de fornecer uma visão geral de dados multivariados complexos (BRO; SMILDE, 2014).

**CAPÍTULO 3:
DEFINIÇÃO DO ESTUDO DE
CASO E PRÉ-PROCESSAMENTO
DOS DADOS DE
MONITORAMENTO**

Este capítulo aborda a definição do estudo de caso, com a descrição da localização da bacia hidrográfica, de suas principais características e estações de monitoramento, e o detalhamento dos bancos de dados que foram utilizados na pesquisa (dados de qualidade da água, fluviométricos e pluviométricos). A metodologia e os resultados e discussões, referentes à organização e pré-tratamento desses bancos de dados, também são apresentados nesse capítulo. Destaca-se que os bancos de dados fluviométricos e pluviométricos foram utilizados somente no capítulo 6.

3.1 ÁREA DE ESTUDO

3.1.1 Localização e características gerais

A bacia hidrográfica do rio São Francisco corresponde a 8% do território nacional e é uma bacia importante para o país, principalmente devido ao volume de água transportado em uma região semiárida, pelo potencial hídrico passível de aproveitamento e por sua contribuição histórica e econômica para a região (CBHSF, 2020, 2016).

Na região hidrográfica do rio São Francisco, os principais fatores de pressão sobre a qualidade e quantidade de água são o desmatamento, a urbanização, a expansão da indústria e a mecanização na agricultura. Dentre os problemas identificados, alguns são específicos ou predominantes de determinadas regiões e outros ocorrem em toda a sua extensão (MMA, 2009).

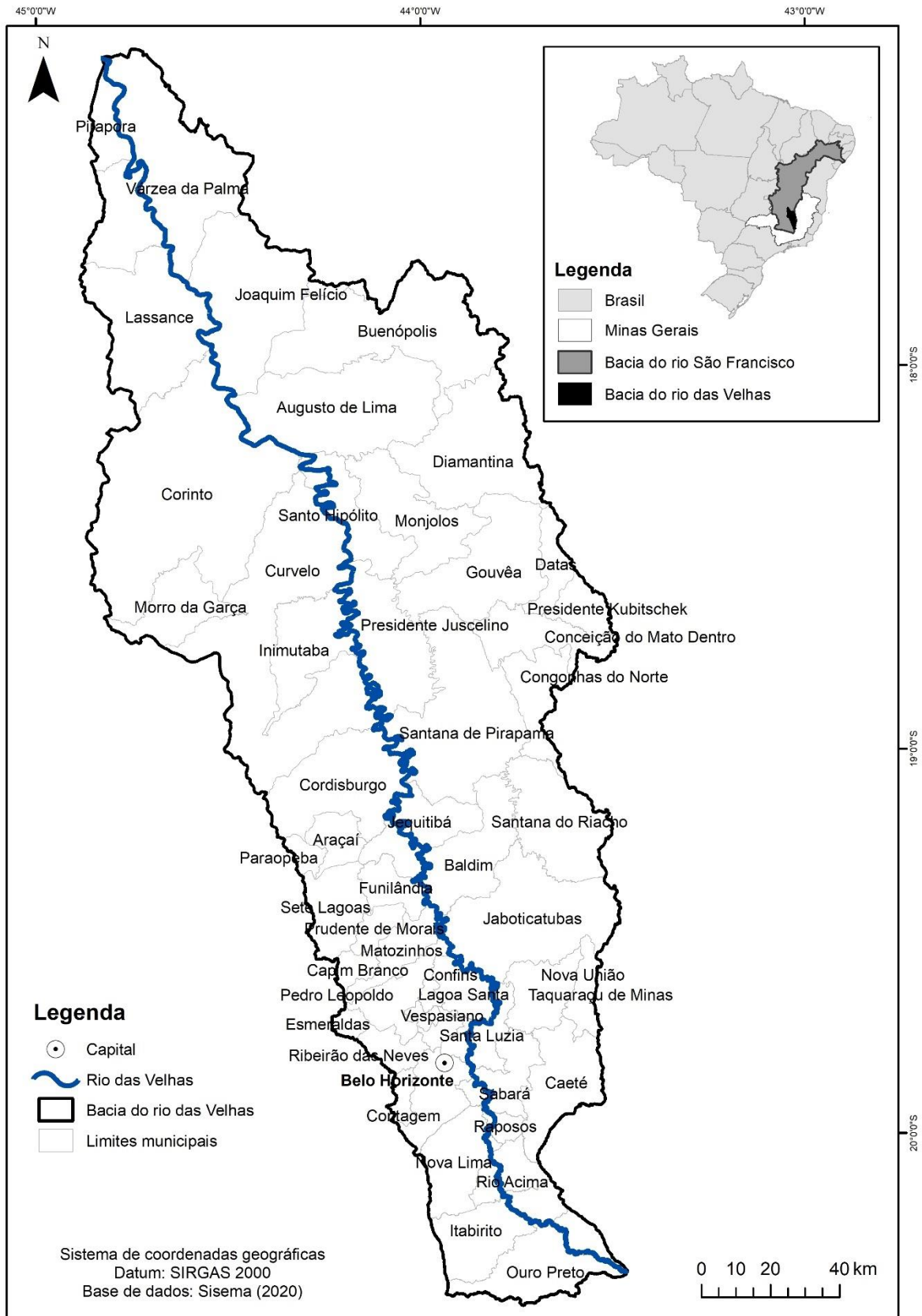
De acordo com Costa et al. (2017) e Trindade et al. (2017) a sub-bacia do rio das Velhas, localizada na região do Alto São Francisco (Figura 3.1), é a mais degradada em relação à qualidade das águas superficiais, sendo, portanto, uma das principais responsáveis pela deterioração da qualidade das águas superficiais do rio São Francisco. De acordo com o diagnóstico do Comitê da Bacia Hidrográfica do Rio São Francisco - CBHSF (2016), no Alto São Francisco, o rio das Velhas é o curso de água que apresenta a pior condição de qualidade da água, inclusive com contaminação por substâncias tóxicas.

A sub-bacia hidrográfica do rio das Velhas está localizada na região central do Estado de Minas Gerais e corresponde à Circunscrição Hidrográfica (CH) SF5 (CBH VELHAS, 2015a). Essa sub-bacia possui o maior afluente em extensão da bacia do

São Francisco - o rio das Velhas - com 806,84 km (CBH VELHAS, 2019) e nascente principal no município de Ouro Preto (região do quadrilátero ferrífero), com passagem pela Região Metropolitana de Belo Horizonte (RMBH) (CBH VELHAS, 2015a). Toda a sub-bacia compreende uma área de 27.850 km² e população de aproximadamente 4,4 milhões de habitantes (CBH VELHAS, 2015a). A RMBH ocupa apenas 10% da área territorial da bacia, mas possui mais de 70% de toda a sua população (CBH VELHAS, 2019).

Na sub-bacia do rio das Velhas, os usos da água variam desde a diluição de efluentes domésticos e industriais a outros mais nobres, tais como abastecimento de água, preservação ambiental, pesca e recreação de contato primário (CBH VELHAS, 2015a).

Figura 3.1 - Localização geográfica da bacia hidrográfica do rio das Velhas e seus municípios

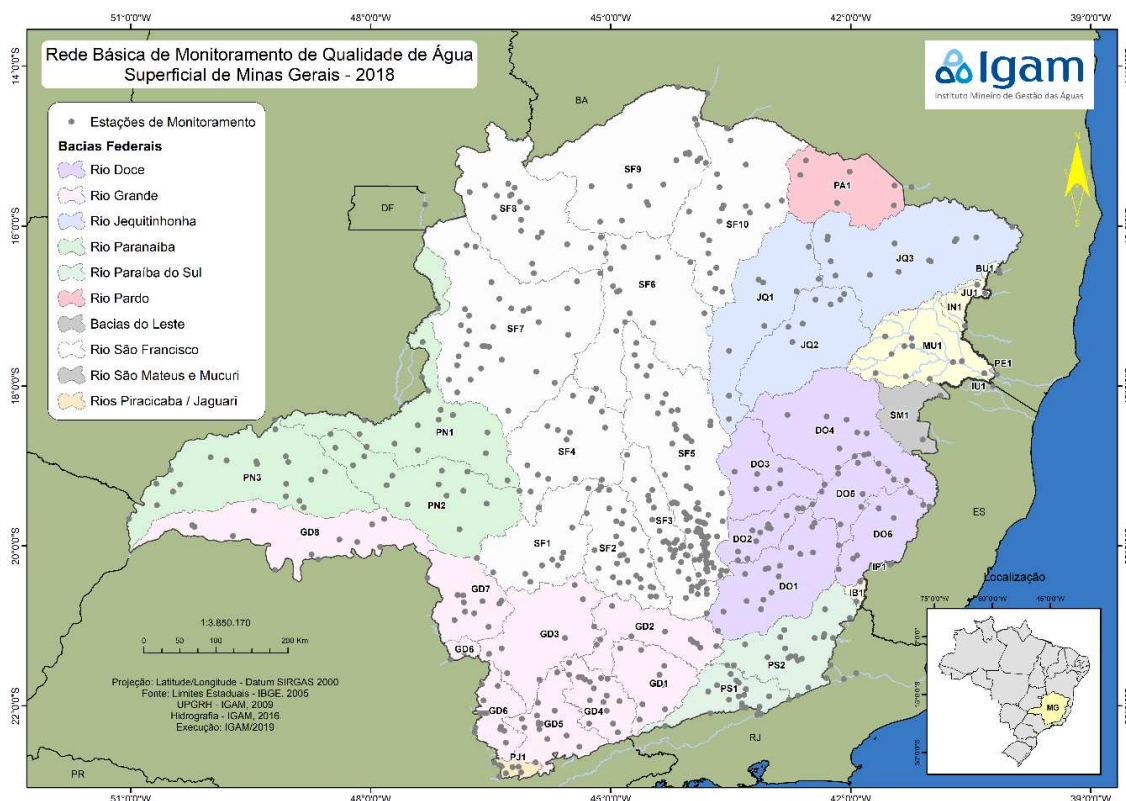


Fonte: Próprio autor.

3.1.2 Monitoramento da qualidade das águas superficiais em Minas Gerais

No estado de Minas Gerais, a rede básica de monitoramento de qualidade de água do Igam possuía, em 2018, 620 estações de amostragem (IGAM, 2019a). A distribuição geográfica dessas estações é apresentada na Figura 3.2.

Figura 3.2 - Pontos de monitoramento de qualidade da água superficial da rede básica do Programa Águas de Minas em operação no ano de 2018



Fonte: Igam (2019a)

As campanhas de amostragens realizadas pelo Igam, para a maioria das estações de monitoramento, são trimestrais. A exceção é para as estações localizadas nas calhas dos rios das Velhas e Doce, nas quais as campanhas são mensais. Os parâmetros de qualidade de água analisados são apresentados no Quadro 3.1. Dois tipos de campanhas são realizadas: as completas, executadas no primeiro e terceiro trimestres do ano, nas quais são monitorados 51 parâmetros comuns ao conjunto de pontos de amostragem; e as campanhas intermediárias, realizadas no segundo e quarto trimestres, quando são analisados 19 parâmetros genéricos em todos os pontos, além daqueles característicos das fontes poluidoras que contribuem para a área de drenagem da estação de coleta (IGAM, 2019a).

Quadro 3.1 - Parâmetros de qualidade da água analisados nas estações de amostragem do Igam, no Programa Águas de Minas.

Alcalinidade bicarbonato	Densidade de cianobactérias**	pH <i>in loco</i> *
Alcalinidade total	Dureza (cálcio)	Potássio
Alumínio dissolvido	Dureza (magnésio)	Selênio total
Arsênio total	Dureza total	Sólidos dissolvidos*
Bário total	Ensaio de Toxicidade Crônica**	Sólidos em suspensão*
Boro total	Estanho total	Sólidos totais*
Cádmio total	Fenóis Totais	Sólidos sedimentáveis
Cálcio	Feofitina*	Substâncias tensoativas
Chumbo total	Ferro Dissolvido	Sulfatos
Cianeto livre	Fósforo Total*	Sulfetos
Cianotoxinas**	Macroinvertebrados bentônicos**	Temperatura da água/ar*
Cloreto total*	Magnésio Total	Transparência da água
Clorofila <i>a</i> *	Manganês Total	Turbidez*
Cobre dissolvido	Mercúrio Total	Zinco Total
Coliformes termotolerantes/ <i>E. coli</i> ¹ *	Níquel Total	COT em sedimentos
Coliformes totais*	Nitrato*	
Condutividade elétrica <i>in loco</i> *	Nitrito	
Cor verdadeira	Nitrogênio Amoniacal Total*	
Cromo total	Nitrogênio Orgânico	
Demanda Bioquímica de Oxigênio*	Óleos e Graxas	
Demanda Química de Oxigênio*	Oxigênio Dissolvido – OD*	

* Parâmetros comuns a todos os pontos nas campanhas intermediárias

** Parâmetros analisados apenas em pontos específicos

¹ O parâmetro *Escherichia coli* passou a ser avaliado em substituição aos coliformes termotolerantes, a partir da primeira campanha de 2013 (IGAM, 2019a)

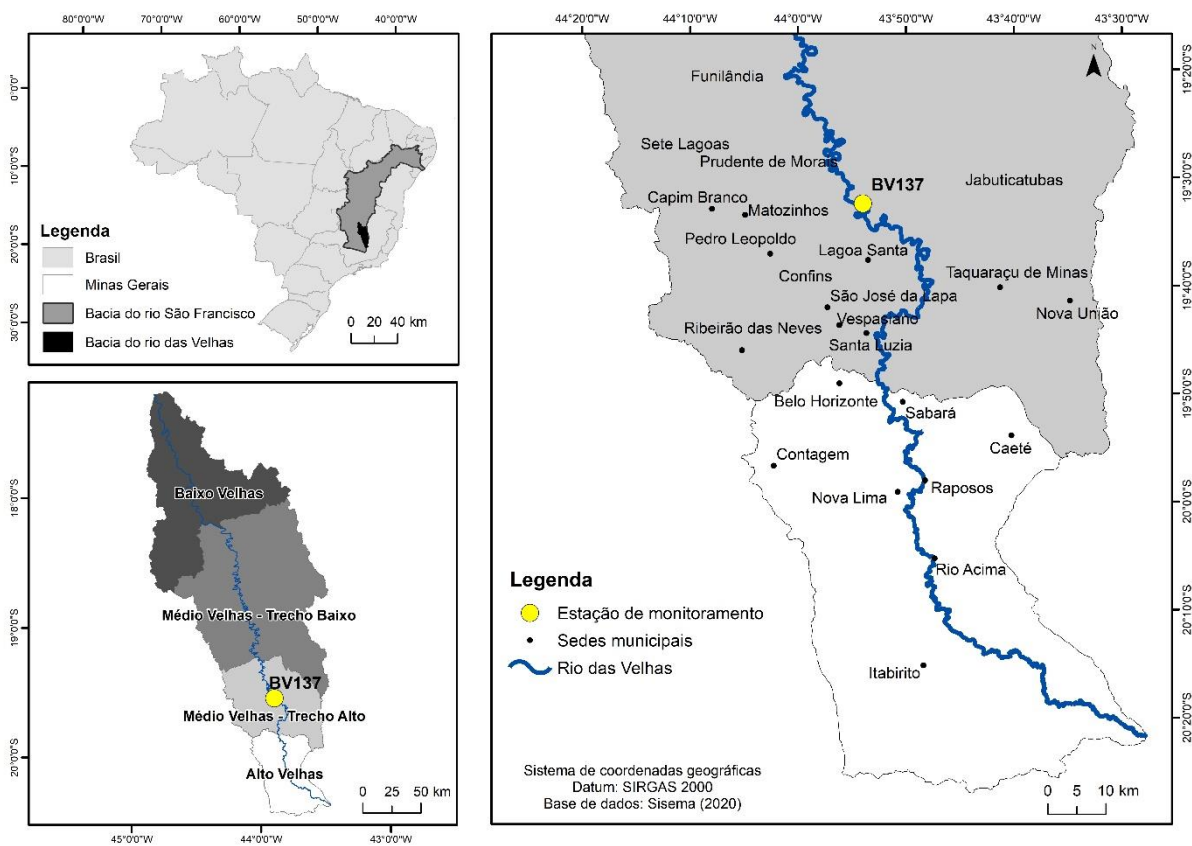
Fonte: Igam (2019a)

A amostragem dos parâmetros de qualidade da água, bem como o transporte e as análises laboratoriais são realizadas pelo Instituto SENAI de Tecnologia em Meio Ambiente, laboratório acreditado na Norma ABNT NBR ISO/IEC 17.025/2005 pelo Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO). Os resultados do monitoramento estão disponíveis no Portal InfoHidro (IGAM, 2020).

3.1.2.1 Monitoramento da qualidade das águas superficiais na bacia do rio das Velhas

No âmbito do programa “Águas de Minas”, a bacia hidrográfica do rio das Velhas possuía, no ano de 2018, 81 estações de monitoramento ativas (IGAM, 2019a). Para o estudo de caso desta pesquisa, foi selecionada uma estação de monitoramento (BV137) localizada na calha do rio das Velhas, na ponte Raul Soares, no município de Lagoa Santa (município que faz parte da Região Metropolitana de Belo Horizonte). Essa estação foi identificada como prioritária para o monitoramento e controle da poluição do rio das Velhas no estudo de Calazans et al. (2018b) e Cordova (2020). A Figura 3.3 mostra a localização da área de estudo e a estação de monitoramento da qualidade da água selecionada. Na Figura 3.4, são apresentadas as fotos das regiões a montante e a jusante da estação de monitoramento BV137.

Figura 3.3 - Localização da estação de monitoramento de qualidade da água superficial BV137



Fonte: Próprio autor.

Figura 3.4 - Registro fotográfico das regiões (a) a montante e (b) a jusante da estação de monitoramento BV137



(a) montante



(b) jusante

Fonte: Igam (2021)

A estação BV137 foi implantada em 1997, inicialmente com frequência de coletas trimestrais, passando a ser mensal a partir de 2008. De acordo com o banco de dados disponibilizado pelo Igam, no ano de 2018, eram monitorados 52 parâmetros físico-químicos e microbiológicos. Desde a implantação dessa estação até o ano de 2018 (1º semestre), foram realizadas 148 coletas, com um total de 4.421 dados válidos. Entretanto, o banco de dados não é homogêneo, pois nem todos os parâmetros são amostrados em todas as campanhas.

3.1.3 Uso e ocupação do solo

A sub-bacia do rio das Velhas é dividida em quatro macro regiões - Alto, Médio Alto, Médio Baixo e Baixo rio das Velhas (Quadro 3.2 e Figura 3.3) - em função da análise dos seguintes temas: hidrografia, tipologias de relevo, ocupação da bacia e a presença de região metropolitana com seus impactos sobre os recursos hídricos (CBH VELHAS, 2015a).

Quadro 3.2 - Regiões da bacia hidrográfica do rio das Velhas e principais características

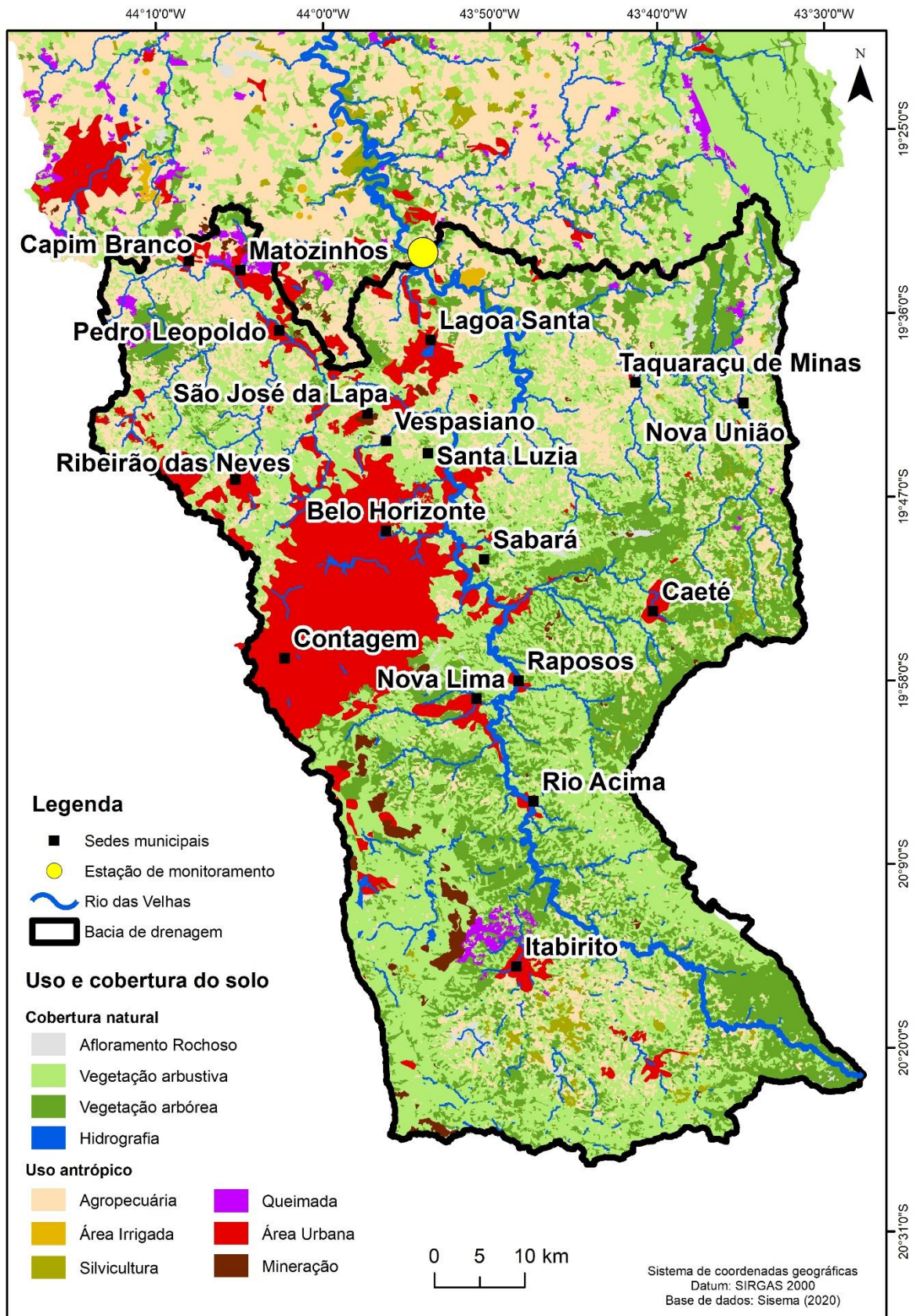
Região	Área total (km ²)	Percentual na bacia	Principais características	Principais fatores de pressão
Alto	2.739,74	9,8%	Apresenta maior contingente populacional, com uma expressiva atividade econômica concentrada, em especial, na RMBH. Nessa região encontra-se o sistema de abastecimento integrado rio das Velhas. Compreende toda a região denominada Quadrilátero Ferrífero (maior produção nacional de minério de ferro).	Esgotos industriais e domésticos não tratados e os efluentes gerados pelas atividades minerárias clandestinas atuantes nesta parte da bacia.
Médio Alto	4.276,01	15,4%	Apresenta menor concentração populacional que o Alto rio das Velhas.	Predomínio das atividades agrícolas e pecuárias.
Médio Baixo	12.204,16	43,8%	Representa a maior porção dentro da bacia do rio das Velhas.	
Baixo	8.630,07	31,0%	O aspecto diferenciado desta região é que nenhum dos municípios tem 100% do território inserido na bacia. Baixa concentração populacional.	

Fonte: Adaptado CBH Velhas (2015a)

A estação de monitoramento BV137 está localizada na região do Médio Alto rio das Velhas, que apresenta uma menor concentração populacional se comparado ao trecho Alto, com o predomínio das atividades agrícolas e pecuárias.

O levantamento dos fatores de pressão que influenciam a qualidade da água na estação BV137 foi realizado considerando a bacia de drenagem dessa estação. As classes de uso e cobertura do solo consideradas foram: cobertura natural (afloramento rochoso, vegetação arbustiva, vegetação arbórea, hidrografia) e uso antrópico (agropecuária, agricultura irrigada, silvicultura, queimada, área urbana e mineração), conforme apresentado na Figura 3.5 (CBH VELHAS, 2015a).

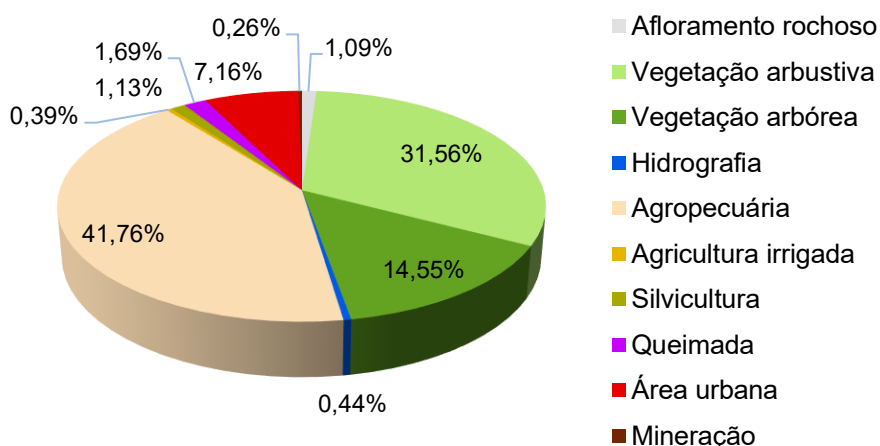
Figura 3.5 - Mapa de uso e ocupação do solo da bacia do rio das Velhas, com foco na bacia de drenagem da estação de monitoramento BV137



Fonte: Próprio autor.

A Figura 3.6 apresenta a síntese da distribuição percentual das categorias de uso e cobertura do solo para a região do Médio Alto rio das Velhas. Essa é a região que apresenta o maior equilíbrio entre as categorias de cobertura natural (47,61%) e uso antrópico (52,39%). A classe com maior representação na região é a de agropecuária, com 41,76% (CBH VELHAS, 2015b).

Figura 3.6 - Distribuição percentual das categorias de uso e cobertura do solo na região do Médio Alto rio das Velhas



Fonte: Adaptado CBH Velhas (2015b)

3.1.4 Áreas prioritárias para a conservação

As “Áreas Prioritárias para a Conservação, Utilização Sustentável e Repartição dos Benefícios da Biodiversidade” são áreas definidas geograficamente com base em informações espacializadas sobre a ocorrência dos alvos de conservação, as atividades antrópicas e as atividades que favorecem a conservação e o uso sustentável da biodiversidade (MMA, 2019b).

A bacia do rio das Velhas possui 46,1% do seu território definido como área prioritária para a conservação da biodiversidade, conforme apresentado na Tabela 3.1 (CBH VELHAS, 2015b). A localização geográfica dessas áreas prioritárias relativas à bacia do rio das Velhas pode ser verificada na Figura 3.7.

As áreas prioritárias estão concentradas, principalmente, na vertente leste da bacia, nas áreas “Espinhaço Meridional” e “Serra do Cabral”, que são especialmente significativas do ponto de vista da conservação da biodiversidade, devido à variedade de espécies e o alto grau de endemismos. A área “Quadrilátero” é a segunda maior

em extensão e ocupa a porção de nascentes, no limite sul da bacia. Está localizada a montante da área metropolitana de Belo Horizonte, em uma região onde uma atividade bastante significativa é a mineração (CBH VELHAS, 2015b).

Destaca-se o fato de que a criação de unidades de conservação é recomendada para a maioria das áreas prioritárias (34,1% do total da bacia), demonstrando a importância da bacia quanto à proteção da vida silvestre e dos processos ecológicos a ela associados (CBH VELHAS, 2015b).

A estação de monitoramento BV137 encontra-se na área “Província Cárstica de Lagoa Santa”, de importância “extrema”, e para a qual se preconiza como ação prioritária a criação de unidade de conservação.

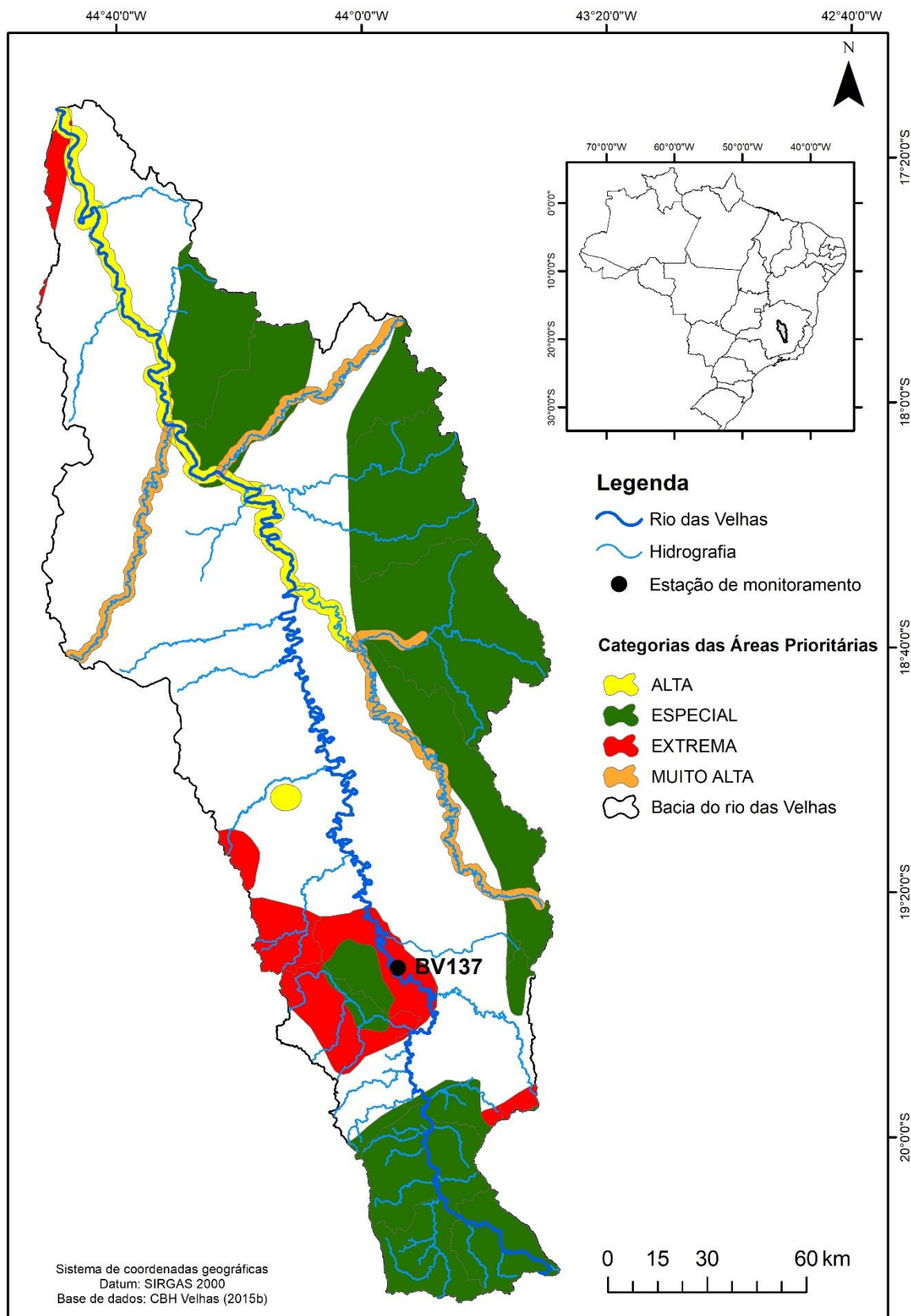
Tabela 3.1 - Áreas Prioritárias para a conservação na bacia hidrográfica do rio das Velhas

Nome	Categoria	Ação Prioritária	% da bacia
Espinhaço Meridional	Especial	Criação de UC ¹	17,8
Quadrilátero	Especial	Investigação científica	8,1
Serra do Cabral	Especial	Criação de UC	5,8
Província Cárstica de Lagoa Santa	Extrema	Criação de UC	4,9
São Francisco e Grandes Afluentes	Alta	Criação de UC	3,5
Tributários do rio das Velhas	Muito Alta	Recuperação/Reabilitação	3,4
Área Peter Lund	Especial	Criação de UC	1,2
Paraopeba	Extrema	Criação de UC	0,4
Buritizeiro/Pirapora	Extrema	Criação de UC	0,4
Florestas da Borda Leste do Quadrilátero	Extrema	Manejo	0,3
Caverna do Salitre	Alta	Investigação científica	0,2
Total			46,1

¹ Unidade de Conservação

Fonte: CBH Velhas (2015b)

Figura 3.7 - Mapa das Áreas Prioritárias para Conservação na bacia hidrográfica do rio das Velhas



Fonte: Próprio autor.

3.1.5 Monitoramento pluviométrico e fluviométrico

Uma das atribuições da ANA é a de promover a coordenação das atividades desenvolvidas no âmbito da rede hidrometeorológica nacional, em articulação com órgãos e entidades públicas ou privadas que a integram, ou que dela sejam usuárias (BRASIL, 2000). Assim, cabe à ANA manter e disponibilizar o cadastro atualizado das estações hidrometeorológicas do País (ANA, 2009).

A Rede Hidrometeorológica Nacional (RHN), em 2019, possuía aproximadamente 23 mil estações sob responsabilidade de várias entidades. A ANA gerencia diretamente 4.807 estações, sendo 2.808 pluviométricas e 1.999 estações fluviométricas. Do total de estações fluviométricas, em 1.485 delas, há medição de vazão de água (descarga líquida) (ANA, 2020a).

O acesso aos dados e informações do monitoramento ocorre pelo Sistema de Informações Hidrológicas (HidroWeb), portal que é vinculado ao Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH), que reúne dados de níveis fluviais, vazões, chuvas, climatologia, qualidade da água e sedimentos (ANA, 2020a, 2020b).

Na área da bacia de drenagem da estação BV137, existem 15 estações pluviométricas que se encontram em operação (Tabela 3.2), com dados disponíveis no Portal HidroWeb (ANA, 2020b).

Em relação às estações fluviométricas localizadas na região de estudo, uma delas - a estação 41340000 - está localizada na calha do rio das Velhas, a uma distância de 4,8 km da estação de monitoramento BV137, sendo, portanto, a estação selecionada para este estudo. A distribuição geográfica das estações pluviométricas e da estação fluviométrica de interesse está apresentada na Figura 3.8.

Tabela 3.2 - Descrição das estações pluviométricas localizadas na bacia de drenagem da estação BV137

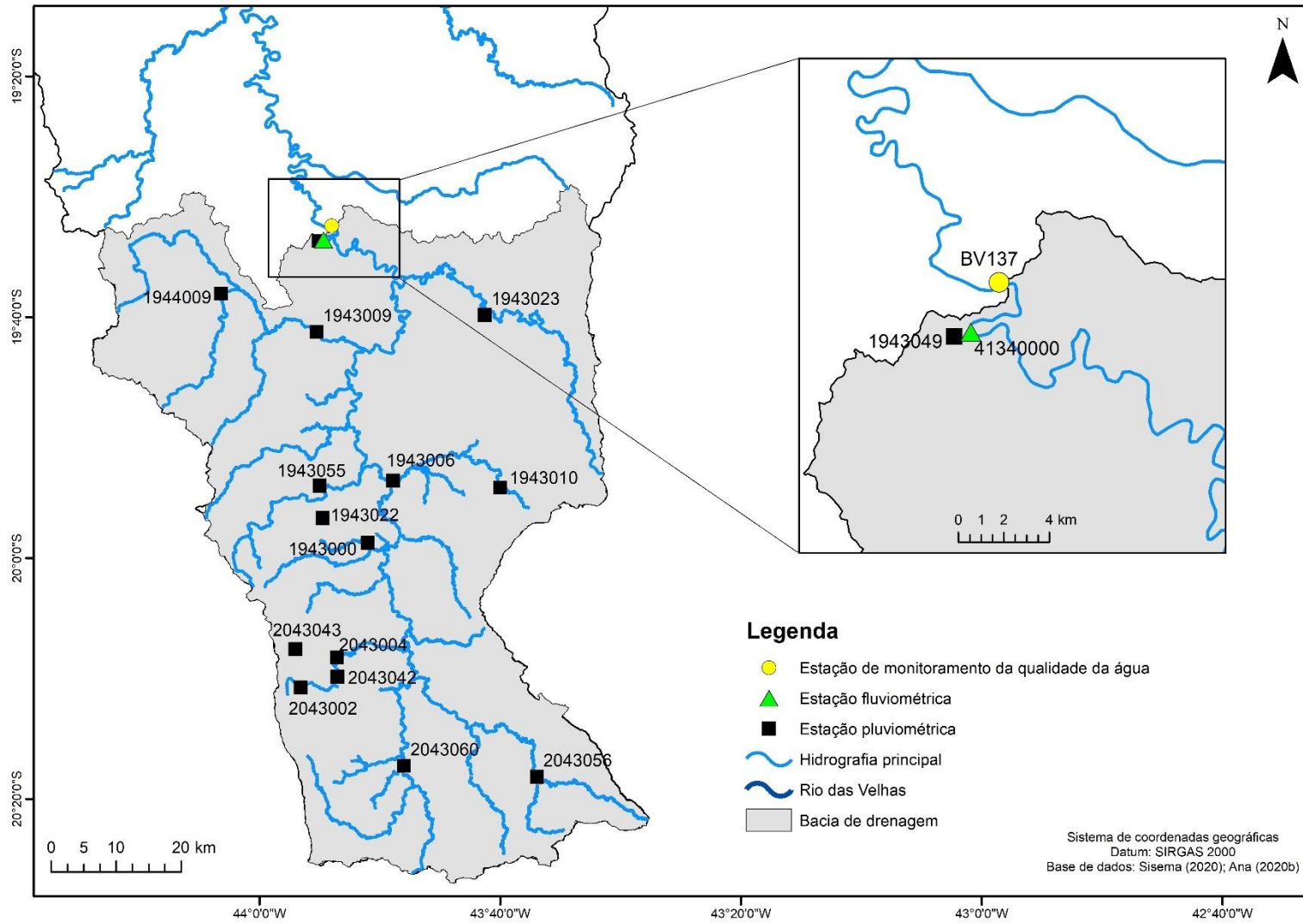
Município	Entidade responsável	Código	Nome da estação	Latitude	Longitude
Nova Lima	ANA	1943000	Mineração Morro Velho	-19,9792	-43,85
Sabará	ANA	1943006	Sabará	-19,8931	-43,815
Vespasiano	ANA	1943009	Vespasiano	-19,6872	-43,9208
Caeté	ANA	1943010	Caeté	-19,9028	-43,6664
Belo Horizonte	ANA	1943022	Caixa de Areia	-19,945	-43,9125
Taquaraçu de Minas	ANA	1943023	Taquaraçu	-19,6639	-43,6881
Lagoa Santa	ANA	1943049	Ponte Raul Soares	-19,5614	-43,9178
Belo Horizonte	INMET ¹	1943055	Belo Horizonte (Horto)	-19,9	-43,9167
Pedro Leopoldo	ANA	1944009	Pedro Leopoldo	-19,6344	-44,0533
Nova Lima	ANA	2043002	Lagoa Grande (MMV ²)	-20,1792	-43,9428
Nova Lima	ANA	2043004	Rio do Peixe (MMV)	-20,1378	-43,8925
Nova Lima	ANA	2043042	Represa das Codornas (MMV)	-20,1647	-43,8919
Nova Lima	ANA	2043043	Represa do Miguelão (MMV)	-20,1261	-43,9503
Ouro Preto	ANA	2043056	Fazenda Água Limpa Jusante	-20,3031	-43,6158
Itabirito	ANA	2043060	Itabirito Linígrafo	-20,2878	-43,8

¹ Instituto Nacional de Meteorologia

² Mineração Morro Velho S.A.

Fonte: ANA (2020b).

Figura 3.8 - Localização geográfica das estações de monitoramento pluviométrico, fluviométrico e de qualidade da água superficial, utilizadas neste estudo



Fonte: Próprio autor.

3.2 DEFINIÇÃO DO BANCO DE DADOS DE MONITORAMENTO DE QUALIDADE DA ÁGUA

A partir do banco de dados da estação BV137, disponibilizado pelo Igam, foi realizada, em duas etapas, a definição dos dados utilizados nesse estudo: seleção do período e dos parâmetros de qualidade da água. Na primeira etapa, selecionou-se todo o período de dados disponível para a estação de monitoramento.

Na segunda etapa, foram calculados os percentuais de dados faltantes e de dados censurados (valores abaixo ou acima do limite de detecção do método analítico), considerando-se os dados do período selecionado, para verificar a necessidade de exclusão de parâmetros. Em muitas análises multivariadas, a presença de dados faltantes pode eliminar tantas observações que uma amostra que era adequada fica reduzida a uma amostra inadequada (HAIR et al., 2009). Assim, foram calculados os percentuais de dados faltantes para todos os parâmetros, e após um exame minucioso desses percentuais, optou-se por eliminar aqueles que apresentaram um percentual superior a 5% para evitar a redução do tamanho da amostra disponível para a análise, visto que ela já era reduzida (148 amostras). De acordo com Reimann et al. (2008), a proporção de amostras que possuem dados faltantes deve ser a mais baixa possível.

Para os parâmetros restantes, foram calculados os percentuais de dados censurados, pois quase todos os tipos de análises estatísticas sofrem interferência desse tipo de dado (CHRISTOFARO; LEÃO, 2014). O maior percentual de dados censurados obtido foi de 11,5% para o parâmetro coliformes termotolerantes. Segundo Reimann et al. (2008), quando métodos multivariados padrão forem utilizados, a proporção total de amostras com dados censurados provavelmente não deve exceder 10%. Entretanto, devido à importância ambiental desse parâmetro, mesmo com um percentual superior a 10%, ele foi mantido nas análises e, dessa forma, nenhum parâmetro foi eliminado devido ao percentual de dados censurados.

Ressalta-se que os resultados de coliformes termotolerantes foram utilizados em conjunto com os resultados de *Escherichia coli* para complementar a série histórica desse indicador de contaminação fecal, visto que o parâmetro *Escherichia coli* passou a ser avaliado em substituição aos coliformes termotolerantes a partir da primeira campanha de 2013.

3.3 PRÉ-PROCESSAMENTO DOS DADOS

A etapa de pré-processamento é crucial em qualquer atividade de análise de dados, e visa conhecer as principais características do conjunto de observações e gerar um banco de trabalho a partir dos dados brutos. Apesar de essencial, as práticas de pré-tratamento de dados ambientais ainda são, algumas vezes, limitadas em conhecimento e uso, conforme relatado por Gholizadeh, Melesse e Reddi (2016).

Com o banco de dados definido no Item 3.2, inicialmente, efetuou-se uma inspeção preliminar, a partir de histogramas e gráficos de dispersão e de séries temporais, com o objetivo de identificar dados faltantes e registros atípicos, 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 foram substituídos pelo próprio LD (limite de detecção), o que caracteriza o pior cenário ambiental possível. No caso dos dados faltantes, toda a amostra foi excluída do banco de dados.

Em sequência, realizou-se a análise descritiva, por parâmetro, a partir do cálculo de média, desvio-padrão, coeficiente de variação, mínimo, máximo, percentis 25, 50 e 75, e o coeficiente de assimetria. Para verificar os níveis de degradação, o percentual de violação dos parâmetros de qualidade da água, em relação aos limites preconizados pela Deliberação Normativa Conjunta do Conselho Estadual de Política Ambiental e Conselho Estadual de Recursos Hídricos (DN COPAM/CERH-MG) nº 01/2008 (MINAS GERAIS, 2008) foi calculado segundo a classe de enquadramento do corpo d'água no local da estação BV137 (classe 3). Os parâmetros de qualidade da água que apresentaram os maiores percentuais de violação foram destacados.

Além disso, a correlação de Spearman (SPEARMAN, 1904) foi calculada para estimar o nível de correlação entre os parâmetros. O coeficiente de correlação de Spearman (ρ) é um caso especial do coeficiente de Pearson. Trata-se de uma medida não-paramétrica de correlação, no qual os dados são convertidos em ranks antes de se calcular o coeficiente. O teste de Shapiro-Wilk, com um nível de significância (α) de 5%, foi aplicado para verificar se os dados seguiam ou não a distribuição normal (REIMANN et al., 2008).

Num segundo momento, identificaram-se, de modo quantitativo, os dados discrepantes (*outliers*), definidos como observações atípicas que possuem características notavelmente diferentes das demais (HAIR et al., 2009).

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; GATHER, 1993; LIN et al., 2007). Nesse caso, a estatística MAD (*median absolute deviation*) (Equação 3.1) é calculada 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 3.2) e superior (Equação 3.3), são mais estreitos em relação à abordagem tradicional.

$$MAD = 1,4826 \cdot \text{Mediana}\{|x_i - \tilde{X}|\} \quad (3.1)$$

$$LI (\text{Limite Inferior}) = \tilde{X} - 3 \cdot MAD \quad (3.2)$$

$$LS (\text{Limite Superior}) = \tilde{X} + 3 \cdot MAD \quad (3.3)$$

A estatística MAD é um método robusto a valores extremos presentes no conjunto de dados; conseqüentemente, é menos sensível a desvios consideráveis da distribuição normal (REIMANN et al., 2008). O fator 1,4826 é empregado para que o MAD esperado seja equivalente ao desvio-padrão σ para dados normalmente distribuídos (LIN et al., 2007).

Identificados os dados discrepantes, conforme sugerido por Hair et al. (2009), eles foram analisados individualmente para a decisão de mantê-los ou retirá-los das análises. Para cada dado discrepante, verificaram-se os resultados de todos os parâmetros obtidos na mesma data, as coletas realizadas em datas próximas, e qual a condição do tempo na data da coleta, para verificar se esses valores poderiam ou não serem considerados erros de medição ou de registro. Caso nenhum erro fosse encontrado, o dado discrepante não era rejeitado.

As análises estatísticas realizadas foram implementadas em linguagem de programação Python (<https://www.python.org/>), versão 2014, e foi utilizado também o Microsoft Excel (2019).

3.4 RESULTADOS E DISCUSSÃO

3.4.1 Definição do banco de dados de monitoramento de qualidade da água

O banco de dados da estação BV137 era composto inicialmente por 52 parâmetros de qualidade da água monitorados no período de 1997 a 2018 (1º semestre). Quanto à seleção do período de estudo, foram utilizados os dados do período completo.

Em relação aos parâmetros monitorados, como o banco de dados era muito heterogêneo, com frequência de amostragem mensal, trimestral e semestral, que variava entre os parâmetros, o banco de dados de estudo selecionado foi bastante reduzido. Do total de parâmetros monitorados, apenas 12 (aproximadamente 23% do total) apresentaram menos de 5% de dados faltantes (Tabela 3.3). Considerando esses parâmetros, nenhum deles foi excluído do banco de dados devido ao percentual de dados censurados, sendo, portanto, todos considerados nas análises.

Em resumo, a base de dados selecionada para análise foi composta por 12 parâmetros físico-químicos e microbiológicos de qualidade da água, monitorados de 1997 a 2018 (monitoramento inicialmente trimestral, e a partir de 2008, mensal), formando uma matriz de dados brutos $X_{[n=148, p=12]}$, onde n é o número de observações por variável e p é o número de variáveis, contendo 1.774 dados válidos.

Tabela 3.3 - Parâmetros físico-químicos e microbiológicos da qualidade da água com respectiva abreviação, unidade e técnica analítica

Parâmetros	Abreviação	Unidade	Técnica analítica
Arsênio total	As _T	mg L ⁻¹ As	Espectrometria de absorção atômica com geração de hidreto
Cloreto total	Cl ⁻	mg L ⁻¹ Cl	Colorimétrico
Condutividade elétrica <i>in loco</i>	CE	μS.cm ⁻¹	Conductimétrico
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 totais	ST	mg L ⁻¹	Gravimétrico
Temperatura da água	T _{água}	°C	Termométrico
Turbidez	Turb.	UNT ^b	Turbidimétrico

^a Número mais provável.

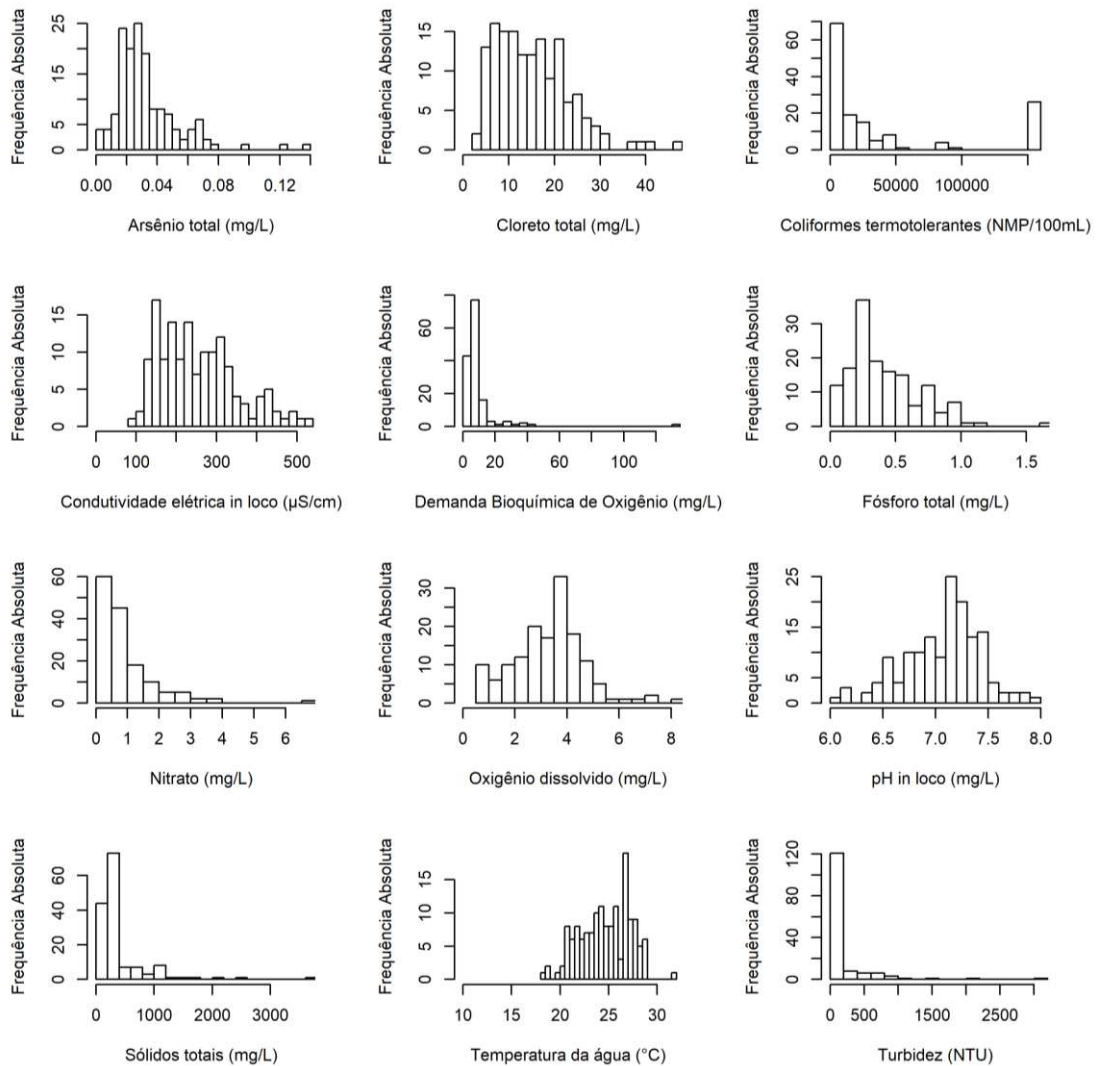
^b Unidade nefelométrica de turbidez

Fonte: Próprio autor.

3.4.2 Pré-processamento dos dados

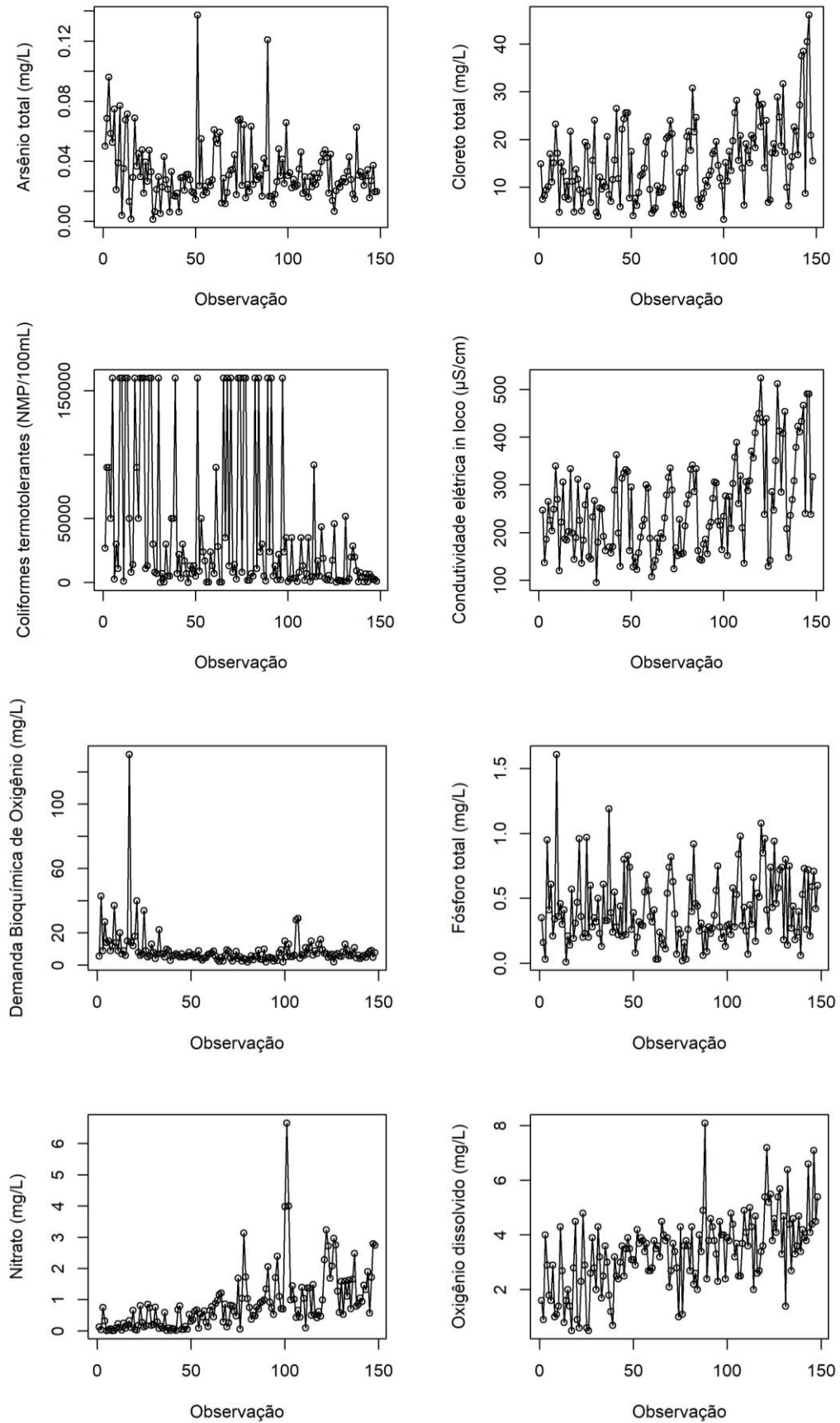
A partir da matriz de dados selecionada $X_{[n=148, p=12]}$, as análises preliminares dos dados consistiram na elaboração de histogramas e gráficos de dispersão e de séries temporais para conhecer o conjunto de dados. A Figura 3.9 e a Figura 3.10 exemplificam, respectivamente, o histograma e o gráfico temporal para todos os parâmetros selecionados.

Figura 3.9 - Histograma dos parâmetros de qualidade de água analisados no período de 1997 a 2018



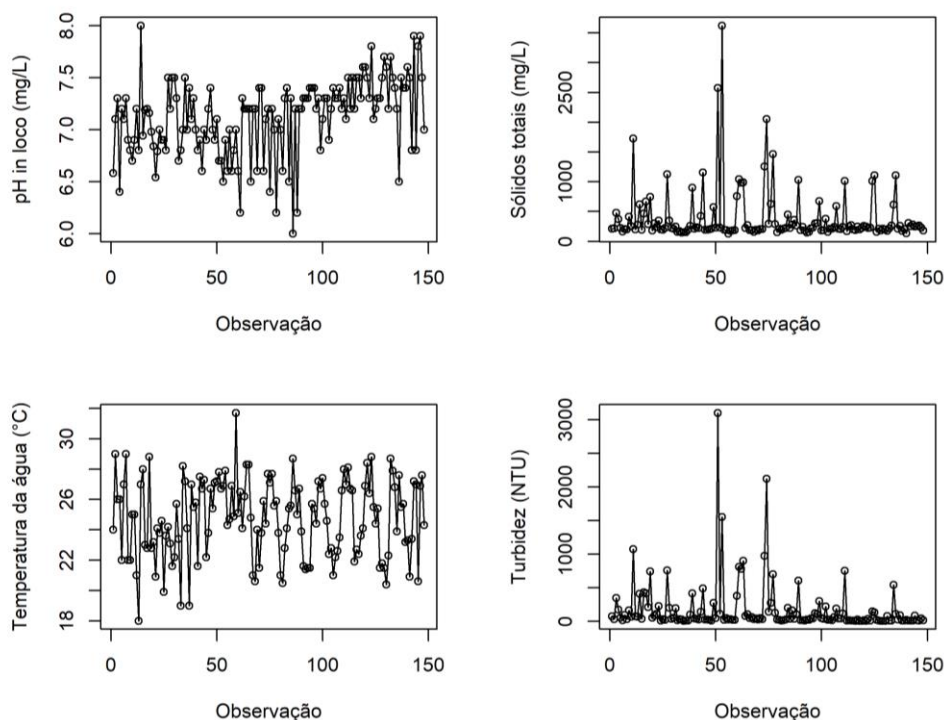
Fonte: Próprio autor.

Figura 3.10 - Gráfico de séries temporais dos parâmetros de qualidade de água analisados no período de 1997 a 2018



Fonte: Próprio autor.

Figura 3.10 - Gráfico de séries temporais dos parâmetros de qualidade de água analisados no período de 1997 a 2018 (continuação)



Fonte: Próprio autor.

Duas amostras foram excluídas em função dos dados faltantes, e o conjunto de dados de trabalho passou a conter 146 registros e 1.752 observações válidas. A Tabela 3.4 apresenta a estatística descritiva dos parâmetros de qualidade da água.

Tabela 3.4 - Estatística descritiva do banco de dados de monitoramento da estação BV137, no período de 1997 a 2018

Parâmetro	Limite ^a (classe 3)	Mín.	Máx.	Média	Desvio- padrão	Perc. 25	Perc. 50	Perc. 75	CV ^b (%)	Assimetria
As _T	0,033	0,001	0,138	0,033	0,021	0,020	0,029	0,040	62,5	1,93
Cl ⁻	250	3,27	46,10	15,28	8,12	8,88	14,05	20,53	53,0	0,93
Colif. term.	4000	50	160000	41640	58385	3300	13000	49028	139,7	1,37
CE	-	96	524	255	97	177	238	311	38,1	0,68
DBO	10	2,00	131,00	9,29	12,26	4,93	6,45	9,00	131,5	7,22
P _T	0,15	0,010	1,610	0,412	0,267	0,230	0,335	0,568	64,7	1,10
N-NO ₃ ⁻	10	0,010	6,660	0,893	0,951	0,235	0,665	1,118	106,1	2,47
OD	>4	0,50	8,10	3,38	1,38	2,60	3,55	4,18	40,6	0,23
pH	6 a 9	6,00	8,00	7,12	0,37	6,90	7,20	7,38	5,1	-0,45
S _T	-	130	3622	395	460	194	226	304	116,1	3,93
T _{água}	-	18,00	31,70	24,80	2,54	22,80	25,00	26,90	10,2	-0,17
Turb.	100	4	3100	174	382	17	40	136	218,8	4,68

^aLimites da DN COPAM/CERH-MG nº 01/2008

^bCV - coeficiente de variação

Fonte: Próprio autor.

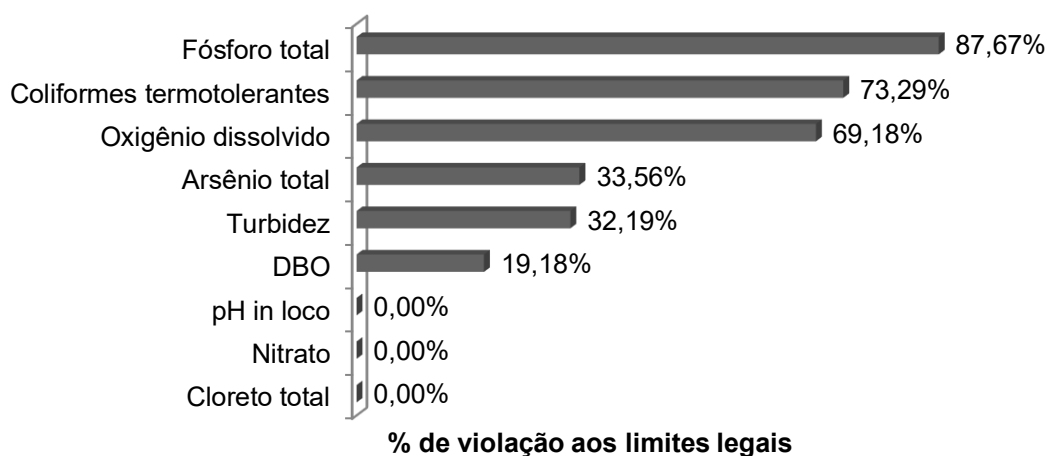
Ressaltam-se os elevados resultados de coliformes termotolerantes e concentração de fósforo total, com valores medianos de 13.000 NMP/100 mL e 0,335 mg/L, respectivamente, e as baixas concentrações de oxigênio dissolvido, com valor mediano igual a 3,55 mg/L, considerado baixo para a classe de enquadramento do curso d'água.

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, $T_{\text{água}}$, CE e OD, indicando menor variabilidade nos dados desses parâmetros. A maioria dos parâmetros apresentaram valores de CV altos indicando grande variabilidade dos resultados, uma 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 Follador (2010) e Toledo e Nicolella (2002).

O coeficiente de assimetria caracteriza o grau de assimetria de uma distribuição em torno de sua média e pode assumir valores positivos (distribuição enviesada para a direita), negativos (distribuição enviesada para a esquerda), ou zero (distribuição normal perfeita) (VON SPERLING; VERBYLA; OLIVEIRA, 2020). Observa-se que nenhum parâmetro possui distribuição normal perfeita e que a distribuição das concentrações para a maioria dos parâmetros é positivamente assimétrica, exceto para pH e temperatura da água; resultado comum para dados de monitoramento da qualidade da água (VON SPERLING; VERBYLA; OLIVEIRA, 2020).

Os valores dos parâmetros de qualidade da água selecionados também foram comparados com os padrões definidos na DN COPAM/CERH-MG nº 01/2008, do estado de Minas Gerais (MINAS GERAIS, 2008) (Anexo 1). Os parâmetros Colif. term., P_T e OD, foram os que apresentaram os maiores percentuais de violação aos limites legais (acima de 50%), conforme mostrado na Figura 3.11. Dos resultados de Colif. term., 11,5% encontram-se acima do limite de detecção do método analítico (>160.000 NMP/100 mL), sendo este o parâmetro com o maior percentual de dados censurados registrado no banco de dados. Esses resultados são um indicativo de degradação da qualidade das águas nessa região, consequência, sobretudo, dos lançamentos de efluentes domésticos e industriais dos municípios localizados a montante desta estação de monitoramento.

Figura 3.11 - Percentual de violação aos limites legais, considerando os dados no período de 1997 a 2018



Fonte: Próprio autor.

Para medir o grau de correlação entre as variáveis de qualidade de água, foi utilizado o coeficiente de correlação de Spearman. Na Figura 3.12, os pontos vermelhos e azuis correspondem a correlações negativas e positivas, respectivamente. Pequenos pontos com intensidade de cor clara representam baixas correlações, enquanto grandes pontos com cores mais escuras correspondem a correlações mais altas.

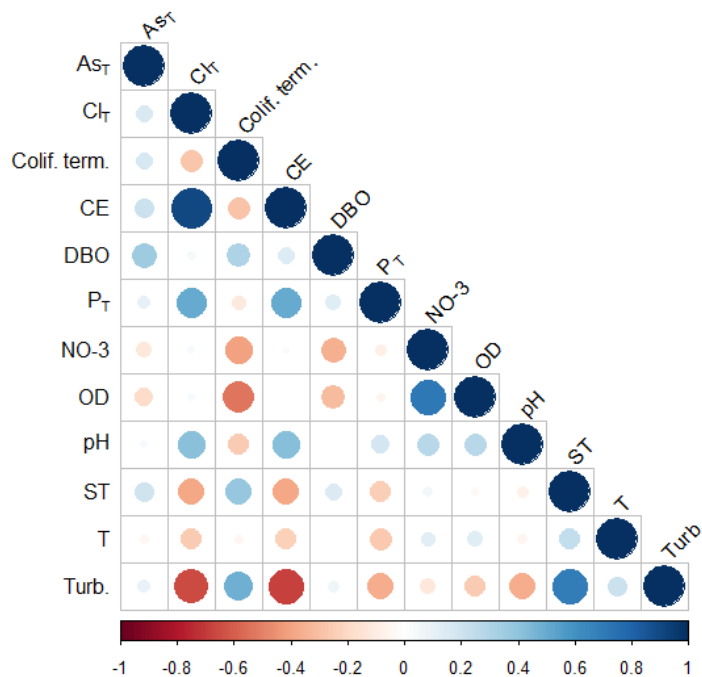
Observa-se que todas as variáveis apresentaram pelo menos uma correlação significativa com outra variável, confirmando o padrão de inter-relação entre os parâmetros de qualidade da água. Os valores de correlação variaram, em módulo, de 0,001 a 0,907. As maiores correlações significativas (valor- $p < 0,05$) foram: condutividade elétrica está positivamente correlacionada com cloreto total, o que pode ser explicado pelo fato do parâmetro condutividade elétrica ser influenciado pela presença de diferentes íons (APHA; AWWA; WEF, 2017). Por sua vez, turbidez está positivamente correlacionada com sólidos totais, que é um resultado esperado, tendo em vista que o tipo e a concentração da matéria em suspensão influenciam a turbidez e a transparência da água (CHAPMAN et al., 1992).

O oxigênio dissolvido (OD) está negativamente correlacionada com coliformes termotolerantes e DBO, e positivamente correlacionado com nitrato e pH. As descargas de efluentes com alto teor de matéria orgânica e nutrientes, nos cursos de água, podem levar a diminuições nas concentrações de OD, como resultado do

aumento da atividade microbiana que ocorre durante a degradação da matéria orgânica (CHAPMAN et al., 1992), justificando as correlações negativas entre OD e DBO.

Como as variáveis são correlacionadas, é possível simplificar o banco de dados e pode-se concluir que alguns fatores podem explicar a principal variância de todos os parâmetros e serem usados para encontrar uma estrutura interna e auxiliar para estudos de qualidade da água (PARINET; LHOTE; LEGUBE, 2004).

Figura 3.12 - Representação gráfica da matriz de correlação de Spearman relacionando todas as variáveis de qualidade de água selecionadas



Fonte: Próprio autor.

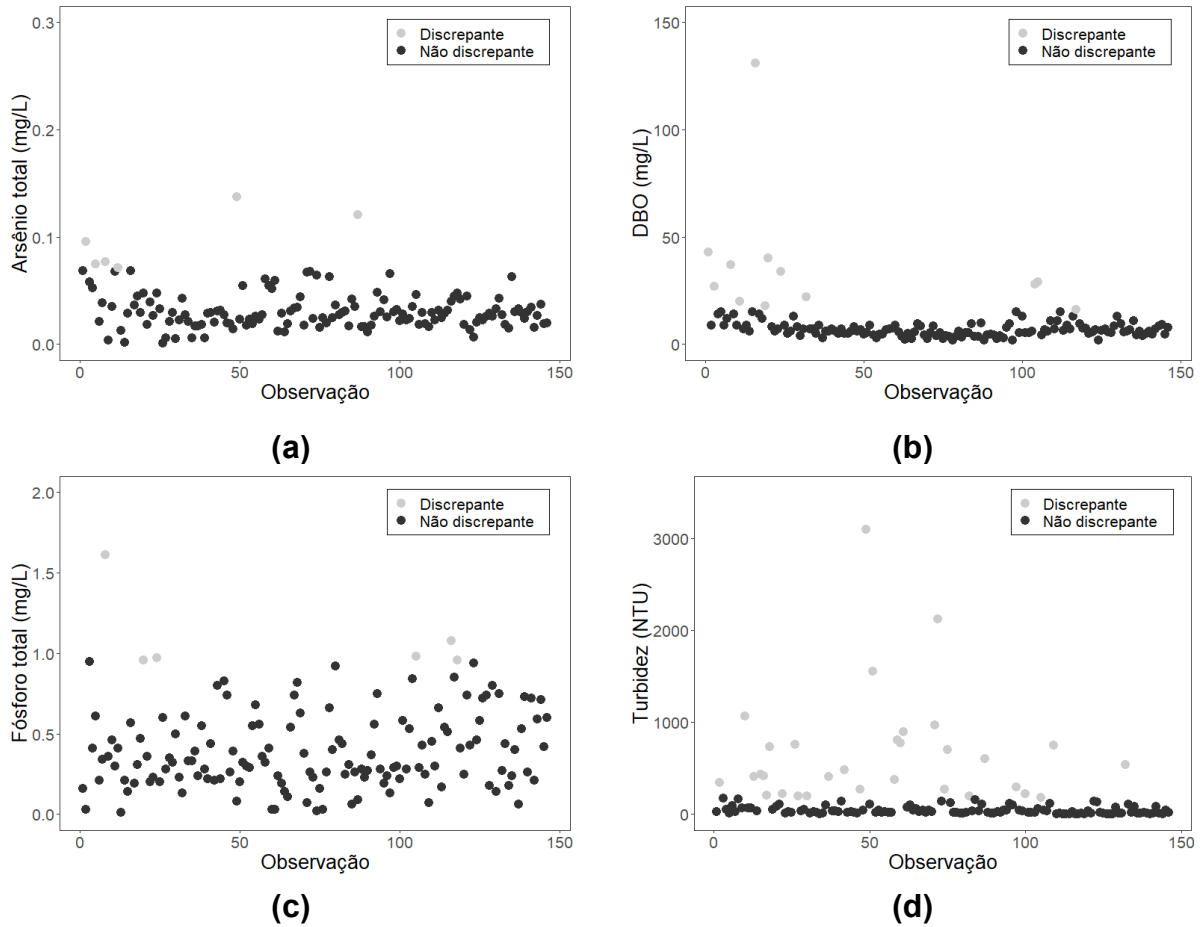
Os testes de normalidade de Shapiro-Wilk foram aplicados para cada parâmetro de qualidade e nenhum deles apresentou aderência à distribuição normal, ao nível de significância (α) de 5%. Resultados já esperados, visto que os dados de qualidade de água geralmente são caracterizados por uma distribuição não normal (HELSEL; HIRSCH, 2002). Os testes de aderência de Qui-quadrado mostraram que os parâmetros Turb., T, pH, CE, Cl_T e As_T, podem ser modelados pela distribuição lognormal, ao nível de significância de 5%. Os parâmetros ST, NO₃⁻, DBO e Colif. term., apresentaram também acentuada assimetria à direita, mas o teste não

confirmou a aderência para (α) de 5%. Apenas as distribuições de P_T e OD foram indefinidas (multimodal).

Por fim, de acordo com a metodologia, os dados discrepantes foram identificados após aplicação do filtro de Hampel. Eles foram analisados individualmente e nenhum deles foi excluído do banco de dados. Todos os dados utilizados nessa pesquisa foram consistidos e confirmados, sendo os *outliers* utilizados como indicativos de condições críticas de qualidade da água. A Figura 3.13 (a-d) apresenta os gráficos temporais, com a identificação desses dois tipos de observações (discrepantes e não discrepantes), para arsênio total, DBO, fósforo total e turbidez, respectivamente.

Para um conjunto de dados ambientais, grandes variações nos dados podem ser consequência da variabilidade espacial ou temporal inerente ao processo ambiental ou algum fator não identificado relacionado à grandeza analisada (CARVALHO, 2003).

Figura 3.13 - Gráfico temporal com dados discrepantes e não discrepantes identificados pelo filtro de Hampel para os parâmetros (a) arsênio total, (b) DBO, (c) fósforo total e (d) turbidez



Fonte: Próprio autor.

As informações apresentadas neste capítulo (área de estudo, definição do banco de dados de monitoramento da qualidade da água e pré-processamento) são a base para o desenvolvimento das metodologias e aplicações dos próximos capítulos.

**CAPÍTULO 4:
UTILIZAÇÃO DE MODELO
BASEADO EM ANÁLISE POR
COMPONENTES PRINCIPAIS
PARA IDENTIFICAÇÃO DE
CONDIÇÕES CRÍTICAS DE
QUALIDADE DE ÁGUA
SUPERFICIAL**

RESUMO

Uma das intervenções em bacias hidrográficas visando à restauração de mananciais são os Programas de Monitoramento da Qualidade da Água (PMQA). Eles são de suma importância para a proteção do recurso hídrico, a manutenção da capacidade ambiental e o controle da poluição. Uma maneira de melhorar o processo de tomada de decisão sobre esses aspectos é incrementar as análises dos dados coletados. 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), contendo doze parâmetros físicos, químicos e microbiológicos. O modelo ACP utilizado como sistema de monitoramento foi capaz de discriminar eventos críticos de poluição. 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 utilizada em sistemas de suporte à decisão em atividades de monitoramento e gerenciamento de qualidade de água.

Palavras-chave: Monitoramento; Qualidade da água; Análise por Componentes Principais; Eventos críticos; Fontes de poluição; Visualização de dados.

4.1 INTRODUÇÃO

A restauração de rios e a melhoria da qualidade das águas são umas das questões ambientais mais importantes do mundo contemporâneo (VOZA; VUKOVIĆ, 2018). As atividades antrópicas representam desafios para a segurança da qualidade da água, fazendo com que a identificação de fontes potenciais de poluição, e a quantificação de suas contribuições correspondentes, sejam essenciais para a gestão da água e controle da poluição (ZHANG et al., 2020).

No Brasil, vários estudos vêm apontando o lançamento de efluentes domésticos sem tratamento ou com tratamento insuficiente como uma das principais fontes de poluição dos rios (BOLLMANN; EDWIGES, 2008; CALAZANS et al., 2018b; COSTA et al., 2017; OLIVEIRA et al., 2018; PESSOA; ORRICO; LORDÉLO, 2018; PINTO; CALAZANS; OLIVEIRA, 2019; SOUZA; GASTALDINI, 2014). A média de atendimento da população urbana com sistemas coletivos de coleta e tratamento de esgotos é de apenas 46,5% (ANA, 2020c).

O monitoramento da qualidade da água é a base para a gestão da qualidade da água e possibilita a obtenção de informações, por exemplo, para descrever os recursos hídricos e identificar problemas reais e emergentes de poluição da água (BARTRAM et al., 1996). E para facilitar/auxiliar na interpretação das informações obtidas a partir do monitoramento, uma ferramenta comumente empregada é a Análise por Componentes Principais (ACP). A 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 reduzir a dimensionalidade de um conjunto de dados que possui variáveis inter-relacionadas, mantendo o máximo possível da variabilidade presente no conjunto de dados original, a aplicação de ACP a dados de monitoramento de qualidade de água cresceu consideravelmente nos últimos anos, com estudos realizados em diferentes países (Quadro 4.1).

Quadro 4.1 - Informações sobre estudos nos quais a ACP foi aplicada a dados de qualidade de água superficial

Referência	Área de estudo	Objetivos	Resumo da aplicação de ACP
Vega et al. (1998)	Rio Pisuerga, bacia do Duero, Espanha	Avaliar a influência que a poluição e a sazonalidade têm na qualidade da água do rio e discriminar os efeitos individuais do clima e das atividades humanas na hidroquímica do rio. Foram usadas as técnicas de ACP, ANOVA ^a e AC ^b .	3 estações de monitoramento ACP aplicada a todo o conjunto de dados Rotação <i>varimax</i> das componentes principais <i>Biplot</i> dos dois primeiros fatores Análise dos <i>scores</i> considerando as estações e os meses em que as coletas de água ocorreram
Zhang et al. (2010)	Rio Xiangjiang, Província de Hunan, China	Identificar semelhanças ou diferenças entre os locais de amostragem; identificar fatores potenciais que explicam a estrutura do conjunto de dados; avaliar a influência de possíveis fontes de poluição nos parâmetros de qualidade das águas dos rios e estimar a contribuição relativa de possíveis fontes para a concentração dos parâmetros identificados. Foram utilizadas técnicas de ACP e AC.	15 estações de monitoramento ACP aplicada a todo o conjunto de dados Análise das 4 primeiras componentes principais realizada por meio de gráficos de séries temporais e <i>box-plots</i>
Mohamed et al. (2015)	Bacia do rio Klang, Malásia	Avaliar a importância dos parâmetros de qualidade da água e identificar os fatores potenciais que contribuem para as variações na bacia. Foram utilizadas estatísticas descritivas e técnicas estatísticas multivariadas (AF ^c e ACP).	20 estações de monitoramento ACP aplicada a todo o conjunto de dados Aplicação de ACP seguida de AF <i>Biplots</i> da combinação dos três primeiros fatores Análises focadas, principalmente, nos <i>loadings</i>
Bengraïne; Marhaba (2003)	Rio Passaic, Nova Jersey, EUA	Extrair os fatores associados à variabilidade hidroquímica e obter as mudanças espaciais e temporais na qualidade da água. Foi utilizada ACP e AF.	12 estações de monitoramento Aplicação de ACP seguida de AF ACP aplicada a todo o conjunto de dados ACP aplicado ao banco de dados dividido em 4 de acordo com as estações do ano Análises focadas nos <i>loadings</i>
Singh et al. (2004)	Rio Gomti, Índia	Avaliar as variações espaciais e temporais na matriz de dados de qualidade da água, utilizando técnicas estatísticas multivariadas (AC, ACP/AF, AD ^d).	8 estações de monitoramento ACP aplicada a todo o conjunto de dados Rotação <i>varimax</i> das componentes principais Análises focadas nos <i>loadings</i>
Gatica et al. (2012)	Bacia do Rio Cuarto, Córdoba, Argentina	Avaliar a dinâmica da qualidade da água por meio da análise de parâmetros físicos e químicos e avaliar as diferenças sazonais e a determinação dos índices de qualidade de zonas ripárias (<i>riparian quality index</i>). Foram utilizadas as técnicas de ACP, AC e ANOVA.	6 estações de monitoramento ACP aplicado ao banco de dados dividido em 2 de acordo com as estações seca e chuvosa Avaliação das duas primeiras componentes

Quadro 4.1 - Informações sobre estudos em a ACP foi aplicada a dados de qualidade de água superficial (continuação)

Referência	Área de estudo	Objetivos	Resumo da aplicação de ACP
Noori et al. (2010)	Rio Karoon, Irã	Identificar as estações de monitoramento da qualidade da água mais informativas, determinar os parâmetros de qualidade da água mais importantes e explorar a relação entre parâmetros físicos e químicos. Foram utilizadas as técnicas de ACP e ACC ^e .	17 estações de monitoramento ACP aplicada a todo o conjunto de dados ACP aplicada considerando as estações de monitoramento como variáveis para identificar as estações mais importantes ACP aplicada considerando os parâmetros como variáveis, realizando análise de <i>loadings</i> para avaliar os parâmetros mais importantes
Phung et al. (2015)	Rio Mekong, bacia do Delta do Mekong, Vietnã	Examinar as semelhanças ou diferenças entre os locais de amostragem e a estação, identificar as variáveis de qualidade da água responsáveis por variações espaciais e temporais na qualidade da água do rio e avaliar a influência de possíveis fontes nos parâmetros de qualidade da água no local da pesquisa. Foram utilizadas técnicas estatísticas multivariadas nas análises – AC, ACP/AF e AD.	38 estações de monitoramento Rotação <i>varimax</i> das CPs ACP/AF foi aplicada para todo o conjunto de dados e nos agrupamentos individuais delineados pela análise de cluster. Avaliação dos <i>loadings</i> para cada CP rotacionada
Taoufik; Khouni; Ghrabi (2017)	Rio Wadi El Bey, Tunísia.	Extrair informações sobre as semelhanças ou dissimilaridades entre os locais de amostragem, a identificação de variáveis de qualidade da água responsáveis por variações temporais na qualidade da água do rio, os fatores ocultos que explicam a estrutura do banco de dados e a influência das possíveis fontes (naturais e antropogênicas) nos parâmetros de qualidade da água. Foram utilizadas técnicas estatísticas multivariadas nas análises – AC e ACP.	13 estações de monitoramento Aplicação de ACP seguida de AF ACP/AF foi aplicada nos agrupamentos individuais delineados pela análise de cluster. Avaliação dos <i>loadings</i> para cada CP rotacionada
Voza; Vuković (2018)	Bacia do rio Morava, Sérvia	Desenvolver um modelo híbrido para a avaliação e previsão de variações temporais na qualidade das águas superficiais, a fim de criar uma estratégia de monitoramento ideal; reconhecer os parâmetros e fontes que têm maior impacto nas variações temporais da qualidade da água. Foram utilizadas técnicas estatísticas multivariadas nas análises – AC, ACP/AF e AD e redes neurais artificiais.	14 estações de monitoramento Aplicação de ACP seguida de AF ACP/AF foi aplicada nos agrupamentos individuais delineados pela análise de cluster. Avaliação dos <i>loadings</i> para cada CP rotacionada
Cruz et al. (2019)	Rio Siriri, Sergipe, Brasil	Avaliar a influência do uso/cobertura do solo e da sazonalidade na variabilidade da qualidade da água na bacia. Foram utilizadas as técnicas de ACP e ANOVA.	5 estações de monitoramento ACP aplicada a todo o conjunto de dados Rotação <i>varimax</i> das componentes principais Avaliação dos <i>loadings</i> para cada CP rotacionada

Quadro 4.1 - Informações sobre estudos em a ACP foi aplicada a dados de qualidade de água superficial (continuação)

Referência	Área de estudo	Objetivos	Resumo da aplicação de ACP
Singh et al. (2020)	Rio Kali, Índia	Identificar a principal fonte de poluição do rio, examinar os atributos das diferenças espaço-temporais entre o rio principal que flui através das cidades e regiões rurais e quantificar os níveis de contaminação da poluição da água do rio. Foram utilizadas nas análises ACP, AC e índices de qualidade da água.	17 estações de monitoramento Aplicação de ACP e rotação das componentes principais ACP aplicada ao banco de dados dividido em 2: verão e inverno Avaliação dos <i>loadings</i> e agrupamento de estações de amostragem

^aANOVA – Análise de Variância

^bAC – Análise de Cluster

^cAF – Análise Fatorial

^dAD – Análise Discriminante

^eACC – Análise de Correlação Canônica

Fonte: Próprio autor.

Apesar do grande número de aplicações, 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, Chappell e Loftis (2012), sobre a aplicação de ACP a amostras de água, os autores ressaltaram que 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, 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 utilizada com diversos objetivos (BRO; SMILDE, 2014). Nos estudos revisados, verificou-se sua aplicação para a avaliação de variações temporais e espaciais, com maior foco dado às análises de *loadings* e à interpretação de bancos de dados considerando as informações de várias estações simultaneamente, localizadas em uma região, sub-bacia ou curso de água (avaliações globais de qualidade de água).

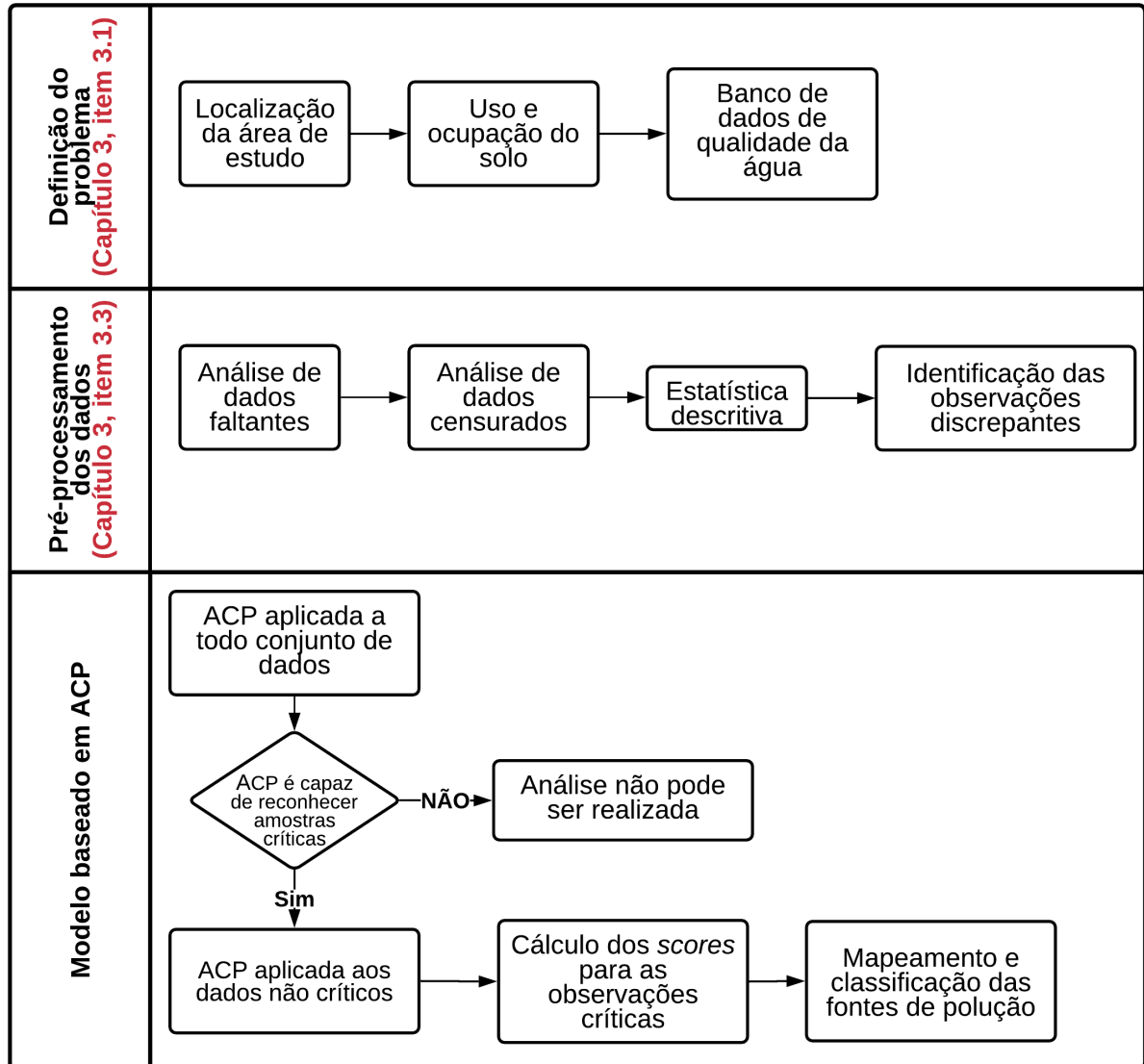
Neste contexto, este estudo tem como objetivo a utilização de ACP, em combinação com o filtro de Hampel, para desenvolver um modelo para o mapeamento e classificação de fontes de poluição determinadas pelos valores críticos dos parâmetros de qualidade de água monitorados. O presente trabalho propõe uma 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. A metodologia proposta foi aplicada 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).

Este capítulo está organizado da seguinte forma: a Seção 4.2 descreve a metodologia utilizada, incluindo a área de estudo e o conjunto de dados usado, uma breve discussão da Análise por Componentes Principais e a construção do modelo de ACP. Em seguida, apresentam-se e discutem-se os resultados na Seção 4.3, e por fim, as conclusões são apresentadas na Seção 4.4.

4.2 METODOLOGIA

O fluxograma da metodologia proposta neste trabalho, para o mapeamento e a classificação das fontes de poluição a partir da análise de ACP, é apresentado na Figura 4.1. Cada etapa é descrita detalhadamente a seguir.

Figura 4.1 - Fluxograma da metodologia adotada para o mapeamento e a classificação das fontes de poluição



Fonte: Próprio autor.

4.2.1 Definição do problema

A primeira etapa da metodologia consistiu em realizar a definição do problema. No capítulo 3, item 3.1, estão apresentadas as informações sobre a localização da

área de estudo, o uso e ocupação do solo e o monitoramento de qualidade da água. A definição do banco de dados utilizado neste estudo também está apresentada no Capítulo 3, item 3.2.

4.2.2 Pré-processamento dos dados

A segunda etapa da metodologia – o pré-processamento dos dados - foi apresentada no capítulo 3, item 3.3.

4.2.3 Definição de modelo baseado em ACP para o monitoramento de condições críticas de qualidade da água

A terceira etapa consistiu na obtenção do modelo ACP para o mapeamento de condições críticas de qualidade da água. A teoria completa sobre a ACP foi apresentada no Capítulo 2 - Revisão Teórica da Análise por Componentes Principais.

Nessa etapa, a identificação dos valores discrepantes, realizada a partir da aplicação do filtro de Hampel, foi utilizada (Capítulo 3, item 3.4.2). Para uma determinada amostra, se ocorreu pelo menos um valor discrepante 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 (74 registros) 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 (72 registros) resultantes foram usadas como comportamento de referência para os parâmetros de qualidade da água.

4.2.3.1 Reconhecimento de observações críticas

Inicialmente, como os parâmetros de qualidade da água possuem diferentes unidades de medida, para evitar interferências nos resultados, os dados foram padronizados em escala z (média 0 e desvio-padrão 1) (SERGEANT et al., 2016).

Em seguida, identificou-se um modelo ACP com o conjunto completo de observações, críticas e não críticas. Os gráficos de *scores*, considerando todas as

combinações possíveis de CPs, foram analisados com o objetivo de se verificar a capacidade do modelo em separar, de modo espacial, as observações segundo a classificação de Hampel. Essa possibilidade de usar ACP para agrupamento forma a base para a classificação subsequente das fontes de poluição.

4.2.3.2 Mapeamento e classificação das fontes de poluição

Dada a capacidade da ACP de identificar amostras críticas e não críticas, um segundo modelo de ACP foi obtido usando apenas amostras não críticas – um modelo de referência. A designação “modelo de referência” não significa uma condição de qualidade de água satisfatória, ou seja, uma condição de referência não perturbada, mas caracteriza amostras com os valores dos parâmetros dentro de um determinado limite, definido pelo filtro de Hampel, que serão utilizadas para detectar mudanças na faixa de valores dos parâmetros.

O modelo de referência foi alimentado com as amostras críticas, as quais, inicialmente, foram padronizadas usando as médias dos parâmetros e desvios-padrão obtidos com as amostras não críticas. Os *scores* foram calculados e plotados nos gráficos de *scores* em relação à região de referência fornecida pelo modelo ACP.

Analisaram-se, então, os gráficos de *scores* considerando as combinações possíveis entre as três primeiras CPs. 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 (Figura 3.5). 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%.

Todas as análises estatísticas realizadas foram implementadas em linguagem de programação Python (<https://www.python.org/>), versão 2014, e R (<https://www.r-project.org/>), versão 4.0.3, ambos de licença livre e código aberto.

4.3 RESULTADOS E DISCUSSÃO

4.3.1 Definição do banco de dados de monitoramento de qualidade da água e pré-processamento dos dados

Os resultados referentes às etapas de definição do banco de dados e pré-processamento estão apresentados no Capítulo 3, item 3.4.

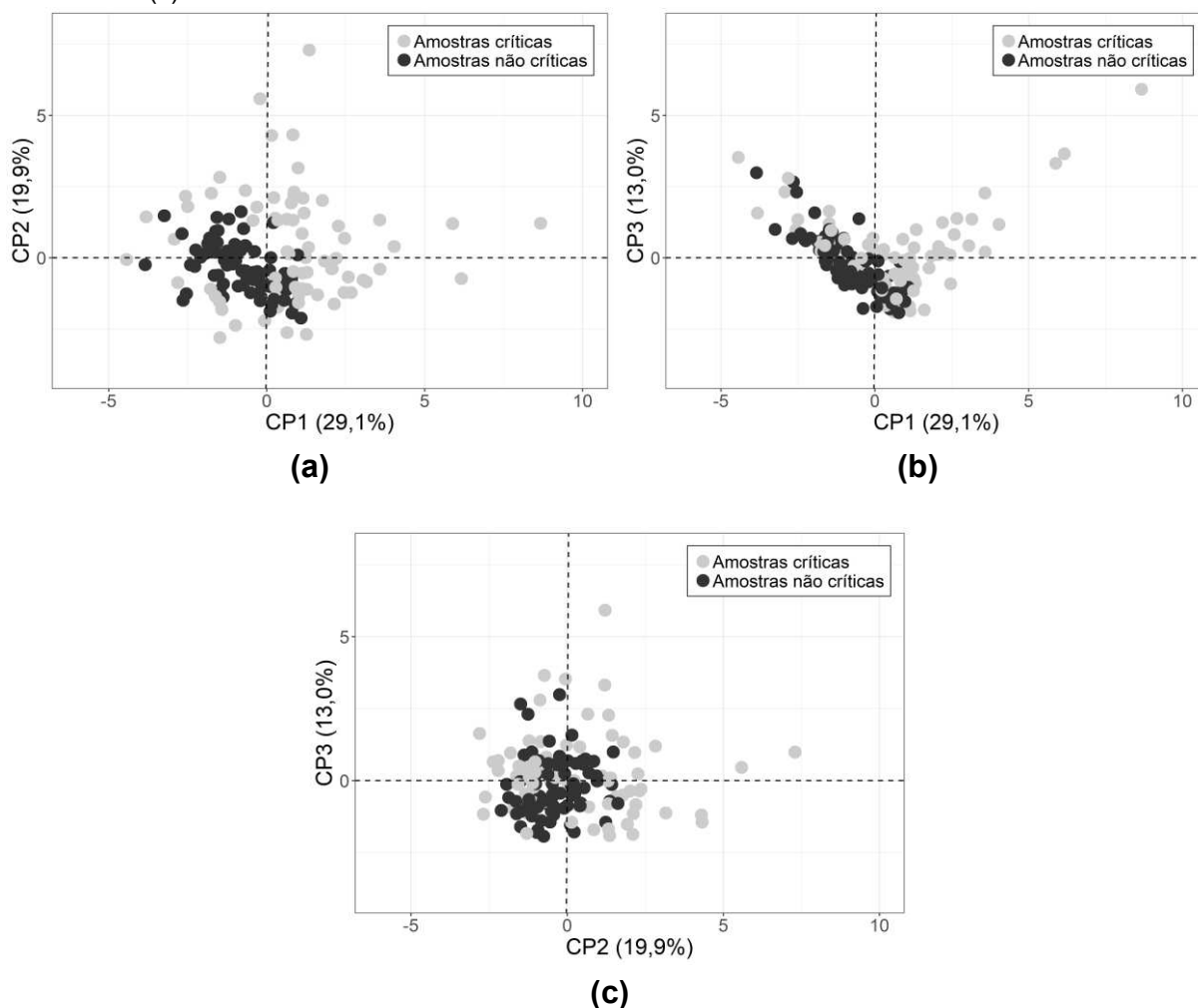
4.3.2 Definição de modelo baseado em ACP para o monitoramento de condições críticas de qualidade da água

4.3.2.1 Reconhecimento de observações críticas

O primeiro modelo de ACP foi construído com o conjunto completo de observações 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.

A Figura 4.2 (a-c) mostra gráficos de *scores* para as componentes principais mais relevantes (CP1, CP2 e CP3), com as respectivas variâncias explicadas entre parênteses. Os resultados são análogos para todas as combinações de pares possíveis. Em geral, observam-se os dados não críticos concentrados numa nuvem de pontos (círculos de cor preta), e os dados críticos no entorno deles, concentrados ou dispersos (círculos de cor cinza). 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.

Figura 4.2 - Gráficos de *scores* para as combinações das componentes principais: (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3

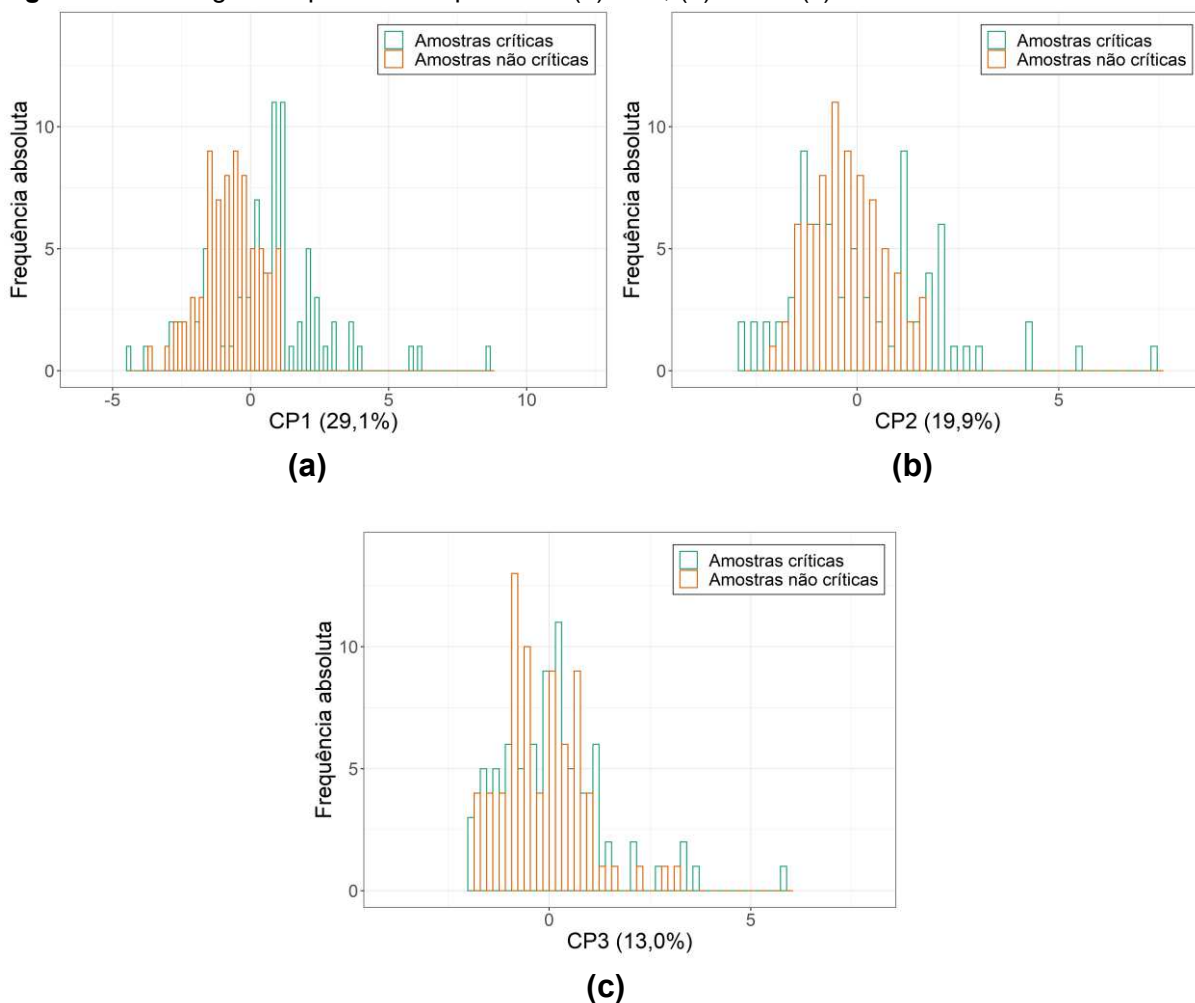


Fonte: Próprio autor.

A Figura 4.3 (a-c) apresenta os histogramas para as componentes CP1, CP2 e CP3, respectivamente. Pode-se observar a diferença entre as distribuições de frequência dos *scores* dos dados críticos e não críticos, principalmente para CP1. As observações classificadas como críticas resultam de valores elevados em pelo menos um dos parâmetros de qualidade de água, sendo um indicativo de uma condição de qualidade de água muito ruim.

A apresentação desses resultados ilustra a possibilidade da utilização de ACP para a detecção de condições críticas de qualidade de água, e consequente identificação das fontes de poluição em atividades de monitoramento de qualidade de água. O ponto principal é que um modelo de ACP de referência pode detectar mudanças na estrutura de correlação espacial dos parâmetros de qualidade da água.

Figura 4.3 - Histogramas para as componentes (a) CP1, (b) CP2 e (c) CP3



Fonte: Próprio autor.

4.3.2.2 Mapeamento e classificação das fontes de poluição

Portanto, confirmada a capacidade de ACP em discriminar as amostras críticas e não críticas, um segundo modelo de ACP, baseado somente nas observações não críticas, foi construído. Os resultados estão resumidos na Tabela 4.1, que apresenta a variância em termos absolutos e relativos para todas as componentes principais (CP1, CP2, ..., CP12), bem como a variância explicada acumulada e as cargas (*loadings*) para todos os parâmetros de qualidade da água para cada componente.

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 4.1 - Resultados do modelo ACP obtido a partir de observações 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

Fonte: Próprio autor.

Uma vez que o modelo de ACP foi obtido usando apenas amostras não críticas, seus *scores* podem ser usados como região de referência. Como os dados foram padronizados antes da identificação do modelo, um ponto (*score*) na origem do novo sistema de coordenadas significa valores dos parâmetros em seus respectivos valores médios. Essa região de referência foi envolta por uma elipse de concentração (ou elipse de dados), com nível de confiança de 95% (refere-se à elipse nos gráficos de *scores*). Pode-se observar que a nuvem de dados não críticos se localiza, de modo geral, dentro do limite de controle, conforme desejado.

Em seguida, os *scores* para as observações críticas, utilizando esse modelo ACP de referência, foram calculados e plotados no gráfico de *scores* em relação à região de referência, delimitada pelo limite de controle. Então, três regiões (1, 2 e 3) foram localizadas no conjunto de dados críticos, devido a diferenças relevantes entre um subconjunto de parâmetros distintos entre elas. Como essas três regiões apresentam valores muito elevados de alguns parâmetros de qualidade de água (no caso do parâmetro oxigênio dissolvido, valores muito baixos), elas foram visíveis em todas as combinações possíveis de CPs, dados os valores dos *loadings*. Para cada

uma das regiões, são apresentados os gráficos de *scores* para os pares de combinações entre as três primeiras componentes (CP1-CP2, CP1-CP3 e CP2-CP3) (Figuras 4.4, 4.6 e 4.8), que serão detalhados a seguir. Nessas figuras, os círculos na cor preta representam as amostras não críticas (região de referência); os círculos vazios, amostras críticas em geral; e os círculos na cor vermelha, amostras críticas referentes às regiões espaciais identificadas nos gráficos.

A região 1 (em vermelho) localiza-se em torno do limite de controle no plano CP1-CP2 (Figura 4.4a). Nos planos CP1-CP3 (Figura 4.4b) e CP2-CP3 (Figura 4.4c), essas amostras também se encontram além do limite de controle. Observa-se que não há uma relação significativa com CP2.

Figura 4.4 - Gráficos de *scores* com destaque na região 1 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3

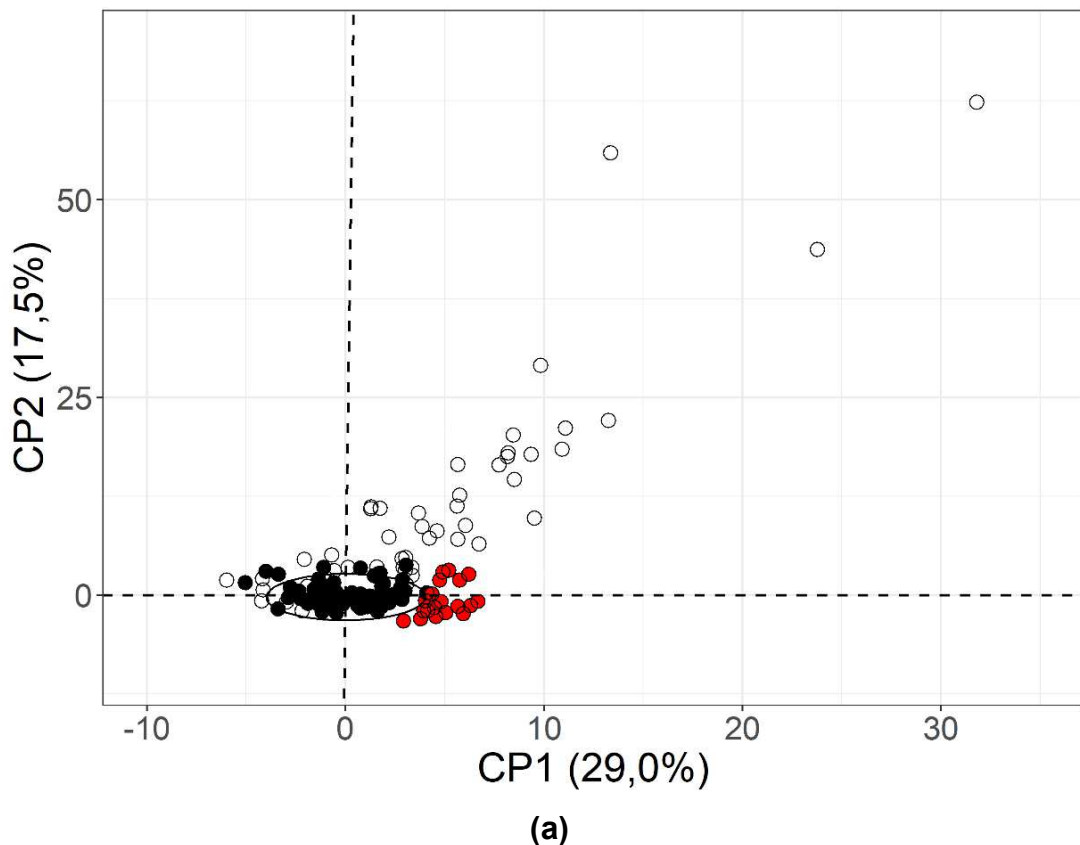
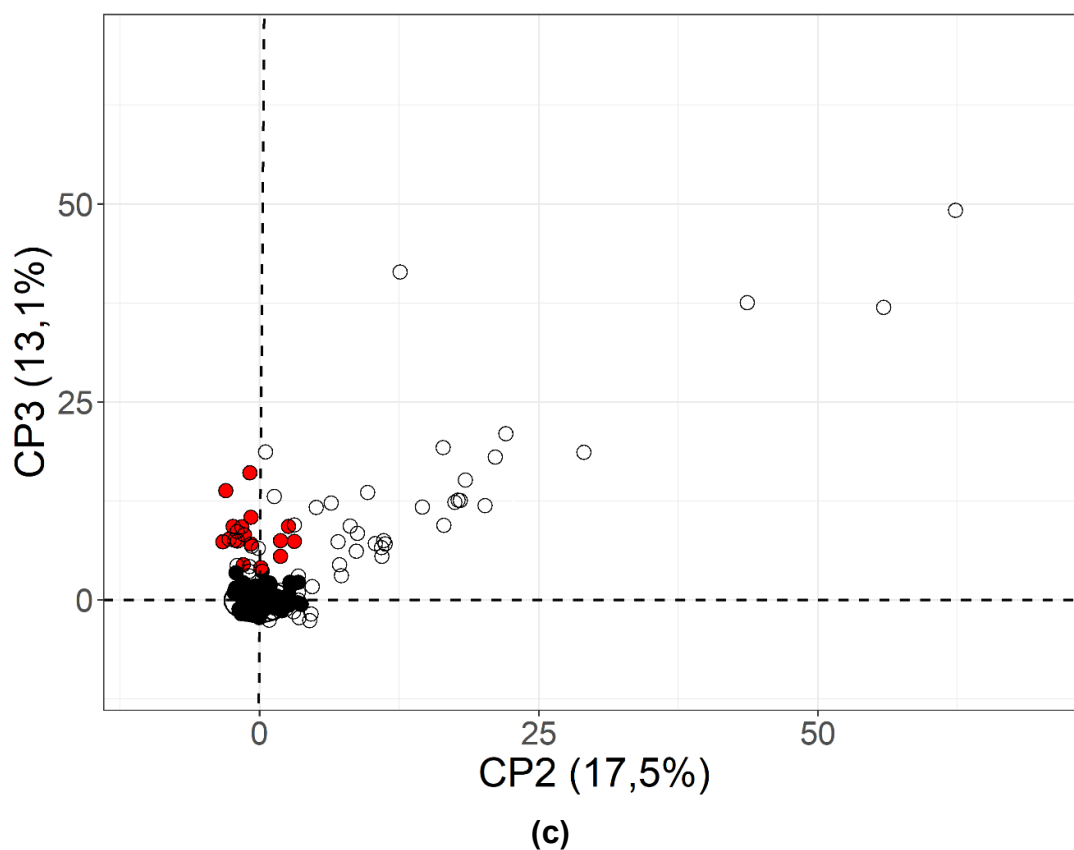
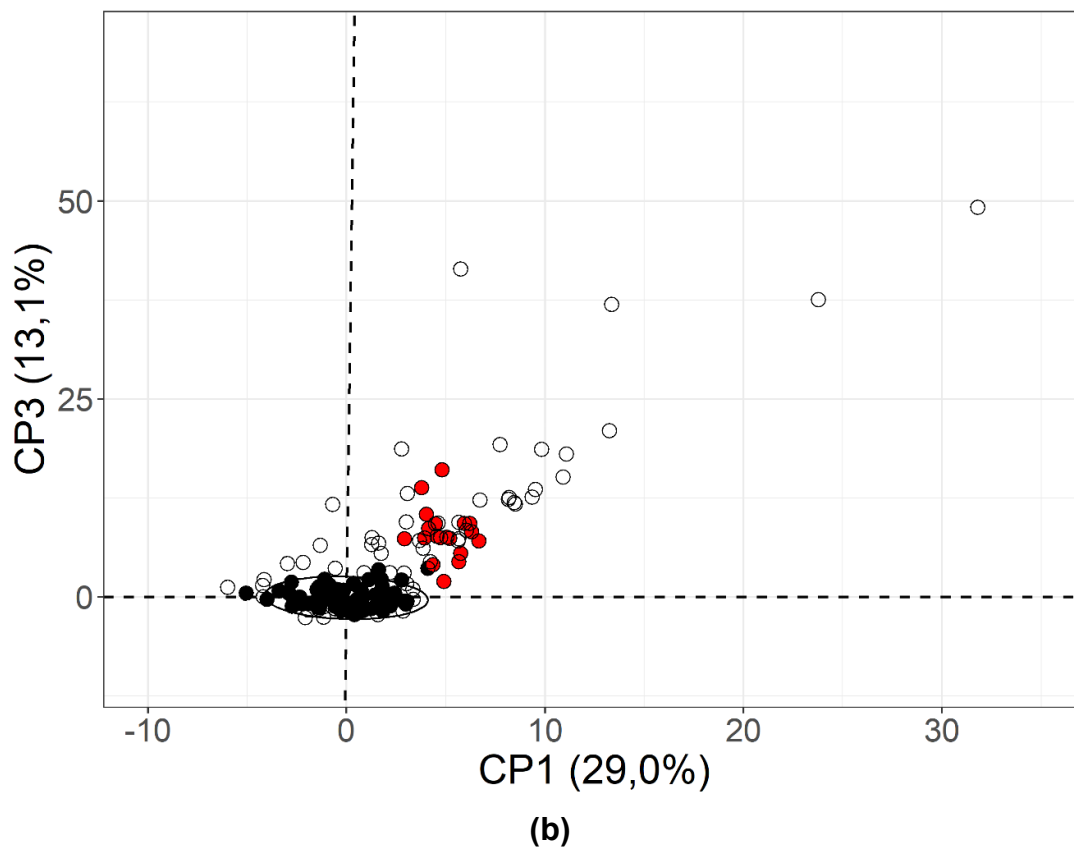


Figura 4.4 - Gráficos de scores com destaque na região 1 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3 (continuação)



Fonte: Próprio autor.

As observações da região 1 foram analisadas separadamente, e na Tabela 4.2 é apresentada a estatística descritiva dessas observações. São observados valores relativamente altos de coliformes termotolerantes e DBO, e relativamente baixos de oxigênio dissolvido.

Tabela 4.2 - Resumo das estatísticas básicas dos parâmetros de qualidade da água para as observações da região 1, entre 1997 e 2018

Parâmetros	Mínimo	Máximo	Média	Desvio-padrão	Percentil 25	Mediana	Percentil 75	CV ^a (%)
As _T	0,004	0,071	0,035	0,017	0,024	0,033	0,044	48,3
Cl ⁻	7,47	24,03	14,44	4,83	10,15	13,73	18,15	32,5
Colif. term.	30000	160000	149474	33078	160000	160000	160000	21,5
CE	156	333	242	55	194	258	283	22,2
DBO	2,40	40,00	12,29	9,99	6,45	9,00	14,00	79,1
P _T	0,110	0,970	0,425	0,272	0,215	0,360	0,455	62,3
N-NO ₃ ⁻	0,010	2,060	0,392	0,530	0,105	0,170	0,425	131,7
OD	0,50	4,60	2,26	1,32	1,15	2,10	3,10	56,8
pH	6,00	7,50	6,93	0,37	6,75	6,90	7,20	5,2
S _T	184	451	249	63	205	237	280	24,5
T _{água}	18,00	28,80	23,42	2,94	21,30	23,20	24,70	12,2
Turb.	12	209	79	58	45	65	105	70,5

^aCV – Coeficiente de variação

Fonte: Próprio autor.

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, os valores medianos das concentrações de Colif. term. (Figura 4.5a) e DBO (Figura 4.5b) são significativamente superiores (valor-p<0,05) em relação às respectivas medianas da região de referência, e o valor mediano das concentrações de OD (Figura 4.5c) é significativamente inferior (valor-p<0,05).

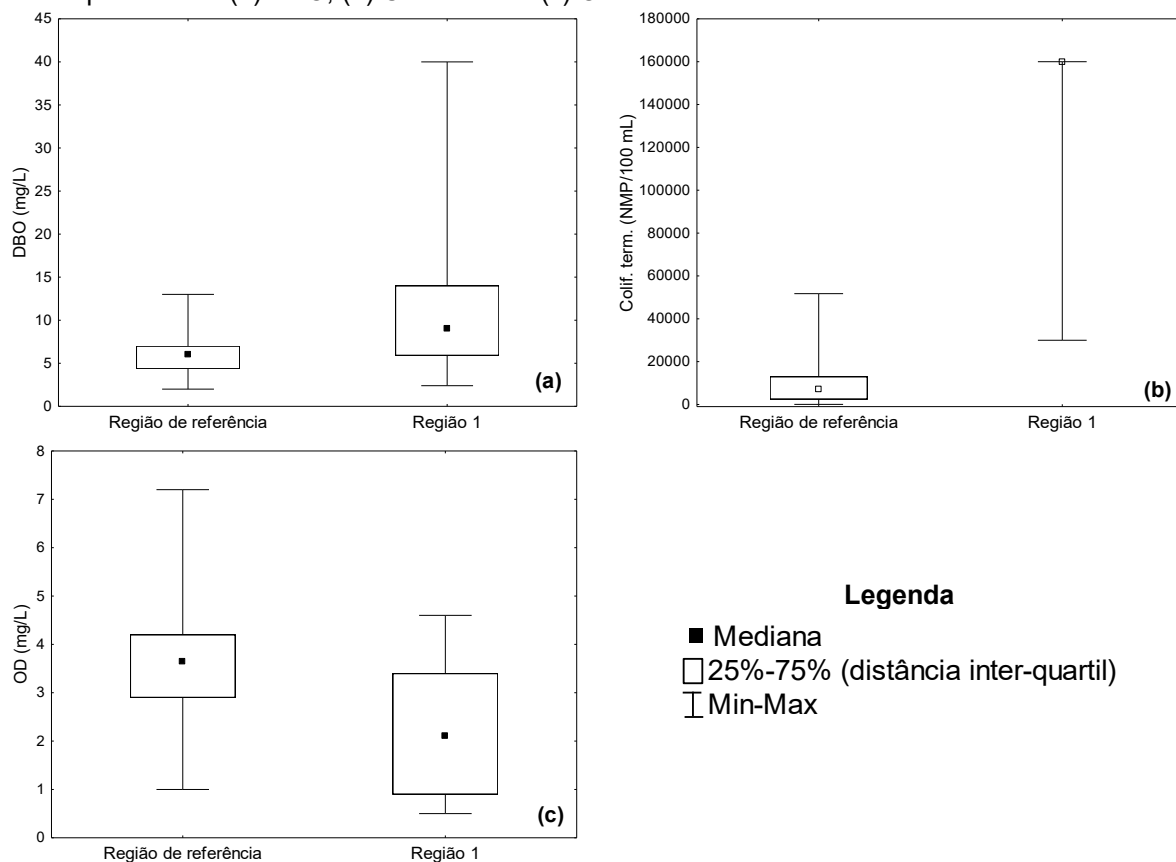
Essa condição é um indicativo, sobretudo, dos lançamentos de efluentes domésticos não tratados ou tratados de forma insuficiente nos cursos de água. Como a estação BV137 está localizada a jusante de municípios da RMBH, ela recebe influência dos lançamentos de esgotos de grandes centros urbanos, como por exemplo, Sabará, Belo Horizonte, Contagem, Santa Luzia e Lagoa Santa (Figura 3.5), sendo um trecho crítico em termos de poluição (IGAM, 2010, 2018). Verificou-se que todas as amostras da região 1 identificadas, à exceção de duas, apresentaram valores de Colif. term. superiores a 160.000 NMP/100 mL (limite máximo de detecção do

método analítico).

O município imediatamente a montante da estação BV137 é o de Lagoa Santa (que faz parte da RMBH). De acordo com levantamento realizado pela ANA, esse município, que possui uma população urbana de 52.520 habitantes (censo demográfico de 2010 do Instituto Brasileiro de Geografia e Estatística - IBGE), gera uma carga orgânica total estimada, associada ao esgoto, de 2.898,3 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 encaminhada para soluções precárias de esgotamento (ANA, 2017a). 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 ao abastecimento humano, à saúde pública e ao equilíbrio do meio ambiente (ANA, 2017b).

Numa avaliação do IQA (Índice de Qualidade de Água), que é um índice de referência particularmente sensível à contaminação por esgotos e outros materiais orgânicos, nutrientes e sólidos, a estação BV137 foi classificada na classe ruim nos anos de 2017 e 2018 (IGAM, 2018, 2019a).

Figura 4.5 - 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



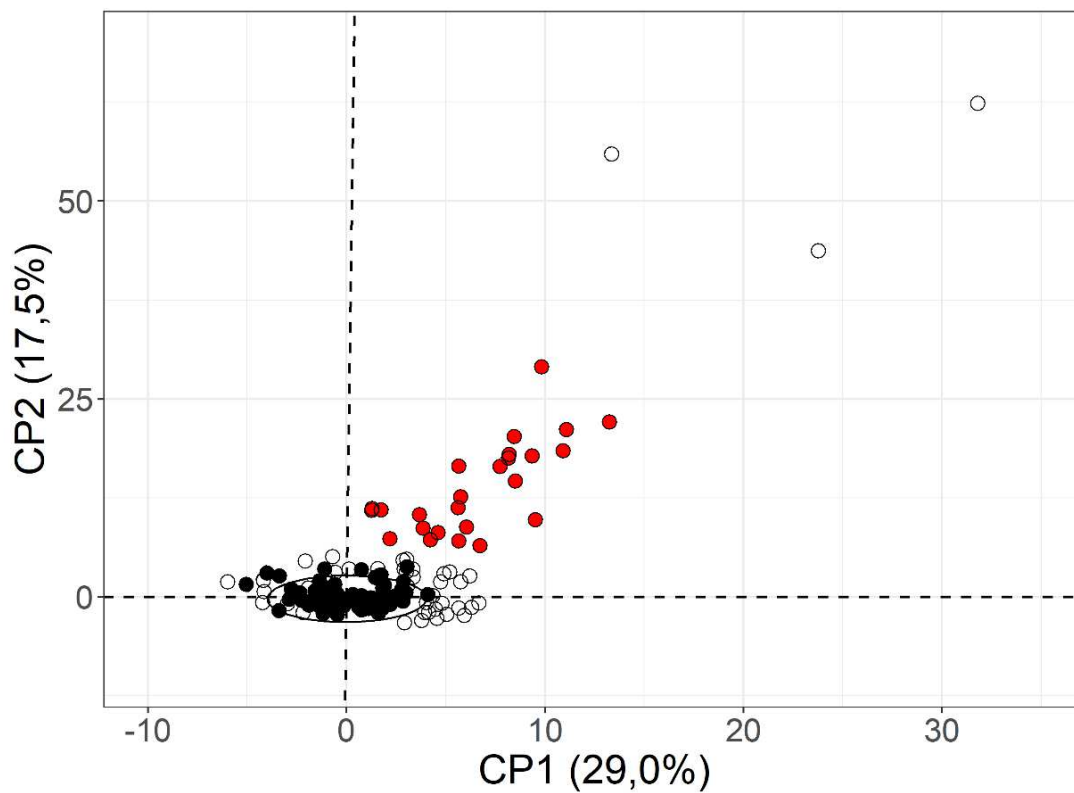
Fonte: Próprio autor.

Os municípios localizados a montante da estação BV137 também contribuem com os lançamentos de efluentes industriais (principalmente indústrias de abate de animais, curtume, fabricação de adubos e fertilizantes, indústrias alimentícias e de laticínios), que resultam 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 (IGAM, 2013).

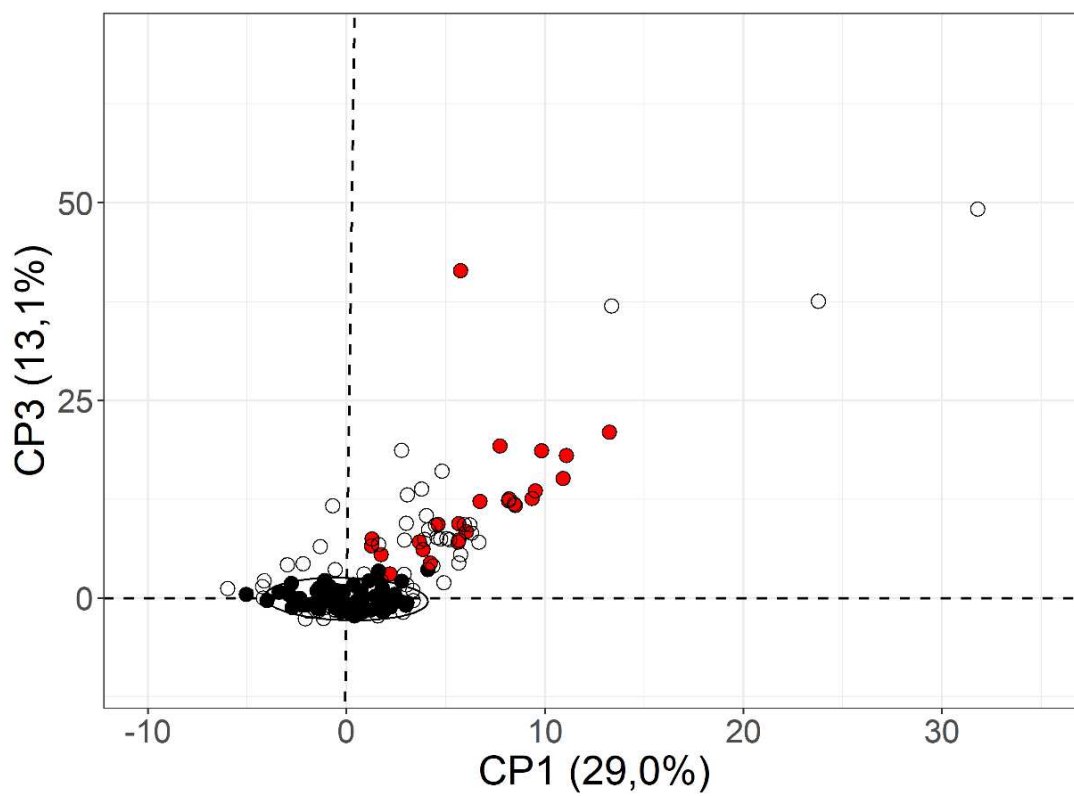
Alguns dos municípios com as condições de qualidade mais críticas, identificados em um estudo do Igam na bacia do rio das Velhas (IGAM, 2013), estão localizados a montante da estação BV137, justificando, assim, os resultados de Colif. term., DBO e OD encontrados nessa estação.

A Figura 4.6 (a-c) 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 para a identificação das fontes de poluição seriam CP1-CP2 e CP2-CP3.

Figura 4.6 - Gráficos de scores com destaque na região 2 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3

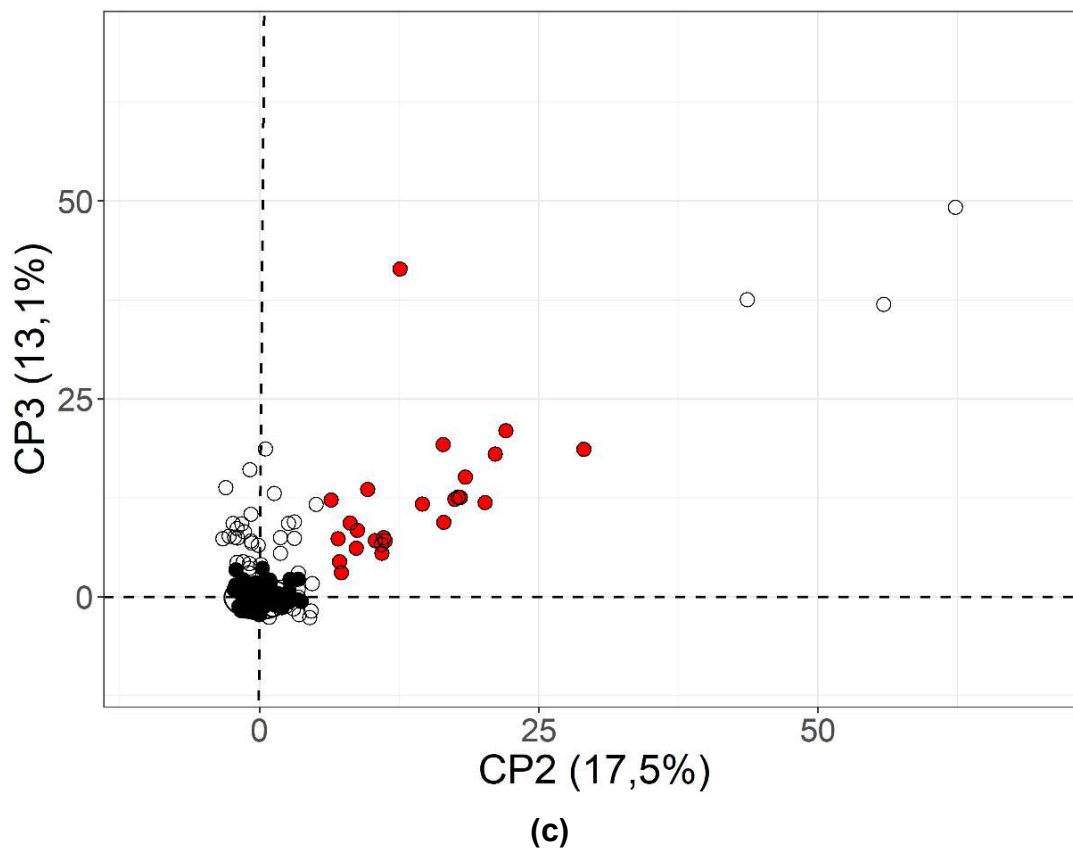


(a)



(b)

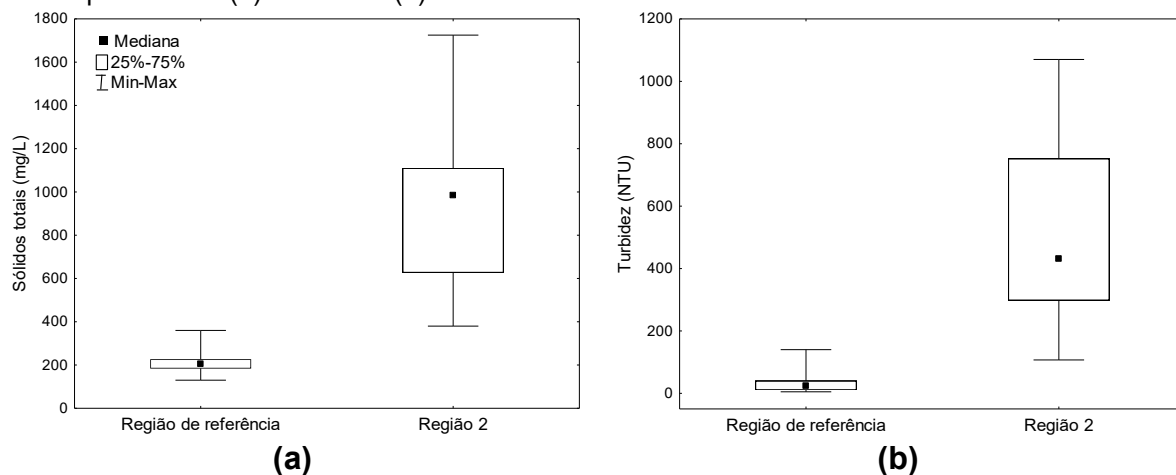
Figura 4.6 - Gráficos de scores com destaque na região 2 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3 (continuação)



Fonte: Próprio autor.

Na região 2, os valores medianos de turbidez e sólidos totais são significativamente superiores (valor- $p < 0,05$) aos respectivos valores medianos da região de referência (Figura 4.7). Esses resultados podem estar relacionados, principalmente, às atividades agropecuárias, que são predominantes nesta área, conforme apresentado no Quadro 3.2 e no mapa de uso e ocupação do solo da área de estudo (Figura 3.5). Essas atividades contribuem para intensificar os processos erosivos que, em combinação com ação do escoamento pluvial, 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 4.7 - *Box-plot* para comparação das concentrações entre a região 2 e a região de referência para os parâmetros (a) turbidez e (b) sólidos totais



Fonte: Próprio autor.

A Figura 4.8 (a-c) apresenta a região 3, que é um exemplo de uma condição extrema, contendo três observações críticas. Essas três observações críticas são as mais afastadas do conjunto de dados de referência, se comparadas às demais. Essa região é característica de concentrações elevadas de arsênio total, sólidos totais e turbidez, em relação à situação usual. Dada a magnitude dessa discrepância, esse evento extremo é reconhecível em qualquer dos planos de componentes principais. Na prática, caso não fosse necessário utilizar a terceira componente principal, dever-se-ia optar pelo plano CP1-CP2, de modo a reduzir o número de gráficos de *scores* a analisar.

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

A região do alto rio das Velhas compreende parte da área denominada Quadrilátero Ferrífero (CBH VELHAS, 2015a), que possui uma anomalia natural de arsênio. O intemperismo das rochas provoca a liberação desse elemento para os sedimentos, água superficial e subterrânea, mas o aumento da sua concentração no ambiente é agravado por atividades de mineração desenvolvidas, sobretudo, na região do Alto Velhas. Além disso, as fontes de contaminação por arsênio, locais e regionais, também estão relacionadas a passivos ambientais (solos contaminados, pilhas de rejeitos antigos, minas antigas), devido às antigas atividades de mineração na região (em razão da presença de depósitos auríferos sulfetados), e à presença de sedimentos enriquecidos com esse elemento nos cursos de água (BORBA;

FIGUEIREDO; MATSCHULLAT, 2003; CHRISTOFARO; LEÃO, 2009; COSTA et al., 2015; DESCHAMPS; MATSCHULLAT, 2007; IGAM, 2010).

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 pela Companhia de Pesquisa de Recursos Mineral – CPRM (2010), também foram relatadas altas concentrações de arsênio em amostras de sedimentos do fundo do rio na área do Quadrilátero Ferrífero.

Figura 4.8 - Gráficos de scores com destaque na região 3 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3

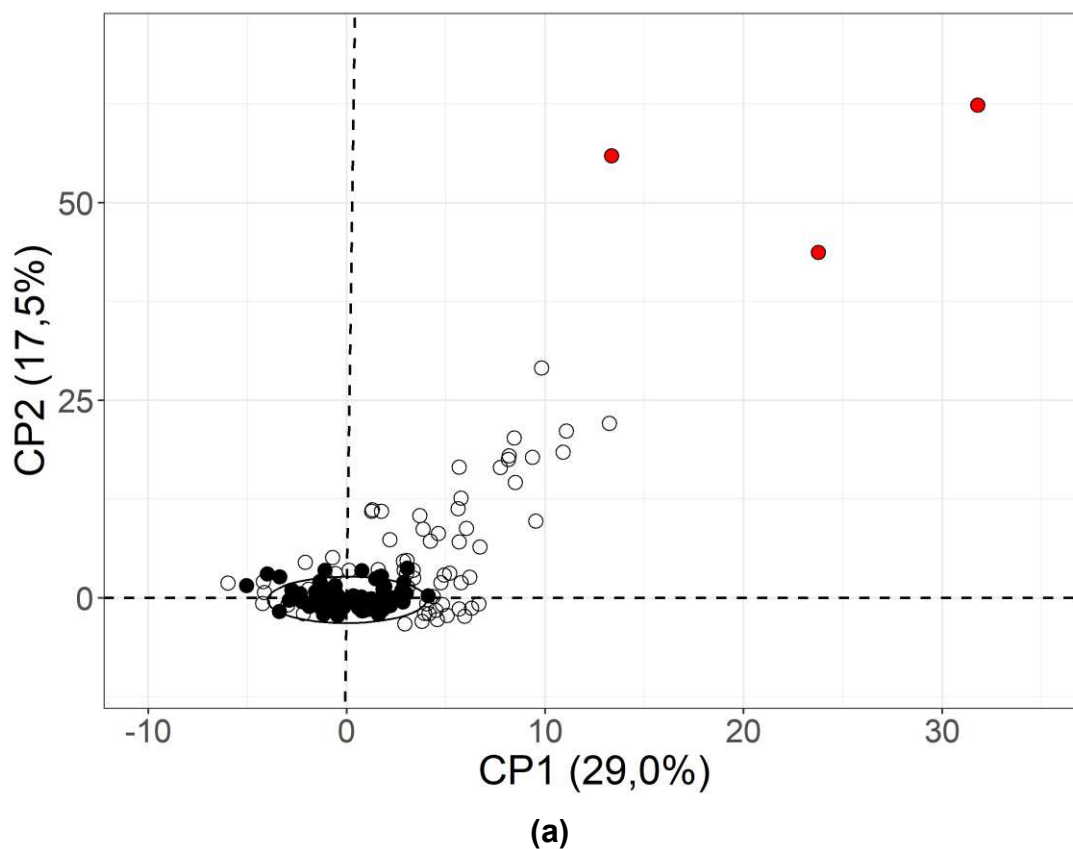
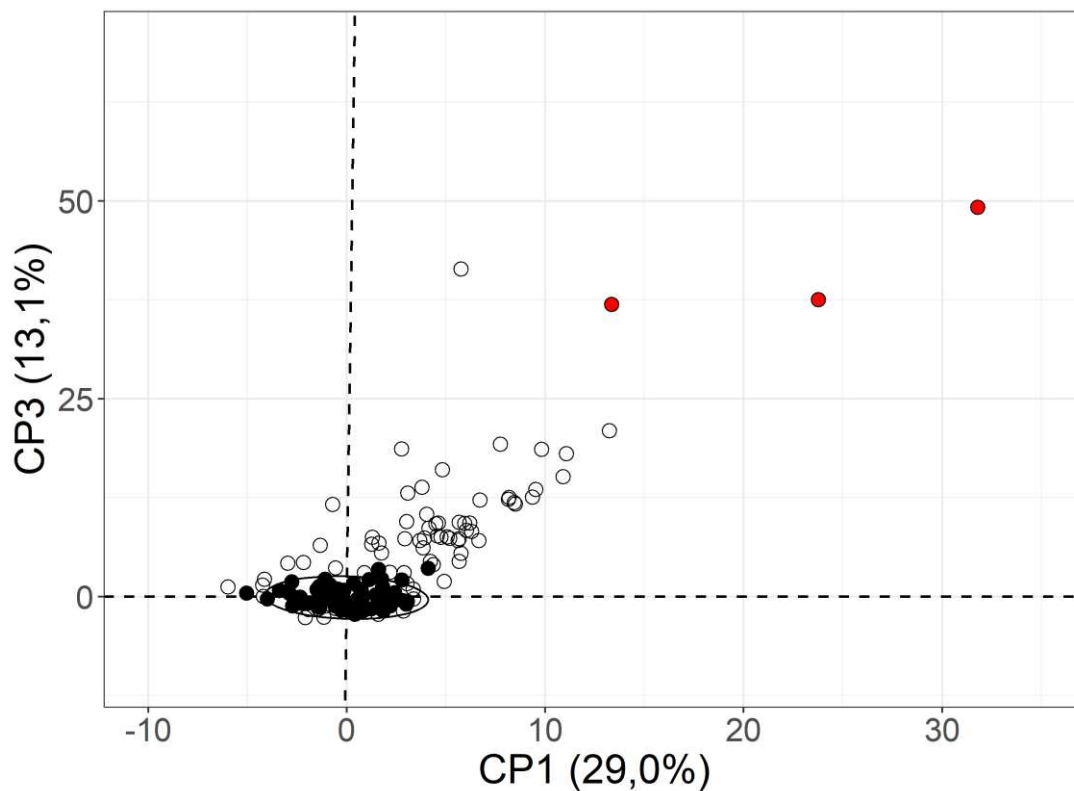
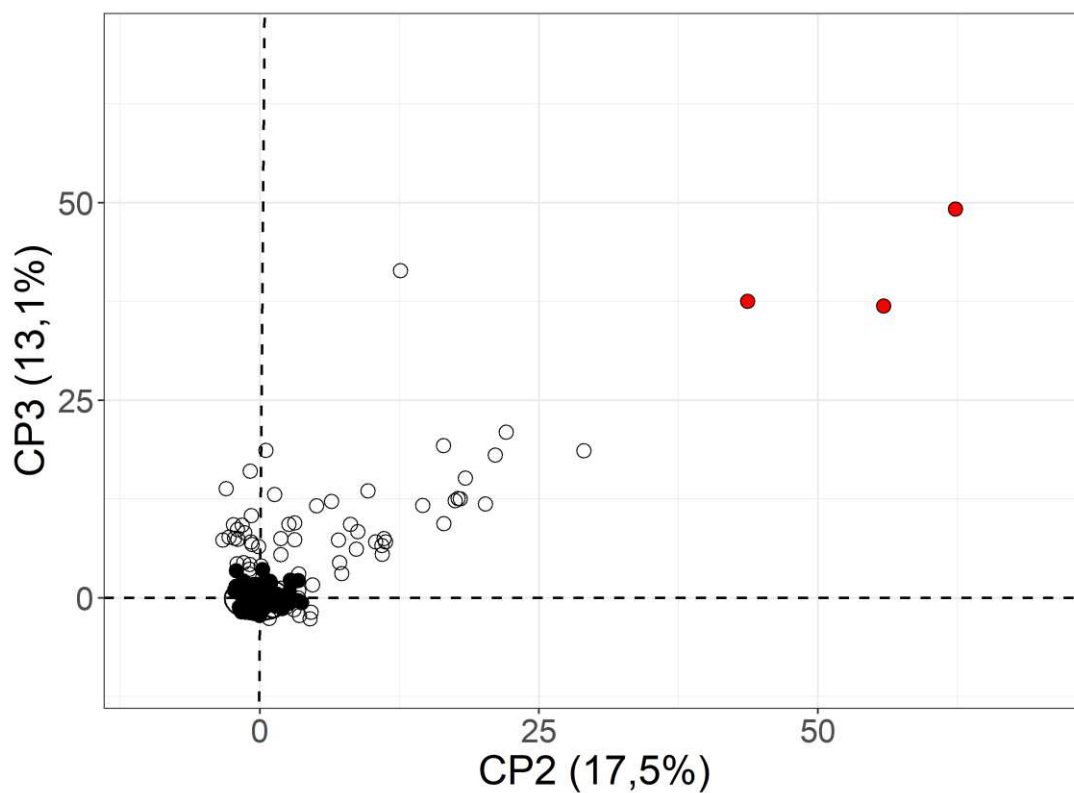


Figura 4.8 - Gráficos de scores com destaque na região 3 para as combinações das componentes (a) CP1-CP2, (b) CP1-CP3 e (c) CP2-CP3 (continuação)



(b)



(c)

Fonte: Próprio autor.

Em um estudo realizado por Calazans et al. (2018b), os autores apontaram a estação BV137 como uma das mais relevantes na calha do rio das Velhas e prioritária para manutenção na rede de monitoramento. Nesse estudo, os autores aplicaram análise de cluster hierárquica 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 do presente trabalho.

Além disso, as fontes de poluição ambiental neste ponto do rio das Velhas corroboram 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. A vantagem desse tipo de análise é o monitoramento a partir de dois ou três gráficos, ao invés, por exemplo, de se ter um gráfico para cada parâmetro (nesse caso, doze), como comumente é realizado. Outra questão é quanto à consideração de correlação espacial, ao se ter uma análise conjunta dos parâmetros, o que é geralmente desconsiderado.

4.4 CONCLUSÕES

Os bancos de dados gerados com os Programas de Monitoramento da Qualidade das Águas (PMQAs) constituem-se numa rica fonte de informação que pode ser explorada a partir da aplicação da Análise por Componentes Principais.

A ACP é uma ferramenta útil a ser empregada em sistemas de suporte à decisão ao monitoramento de qualidade de água, podendo ser utilizada com diferentes objetivos. Neste estudo, explorou-se o uso de um modelo baseado em ACP, associado ao filtro de Hampel. Os gráficos de *scores* foram utilizados 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 e a associação de regiões espaciais nos gráficos de *scores* com eventos críticos de poluição em relação à qualidade da água. As regiões identificadas nos gráficos de *scores* foram associadas a fontes de poluição presentes na área de estudo, nomeadamente descargas de efluentes domésticos e industriais,

atividades agropecuárias e minerárias. 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.

As avaliações realizadas de uma estação de monitoramento demonstraram a importância da abordagem local para um estudo detalhado do trecho de um curso de água. Enquanto uma análise global fornece características gerais de uma sub-bacia, a análise local revela detalhes sutis sobre o comportamento de um determinado trecho de rio. Na prática, a combinação das abordagens global e local podem levar a um melhor entendimento sobre as condições de qualidade da água e, portanto, ao estabelecimento de políticas ambientais mais eficazes. A metodologia utilizada neste estudo pode ser aplicada a outras estações de monitoramento.

**CAPÍTULO 5:
MONITORAMENTO DE
QUALIDADE DE ÁGUA
SUPERFICIAL POR MEIO DE
CONTROLE ESTATÍSTICO DE
PROCESSO MULTIVARIADO**

RESUMO

A deterioração cada vez mais séria da qualidade das águas superficiais tem levado ao desenvolvimento e à utilização de diferentes metodologias e/ou ferramentas estatísticas para avaliação da qualidade da água e gerenciamento sustentável. Neste estudo, uma abordagem combinada que integra Controle Estatístico de Processo Multivariado e Análise por Componentes Principais (CEPM-ACP) foi utilizada para detecção de condições críticas da qualidade da água e fornecimento de alerta, de acordo com os dados históricos de observação. Primeiro, um modelo de ACP foi usado para reduzir a dimensionalidade do problema. Em seguida, as métricas de monitoramento T^2 e Q foram aplicadas e os limites das cartas de controle foram definidos. O modelo desenvolvido foi então aplicado para avaliar as variações da qualidade da água de superfície para alerta em uma estação de monitoramento na calha do rio das Velhas. O banco de dados, composto por doze parâmetros de qualidade da água, consistiu em amostras coletadas ao longo de vinte e um anos (1997 a 2018). O modelo de ACP obtido foi capaz de explicar 77,46% da variância total dos dados originais utilizando somente cinco componentes principais. Utilizou-se o filtro de Hampel para a classificação das observações em discrepantes e não discrepantes, sendo esse o critério para a distinção das condições críticas e não críticas de qualidade de água. A aplicação das cartas de controle para as estatísticas T^2 e Q mostraram-se promissoras para a detecção de condições críticas, podendo servir como sistemas de suporte à decisão em atividades de monitoramento e de gerenciamento de qualidade de água. Assim, decisões estratégicas mais eficientes, em relação à proteção e gestão dos corpos de água, podem ser definidas e implementadas.

Palavras-chave: Monitoramento; Qualidade de água; CEP Multivariado; Análise por Componentes Principais; Estatísticas T^2 e Q .

5.1 INTRODUÇÃO

A qualidade das águas superficiais e subterrâneas sofrem uma deterioração cada vez mais séria devido, principalmente, às atividades humanas intensas, como a descarga de efluentes doméstico e industriais, o uso de agrotóxicos, e as mudanças no uso e cobertura do solo, com consequências específicas para a saúde humana e para o ecossistema (SHI et al., 2018; UNEP; GEMS WATER, 2008; ZHANG et al., 2010). Portanto, é essencial prevenir e controlar a poluição da água e implementar programas regulares de monitoramento (ZHAO et al., 2012).

Por causa das variações espaciais e temporais na qualidade da água, esses programas de longo prazo são necessários para fornecer uma base de dados confiável e representativa da qualidade da água superficial (SIMEONOV et al., 2003; ZHANG et al., 2010).

Normalmente, os programas de monitoramento da qualidade da água incluem a medição de múltiplos parâmetros físicos, químicos e biológicos em diversas estações de monitoramento, resultando em grandes e complexos conjuntos de dados em um período relativamente curto, que podem ser difíceis de entender e inferir para obter informações significativas (CHAPMAN et al., 1992; SIMEONOV et al., 2003). Como usar eficientemente esses conjuntos de dados, para traçar ações, planos e acompanhamento dos dados registrados, é de particular interesse.

Mais do que monitorar a qualidade dos cursos de água, interpretar essas informações e comunicá-las aos tomadores de decisão é imprescindível para que decisões de gestão sejam tomadas e políticas regionais e nacionais sejam definidas (SPRAGUE et al., 2019). Assim, em relação à atividade de monitoramento e diagnóstico, o uso de diferentes abordagens e técnicas estatísticas para analisar dados de qualidade da água podem contribuir para o conhecimento sobre a variabilidade espaço-temporal da qualidade da água e identificação de fontes de poluição, de modo a qualificar melhor o processo de tomada de decisão dos gestores (SIMEONOV et al., 2003; WUNDERLIN et al., 2001).

Nos últimos anos, as cartas de controle univariadas, uma das técnicas de Controle Estatístico de Processo (CEP), e com longa história de uso nas indústrias (MONTGOMERY; RUNGER, 2015), têm sido aplicadas para monitoramento de qualidade de água em diversos estudos.

CEP UNIVARIADO

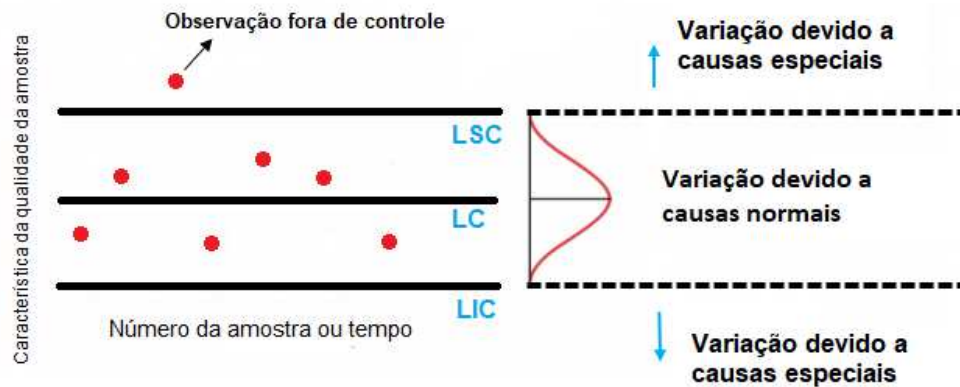
O CEP é uma abordagem essencialmente para entender e controlar a variabilidade do processo e tem sido amplamente utilizada nos campos industrial e não industrial (NOSKIEVIČOVÁ, 2013).

O gráfico de controle é o principal instrumento do CEP para a análise dessa variabilidade ao longo do tempo. Eles são uma representação gráfica da variabilidade do processo e de seus padrões naturais e não naturais (NOSKIEVIČOVÁ, 2013). Existem duas fontes de variabilidade: causas comuns (casual) e causas especiais (atribuídas). As causas comuns dizem respeito à variabilidade natural que sempre existe em qualquer processo e não há como evitá-la. Já as causas especiais não são uma característica inerente ao processo e, portanto, podem ser identificadas e eliminadas (MONTGOMERY, 2016).

A Figura 5.1 apresenta um gráfico típico de controle, que exibe um valor da característica de qualidade (medida ou calculada a partir de uma amostra) *versus* o número da amostra ou tempo. Observam-se três linhas: a linha central (LC), que representa o valor médio da característica da qualidade, e duas outras linhas horizontais que representam o limite superior de controle (LSC) e o limite inferior de controle (LIC) (MONTGOMERY, 2016).

Os LSC e LIC separam as causas comuns e causas especiais de variação. O intervalo entre esses limites define a variação comum. Então, diz-se que um processo está fora de controle estatístico quando uma observação fica fora dos limites de controle (REID; SANDERS, 2013). As causas comuns estão associadas à variabilidade natural ou inerente ao processo, enquanto as causas especiais dizem respeito à variabilidade que não faz parte do padrão de causas comuns (MONTGOMERY; RUNGER, 2015).

Figura 5.1 - Visão geral de um gráfico de controle univariado



Fonte: Adaptado de Reid e Sanders (2013).

Em essência, o gráfico de controle é um teste da hipótese de que o processo está em um estado de controle estatístico (MONTGOMERY; RUNGER, 2015):

$$\begin{cases} H_0: \text{Processo sob controle estatístico.} \\ H_1: \text{Processo fora de controle estatístico.} \end{cases}$$

Um ponto em um gráfico dentro dos limites de controle é equivalente a não rejeitar a hipótese de controle estatístico, e um ponto fora dos limites de controle é equivalente a rejeitar a hipótese de que o processo está sob controle estatístico (MONTGOMERY, 2016).

O modelo de um gráfico de controle é apresentado nas Equações (5.1), (5.2) e (5.3) (MONTGOMERY, 2016).

$$LSC = \mu_w + k\sigma_w \quad (5.1)$$

$$LC = \mu_w \quad (5.2)$$

$$LIC = \mu_w - k\sigma_w \quad (5.3)$$

Onde w é uma estatística da amostra que mede alguma característica da qualidade de interesse, μ_w é a média de w , σ_w é o desvio-padrão de w , e k é a distância dos limites de controle a partir da linha central, expressa em unidades de desvio-padrão. Os limites de controle “3-sigma”, ou seja $k = 3$ é uma escolha comum

(MONTGOMERY; RUNGER, 2015). Os gráficos desenvolvidos de acordo com essa metodologia são chamados de gráficos de controle de Shewhart.

A construção de um gráfico de controle é dividida em duas fases. As observações iniciais são utilizadas para estimar a média populacional, bem como para estabelecer LSC e o LIC provisório ou experimental. Quaisquer pontos fora de controle para os quais as causas especiais podem ser encontradas devem ser removidos ou eliminados antes dos limites de controle de teste serem calculados. Esse primeiro período da análise de dados é chamado de fase I ou período de base. Os valores estimados na fase I são, então, usados para o período subsequente, conhecido como período de monitoramento (fase II). A média populacional estimada, obtida durante o período base, continua a ser utilizada durante este período (CORBETT; PAN, 2002; MONTGOMERY, 2016).

Os gráficos de controle são construídos com base em suposições subjacentes: dados são independentes e normalmente distribuídos com média constante μ e variância constante σ^2 (GILBERT, 1987). No estudo realizado por Korzenowski (2009), foi proposto um *framework* com a organização de procedimentos e técnicas estatísticas para a verificação das suposições dos gráficos de controle.

De acordo com Gilbert (1987), os gráficos de controle são ferramentas gráficas que indicam quando uma mudança nos padrões dos dados ao longo do tempo deve ser examinada para determinar as causas, sendo úteis, então, por fornecerem uma base para a ação. Em um gráfico de controle, mesmo que todos os pontos estejam dentro dos limites de controle, é possível identificar se os dados se comportam de maneira não aleatória ou sistemática, sendo este último uma indicação de que o processo não está sob controle (MONTGOMERY, 2016).

O processo de reconhecimento de padrões não naturais (por exemplo, grandes mudanças na média, tendências, variação sistemática) pode ser realizado por meio de testes estatísticos (regras) que quantificam o comprimento desses padrões. Eles são baseados em cálculos de probabilidade sobre a porção de pontos próximos à linha central, próximos aos limites de controle, etc (NOSKIEVIČOVÁ, 2013). O primeiro conjunto de regras foi formulado pela Western Electric, e a partir de então, diferentes outras regras surgiram, como, por exemplo, regras de Nelson, testes ISO 2859 e conjuntos especiais definidos por grandes empresas (como Boeing). Noskievičová

(2013) realizou uma análise comparativa de conjuntos de regras e suas propriedades estatísticas básicas.

Além dos gráficos de controle de Shewhart, outras cartas de controle também foram desenvolvidas, por exemplo: o gráfico de controle da soma cumulativa (CUSUM) e o gráfico de controle para a média móvel exponencialmente ponderada (MMEP). Cada um deles tem vantagens e desvantagens (MONTGOMERY; RUNGER, 2015).

Dentre os estudos de qualidade de água que utilizaram os gráficos de controle univariados, pode-se citar: Arciszewski et al. (2018) usaram cartas de controle para detectar mudanças nos parâmetros de química da água e destacaram o valor da ferramenta nos programas de monitoramento de longo prazo; Samsudin et al. (2017) aplicaram carta de controle de Shewhart e o cálculo do índice de capacidade de processo para monitoramento da qualidade de água e verificar o risco de poluição da água; Bhasin et al. (2016) utilizaram o gráfico de controle Shewhart para análise do estado da poluição da água durante diferentes estações do ano e em diferentes localidades; Cruz; Oliveira e Filho (2016) empregaram cartas de controle tradicionais e uma metodologia geoestatística para monitoramento do parâmetro pH em um curso de água; Iglesias et al. (2016) exploraram as cartas de controle do tipo Shewhart para controle estatístico de processo com o objetivo de identificar valores anormais de um índice global de qualidade de água; Sancho et al. (2016) usaram gráficos de medidas individuais para analisar os dados de qualidade da água e detectar valores anômalos de parâmetros que possam indicar a ocorrência de episódios passados de poluição; Lee et al. (2013) propuseram um gráfico de controle individual com limites de controle variáveis para monitorar a qualidade do rio; Follador et al. (2012) utilizaram as cartas de controle do tipo Shewhart (gráfico individual), CUSUM (soma cumulativa) e MMEP (média móvel exponencialmente ponderada), no controle de qualidade da água superficial, e a carta MMEP mostrou-se mais eficaz para os dados analisados; Zhou et al. (2008) utilizaram cartas de controle para monitorar parâmetros de qualidade de água em nascentes cársticas; e Zimmerman et al. (1996) aplicaram cartas de medidas individuais para identificar mudanças nos parâmetros de qualidade da água em um estuário.

Como o monitoramento da qualidade de água envolve a análise simultânea de diversas variáveis, usualmente correlacionadas (DU et al., 2017; MORRISON, 2008),

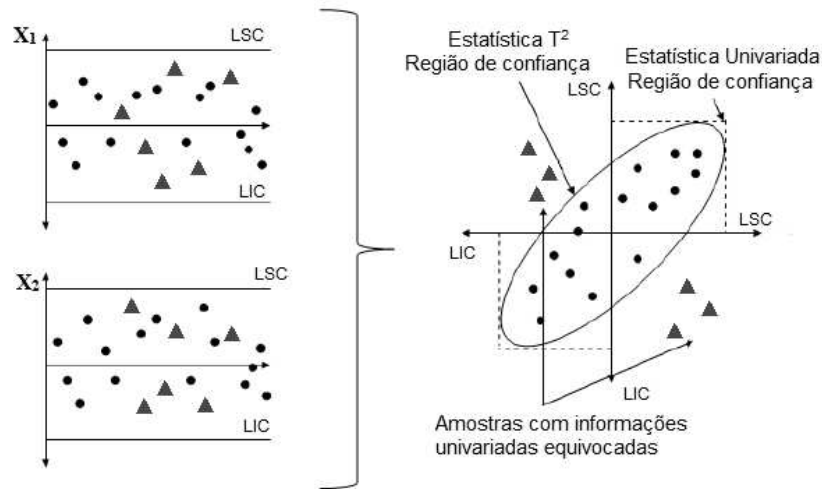
a utilização das cartas de controle univariadas torna-se ineficaz e pode levar a conclusões errôneas (MONTGOMERY, 2016). Quando são utilizadas cartas univariadas, cada variável (parâmetro de qualidade de água) possui um limite e é analisada individualmente, sem considerar as informações contidas nas outras variáveis (RUSSELL; CHIANG; BRAATZ, 2000), que podem ser muito importantes para o entendimento do comportamento do processo (CHEN, 2001). Sabe-se que quando ocorrem eventos especiais em um processo, eles afetam não apenas a magnitude das variáveis, mas também a relação entre elas (FERRER, 2007).

Segundo Camacho et al. (2016), à medida que o número de variáveis a serem monitoradas aumenta, a abordagem para visualizar uma variável de cada vez reduz seu desempenho; então é mais conveniente usar estatísticas e gráficos multivariados, como por exemplo, as cartas de controle multivariadas.

Diante do exposto, como nos programas de monitoramento de qualidade de água muitos parâmetros físicos, químicos e biológicos são monitorados, técnicas estatísticas multivariadas receberam maior atenção em pesquisas recentes, conforme explanado no Capítulo 1.

A Figura 5.2 mostra a diferença entre as abordagens univariada e multivariada, ilustrando o problema do uso de gráficos de controle separados para duas variáveis de qualidade. De acordo com a avaliação independente de cada carta de controle univariada, o processo está sob controle estatístico (círculos e triângulos entre o LIC e o LSC). Porém, segundo a carta de controle multivariada, quando as variáveis são analisadas simultaneamente, há observações sob controle (círculos) e fora de controle estatístico (triângulos). Esse fato é devido à consideração da correlação espacial entre ambas as variáveis (representada pela elipse). A elipse delimita um limite de confiança de $100(1 - \alpha)\%$ da distribuição conjunta, ou seja, quando o processo está sob controle, uma fração α dos pontos ficará fora da elipse (KOURTI, 2002; MONTGOMERY, 2016).

Figura 5.2 - Comparação entre as abordagens de cartas de controle univariada (à esquerda) e multivariada (à direita)



Fonte: Adaptado de Ordóñez (2008)

CEP MULTIVARIADO

O Controle Estatístico de Processo Multivariado (CEPM) é uma extensão do CEP univariado para controlar diversas variáveis ao mesmo tempo (CAMACHO et al., 2016). A ideia principal da abordagem de monitoramento baseada em CEPM é extrair as informações úteis de um conjunto de dados e construir algumas estatísticas para o monitoramento (GE; SONG, 2013). Ele tem sido amplamente utilizado em diferentes campos da ciência: matemática, medicina, química, processos biológicos, dentre outros (ORDÓÑEZ, 2008). O CEPM tornou-se muito popular, principalmente, aplicado ao monitoramento de processos industriais (GE; SONG, 2013). Nesses novos ambientes ricos em dados, os métodos multivariados são importantes por facilitarem a visualização simultânea de muitas variáveis (FERRER, 2007).

A depender da abordagem, o Controle Estatístico de Processo Multivariado pode ter diferentes classificações, por exemplo, a classificação em CEPM convencional e CEPM baseado em variáveis latentes (FERRER, 2007).

O CEPM Convencional é baseado no desenvolvimento de gráficos de controle da estatística T^2 de Hotelling baseados nas variáveis originais registradas (FERRER, 2007), e incluem, principalmente, modelos de controle multivariado de Shewhart, modelos de controle multivariados de CUSUM e modelos multivariados de controle da MMEP (QIU, 2014).

Já CEPM baseado em variáveis latentes utiliza uma metodologia que explora a estrutura de correlação das variáveis originais, revelando os poucos eventos

subjacentes independentes (variáveis latentes) que estão conduzindo o processo a qualquer momento (FERRER, 2007). Embora o CEPM convencional seja bem visto do ponto de vista estatístico, ele sofre de falta de aplicabilidade em ambientes ricos em dados, típicos de processos modernos (FERRER, 2007), além disso a presença de variáveis correlacionadas podem levar à falta de precisão e problemas computacionais devido ao mau condicionamento (KOURTI; MACGREGOR, 1996).

De acordo com Camacho et al. (2016), quando o número de variáveis é muito grande e/ou as variáveis são altamente inter-relacionadas, o uso de métodos de variáveis latentes (ou métodos de projeção) dentro do CEPM é recomendado. Exemplo de método de projeção estatística multivariada é a Análise por Componentes Principais (ACP), uma das principais ferramentas utilizadas no CEPM (CAMACHO et al., 2016; GE; SONG, 2013). ACP é capaz de lidar com processos de alta dimensão contendo variáveis correlacionadas (GE; SONG, 2013). A partir dessa ferramenta, é possível reduzir a dimensionalidade do espaço de monitoramento, projetando as informações nas variáveis originais em subespaços de baixa dimensão definidos por algumas variáveis latentes (JACKSON, 2003). Em CEPM-ACP, é comum monitorar duas estatísticas: a estatística T^2 de Hotelling, que é calculada a partir dos *scores*, e a estatística Q , relacionada aos resíduos (FERRER, 2007; GE; SONG, 2013). A teoria completa de ACP foi apresentada no Capítulo 2.

Na literatura, há diversos estudos da aplicação de ACP a dados de qualidade de água superficial com o foco, principalmente, em *loads*, para alcançar diferentes objetivos, como: identificação de parâmetros com maior contribuição sobre variações espaço-temporal da qualidade da água, detecção de principais fontes de poluição, estimativa da contribuição relativa de possíveis fontes para a concentração dos parâmetros, dentre outros (BHAT et al., 2014; SIMEONOV et al., 2003; SINGH et al., 2004; ZHANG et al., 2010). No Capítulo 4, Quadro 4.1, é apresentado um levantamento de estudos em que a ACP foi aplicada a dados de qualidade de água superficial.

O presente estudo, portanto, utiliza a técnica ACP combinada ao Controle Estatístico de Processo Multivariado (CEPM-ACP), e visa o desenvolvimento desta abordagem para monitorar, de modo simultâneo, um conjunto de parâmetros de qualidade de água para a detecção de eventos críticos de poluição, ou seja, detectar valores anormais de qualquer parâmetro de qualidade da água. O modelo ACP

mapeia as variáveis em um espaço dimensional inferior e rastreia seu comportamento usando as métricas de monitoramento (estatísticas T^2 e Q).

Este capítulo está organizado da seguinte forma: a seção “5.2 Metodologia” apresenta a área de estudo e análises estatísticas preliminares, além das etapas executadas para a aplicação de CEPM-ACP para o monitoramento da qualidade da água. A seção “5.3 Resultados e Discussão” descreve a aplicação da metodologia no local de monitoramento selecionado (estação BV137, localizada na bacia do rio das Velhas), e o desempenho na detecção de condições críticas de poluição é analisado. A última seção apresenta as conclusões do estudo.

5.2 METODOLOGIA

Este trabalho foi desenvolvido utilizando dados, disponibilizados pelo Igam, do monitoramento da estação BV137, localizada na bacia do rio das Velhas.

5.2.1 Área de estudo

A área de estudo está detalhadamente descrita no Capítulo 3, item 3.1

5.2.2 Banco de dados

As informações sobre o banco de dados bruto da estação de monitoramento de qualidade da água localizada na área de estudo (estação BV137) e as análises realizadas para a definição do banco de dados de trabalho estão descritas no Capítulo 3, no item 3.2.

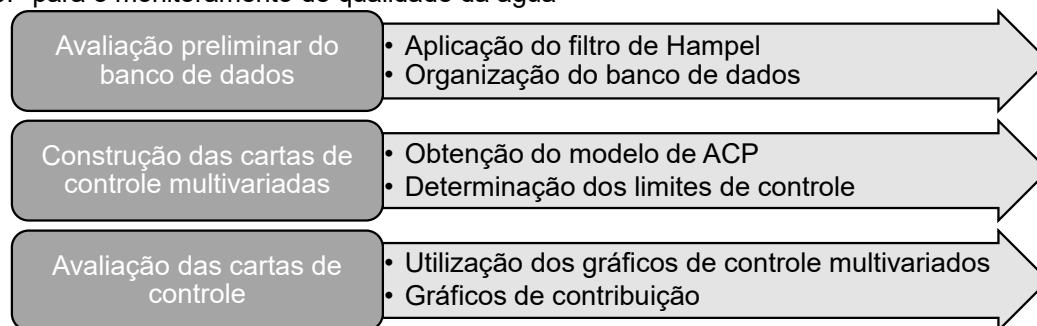
5.2.3 Pré-processamento dos dados

Todas as análises realizadas na etapa de pré-processamento dos dados estão apresentadas no Capítulo 3, item 3.3.

5.2.4 Monitoramento de qualidade da água com Controle Estatístico de Processo Multivariado baseado em ACP (CEPM-ACP)

A metodologia utilizada para o desenvolvimento da abordagem CEPM-ACP para monitorar, de modo simultâneo, os parâmetros de qualidade da água, está apresentada na Figura 5.3. Na sequência, é apresentado um detalhamento de cada etapa.

Figura 5.3 - Metodologia utilizada para o desenvolvimento da abordagem de CEP Multivariado baseado em ACP para o monitoramento de qualidade da água



Fonte: Próprio autor.

5.2.4.1 Etapa 1 – Avaliação preliminar do banco de dados

O filtro de Hampel foi aplicado aos dados, conforme apresentado no item 3.4.2, e os dados de cada parâmetro foram classificados em dois grupos (discrepantes e não discrepantes). A partir disso, para uma dada amostra, se o valor de pelo menos um dos parâmetros se apresentou acima dos limites (dado discrepante), toda a amostra foi classificada como crítica e passou a ser identificada com um rótulo (*label*) do grupo correspondente. As amostras classificadas como não críticas (72 registros) foram consideradas como a base de referência (*baseline*) para essa estação de monitoramento, e as amostras críticas (74 registros) foram utilizadas como indicativo de uma condição extrema de qualidade da água em termos de poluição. Empregou-se essa classificação para a aplicação do CEPM-ACP.

Em seguida, baseado nessa classificação, o conjunto de dados foi dividido em três subconjuntos, sendo dois deles (X_1 e X_2), contendo apenas amostras não críticas (representando uma condição de referência), utilizados para a identificação inicial do modelo baseado em ACP (usado como sistema de monitoramento) e o cálculo dos limites superiores de controle, respectivamente. O terceiro subconjunto de dados (X_3), contendo amostras não críticas restantes e críticas, foi utilizado para teste.

5.2.4.2 Etapa 2 – Construção das cartas de controle multivariadas

Obtenção do modelo de ACP

O modelo de ACP foi desenvolvido para descrever as correlações das variáveis de qualidade da água, obtidas a partir do monitoramento na calha do rio das Velhas.

Inicialmente, o banco de dados X_1 (contendo somente amostras não críticas) foi padronizado em escala-Z, com média 0 e desvio-padrão 1. A matriz de covariâncias foi calculada, e a partir dela foram obtidos o vetor de autovalores ($\lambda_{[px1]}$) e a matriz de autovetores associados ($W_{[p \times p]}$), que descrevem a estrutura das variáveis originais.

Na sequência, foi usado o critério de Kaiser (1958), com exemplos de aplicação em Kowalkowski et al. (2006) e Simeonov et al. (2003), para a definição do número k de componentes principais (CP) a se reter (as componentes com autovalores menores do que 1 (um) foram excluídos). Essa abordagem de seleção de CPs é comumente usada na área de análise de qualidade de água. Desse modo, consegue-se explicar a maior parte da variância total dos dados originais a partir de um número k de componentes principais significativamente menor do que o número p de parâmetros de qualidade da água ($k \ll p$). Baseado nesse critério, foram selecionadas k componentes principais, e obteve-se um modelo com o vetor reduzido de autovalores ($\lambda_{[kx1]}$) e a matriz reduzida de autovetores $W_{[p \times k]}$, que foram utilizados como sistema de monitoramento.

Determinação dos limites de controle

Um novo conjunto de dados contendo somente amostras não críticas (X_2) foi empregado no cálculo dos limites de controle. Os valores de média e desvio-padrão, bem como o modelo ACP obtido anteriormente, foram utilizados nesta etapa. A partir do conjunto de dados padronizados em escala-Z, os *scores* foram calculados e, em seguida, com base nos k componentes principais selecionados, calculou-se a estatística T^2 de Hotelling para cada observação, utilizando a Equação (2.13) apresentada no Capítulo 2. Além desse parâmetro de controle, foi calculada também a estatística Q , que é igual a soma dos quadrados dos resíduos (Equação 2.14).

Para cada observação, tem-se um valor de T^2 e de Q , que podem ser utilizados para o cálculo dos limites de controle. Uma taxa de alarmes falsos equivalente a um nível de confiança de 95% foi utilizada para ambas as métricas de monitoramento.

Informações detalhadas sobre as estatísticas T^2 e Q estão apresentadas no Capítulo 2.

5.2.4.3 Etapa 3 – Avaliação das cartas de controle

Utilização dos gráficos de controle multivariados

Em sequência, o terceiro conjunto de dados (X_3), contendo amostras críticas e não críticas, foi utilizado para a realização de testes. Mais uma vez, os dados foram padronizados em escala-Z, baseando-se na mesma média e no mesmo desvio-padrão do modelo (X_1), e os *scores* foram calculados. Para cada observação, calcularam-se os valores de T^2 e Q , por meio das Equações (5.4) e (5.5), que são apresentados em gráficos (cartas de controle multivariadas) de T^2 versus observação e de Q versus observação. Ambos os *scores* e resíduos são monitorados em um sistema CEPM (CAMACHO et al., 2016).

As observações fora de controle foram comparadas com a classificação das amostras (realizada a partir do filtro de Hampel) para fins de validação. Dessa forma, se a observação foi classificada como crítica, os valores de T^2 e/ou Q devem estar além de seus respectivos limites de controle. Assim, a detecção da situação crítica é dada pelos pontos acima do limite de controle de T^2 e/ou Q . Valores consideravelmente altos de T^2 e/ou Q estão associados a situações extremas de poluição da água.

Gráficos de contribuição

A influência de determinado parâmetro em uma amostra pode ser obtida por meio dos gráficos de contribuição (*contribution plots*). Em outras palavras, esses gráficos são utilizados para indicar quais variáveis fazem a amostra selecionada apresentar um valor significativo de T^2 de Hotelling ou da estatística Q .

Então, quando uma amostra na estatística T^2 ou Q estava acima dos limites de controle, as contribuições de cada parâmetro de qualidade da água para essa estatística foram examinadas e foi possível identificar quais variáveis foram as mais relevantes para a condição crítica de poluição detectada.

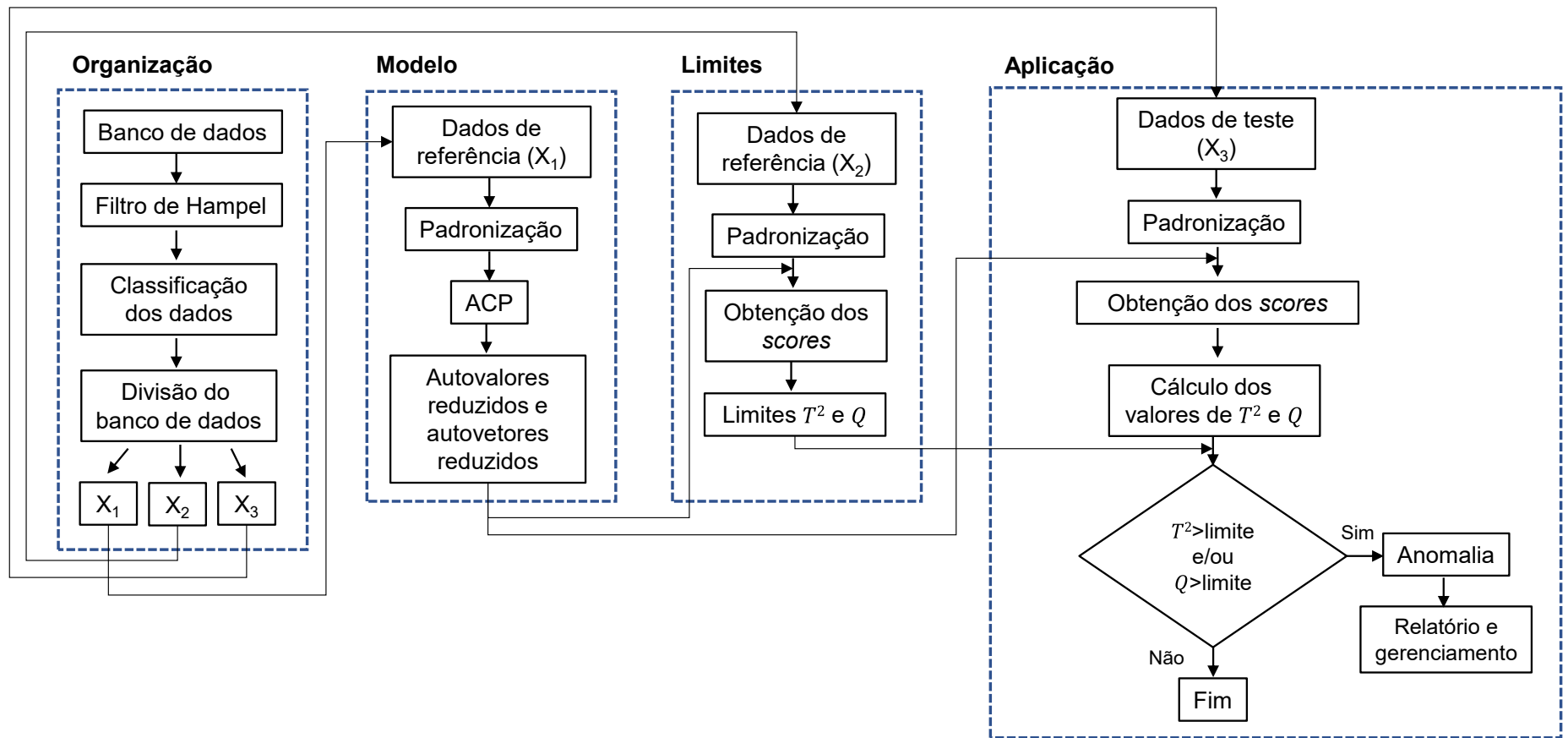
Em relação à estatística Q , o cálculo das contribuições das variáveis é realizado com base no espaço residual. A amostra acima do limite no gráfico da estatística Q representa um novo evento que não foi descrito pelo modelo de processo. Dessa forma, utilizou-se a Equação (2.15), apresentada no Capítulo 2, para calcular a contribuição da j -ésima variável no resíduo Q da i -ésima observação. Quanto maior o

valor de contribuição maior é a importância relativa daquela variável j sobre o desvio da observação i .

Após os cálculos dos percentuais de contribuição de cada parâmetro de qualidade da água na estatística de monitoramento Q , para uma amostra crítica, foi possível determinar os parâmetros mais importantes e/ou relevantes para a caracterização da qualidade das águas superficiais do trecho do curso de água avaliado, pois quanto maior a contribuição, maior a influência da variável.

O fluxograma da Figura 5.4 apresenta um resumo simplificado da metodologia adotada neste estudo, para a aplicação do CEP-ACP.

Figura 5.4 - Fluxograma da metodologia utilizada para a aplicação de CEPM-ACP no monitoramento de qualidade de água



Fonte: Próprio autor

5.3 RESULTADOS E DISCUSSÃO

5.3.1 Banco de dados e pré-processamento dos dados

Os resultados referentes às etapas de definição do banco de dados de trabalho e pré-processamento dos dados estão apresentados no Capítulo 3, itens 3.4.1 e 3.4.2, respectivamente.

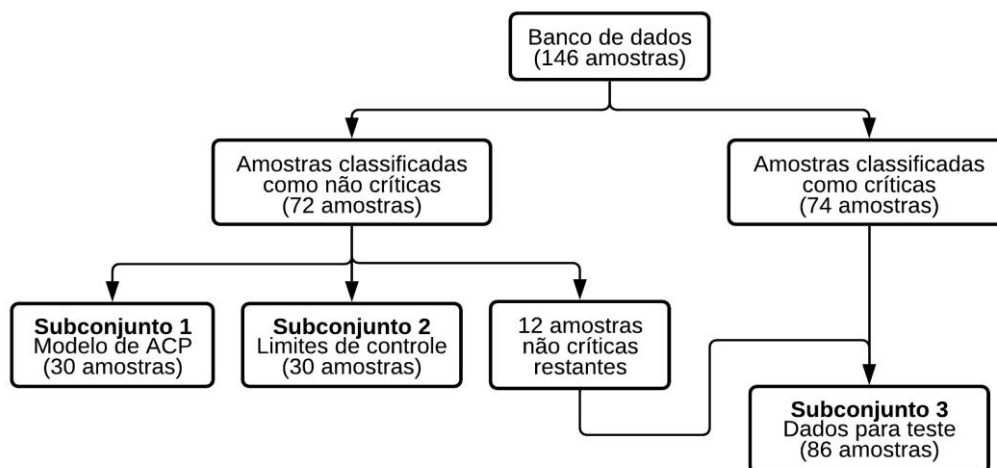
5.3.1 Monitoramento de qualidade da água com Controle Estatístico de Processo Multivariado baseado em ACP (CEPM-ACP)

O conjunto de dados de trabalho com 146 amostras foi dividido em três subconjuntos, de acordo com a metodologia descrita no item 5.2.4. O critério utilizado para a divisão do banco de dados foi a classificação dos dados a partir do filtro de Hampel. As amostras classificadas como críticas (74 registros) foram consideradas indicativas de uma condição de qualidade de água crítica em termos de poluição, e aquelas classificadas como não críticas (72 registros) foram consideradas indicativas de uma condição de referência.

Ressalta-se que a designação “condição de referência” não significa uma condição de qualidade de água satisfatória, mas representa uma condição de qualidade de água que é comumente encontrada naquele local. Quando a estação de monitoramento possui um estado de qualidade de água degradado, como é o caso da estação de monitoramento deste estudo, as concentrações dos parâmetros são elevadas para a maioria das amostras. Dessa forma, após a aplicação do filtro de Hampel, mesmo dados com concentrações relativamente elevadas foram classificadas como amostras de referência.

Na divisão do banco de dados, utilizou-se a mesma quantidade de amostras para a construção do modelo e dos limites de controle, ficando cada um desses subgrupos com um total de 30 amostras. Os dados não críticos restantes (12 amostras) foram utilizados no subconjunto de teste juntamente com os 74 registros de dados críticos. A Figura 5.5 apresenta um resumo da divisão de dados realizada.

Figura 5.5 - Divisão do banco de dados para aplicação de CEPM-ACP



Fonte: Próprio autor.

Então, construiu-se um modelo de ACP para a detecção de dados críticos com um conjunto de 30 observações não críticas (subconjunto 1). A partir do modelo de ACP obtido, observa-se que quatro CPs das doze apresentaram autovalores maiores que 1 (um) (critério de Kaiser (1958)) e explicam 69,59% da variância encontrada na qualidade das águas da bacia (Tabela 5.1).

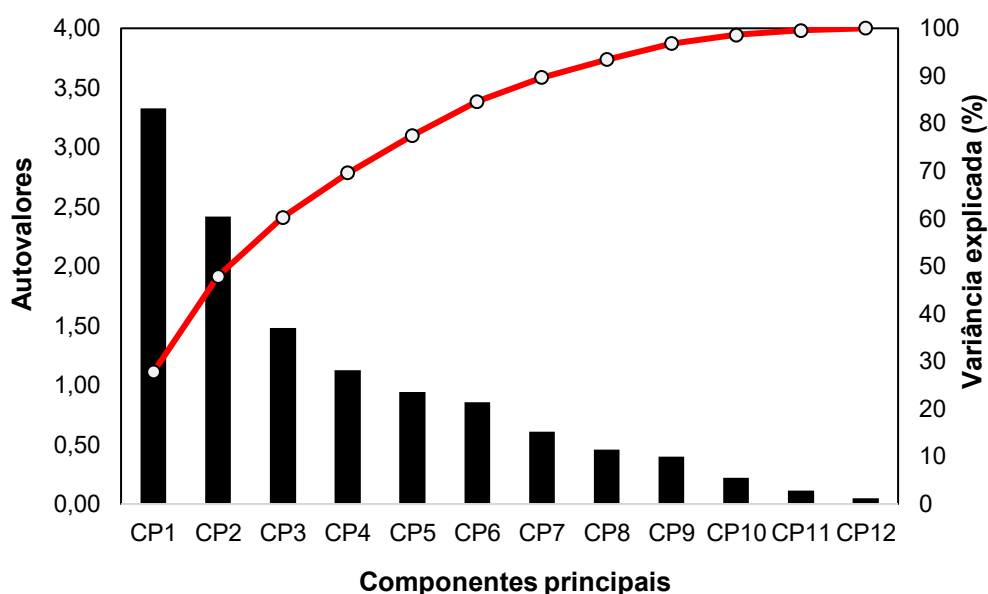
Tabela 5.1 - Modelo de ACP obtido a partir de observações não críticas

Parâmetro /Componente Principal (CP)	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12
A _{ST}	-0,355	0,010	0,101	-0,392	0,064	0,247	0,719	-0,136	0,017	-0,146	0,297	0,003
Cl ⁻	-0,476	0,011	0,298	0,161	-0,144	-0,128	-0,012	0,187	-0,086	0,054	-0,179	0,737
Colif. term.	0,212	0,288	0,135	-0,250	-0,305	0,612	-0,209	0,492	-0,063	-0,144	0,086	0,074
CE	-0,469	0,046	0,251	0,137	-0,269	-0,058	0,028	0,222	-0,232	0,023	-0,295	-0,655
DBO	-0,055	0,505	-0,296	-0,252	0,208	-0,038	0,164	0,176	0,216	0,533	-0,396	0,014
P _T	-0,383	-0,104	0,063	-0,118	-0,051	0,438	-0,450	-0,496	0,357	0,228	-0,028	-0,043
N-NO ₃ ⁻	0,193	-0,420	0,286	-0,392	0,018	-0,030	-0,031	0,029	-0,431	0,599	0,058	0,012
OD	0,032	-0,498	0,219	-0,027	0,266	0,023	0,071	0,472	0,610	-0,052	-0,148	-0,078
pH	-0,250	0,161	0,094	-0,024	0,796	0,065	-0,312	0,152	-0,297	-0,141	0,180	-0,054
S _T	0,066	0,401	0,508	0,004	-0,088	-0,372	-0,087	-0,003	0,334	0,194	0,513	-0,102
T _{ÁGUA}	0,181	0,061	0,208	0,686	0,171	0,455	0,307	-0,065	-0,047	0,335	0,031	-0,007
Turb.	0,308	0,184	0,537	-0,189	0,149	-0,003	0,060	-0,357	-0,020	-0,296	-0,553	0,025
Autovalores	3,33	2,42	1,48	1,12	0,94	0,86	0,61	0,46	0,40	0,22	0,12	0,05
Variância Explicada (%)	27,74	20,16	12,33	9,35	7,87	7,13	5,06	3,80	3,34	1,84	0,96	0,41
Variância Explicada Acumulada (%)	27,74	47,91	60,24	69,59	77,46	84,59	89,65	93,45	96,79	98,62	99,59	100,00

Fonte: Próprio autor.

A Figura 5.6 apresenta o gráfico do tipo *scree*, com os autovalores em ordem decrescente e as respectivas porcentagens da variabilidade acumulada. Observa-se que o modelo com 5 componentes principais explica 77,46% da variância total dos dados originais e o autovalor é próximo a um (0,94). Como a regra de Kaiser deve ser considerada como uma orientação e não como uma lei inviolável (LATTIN; CARROLL; GREEN, 2011), optou-se por um modelo com cinco CPs, mesmo CP5 apresentando um autovalor abaixo de 1,0.

Figura 5.6 - Gráfico *scree* com os autovalores e a variabilidade acumulada das componentes principais obtidas a partir do modelo de ACP



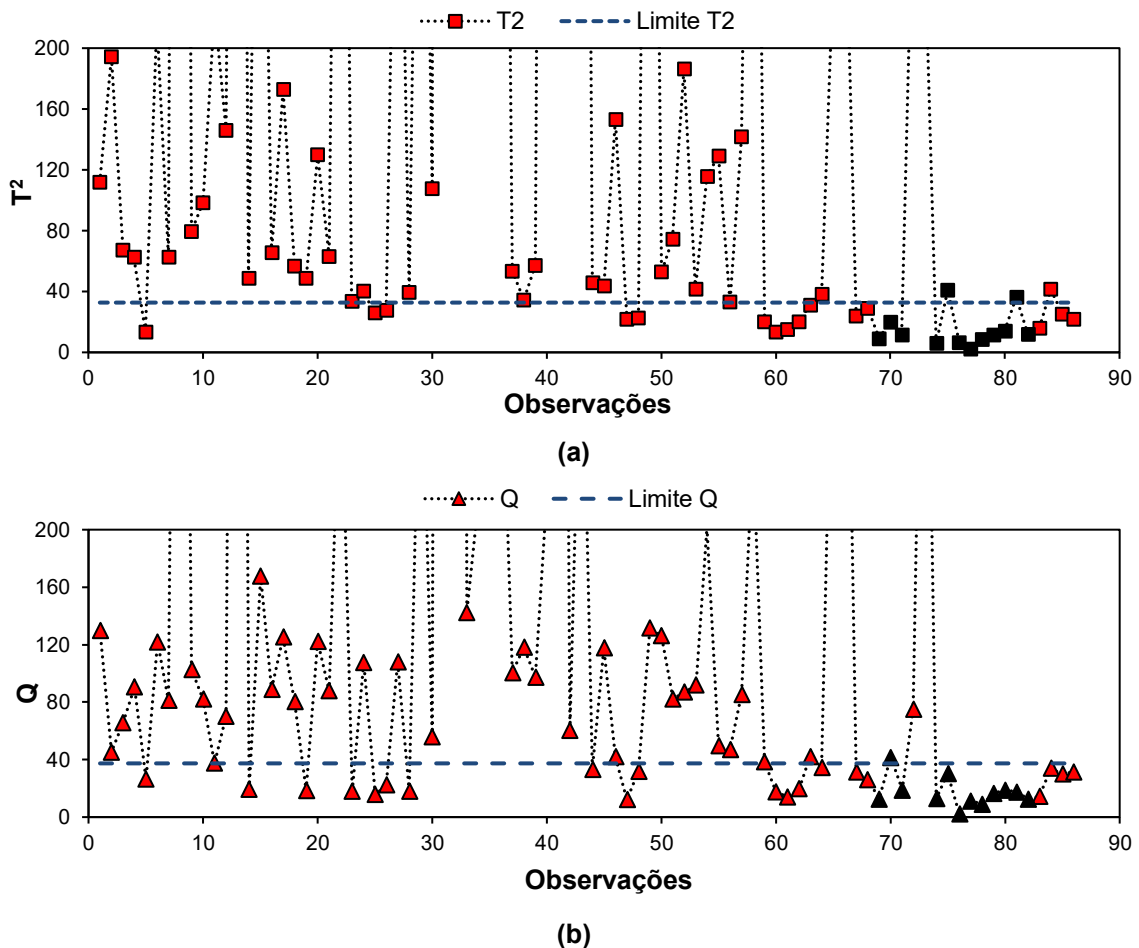
Fonte: Próprio autor.

A partir do modelo de ACP obtido e do segundo banco de dados de referência (subconjunto 2), os limites dos gráficos de controle T^2 e Q foram determinados. Para o cálculo dos limites de controle, utilizou-se um nível de confiança de 95%, que resultou em $T^2_{limite} = 32,73$ e $Q_{limite} = 37,39$. Acima desses valores, as amostras foram consideradas como condição crítica de qualidade da água.

Com o terceiro conjunto de dados (conjunto de teste), utilizaram-se todas as 74 observações classificadas como críticas e 12 observações de referência restantes (subconjunto 3), sendo constituído, então, por 86 observações (Figura 5.5). O objetivo foi verificar se as métricas calculadas seriam capazes de detectar as condições críticas de qualidade da água. A Figura 5.7 apresenta os gráficos de controle para as estatísticas T^2 e Q , calculadas a partir das observações do conjunto de dados de teste.

As duas estatísticas são importantes porque elas contêm informações complementares.

Figura 5.7 - Cartas de controle multivariadas para as estatísticas (a) T^2 e (b) Q , calculadas a partir do conjunto de dados de teste



Nota: As observações na cor vermelha referem-se às amostras críticas e as observações na cor preta são as amostras de referência.

Fonte: Próprio autor.

Na Figura 5.7 (a) e (b), as observações estão dispostas em ordem cronológica. As observações na cor vermelha referem-se às amostras críticas e as observações na cor preta são as amostras de referência. Teoricamente, para uma detecção de 100%, todas as amostras críticas deveriam ter ficado acima dos limites de controle, e as amostras de referência, abaixo. Os valores acima do limite indicam que há um desvio da qualidade da água em relação à condição de referência, que pode ser resultado de fontes de poluição pontuais ou não pontuais. Valores altos das estatísticas T^2 ou Q , indicam amostras longe do centro do modelo de ACP, ou seja, distantes da condição de referência.

Assim, a condição de qualidade de água crítica, em termos de poluição, foi detectada quando as observações críticas, nos gráficos de T^2 e/ou Q , ficaram acima do respectivo limite estabelecido. Na carta de controle de T^2 (Figura 5.7 (a)), a maioria das observações atípicas apresentaram valores superiores ao limite de controle. Essa carta foi capaz de identificar 79,73% dos dados críticos e gerou dois alarmes falsos, ou seja, a amostra era uma condição de referência, mas ficou acima do limite de controle.

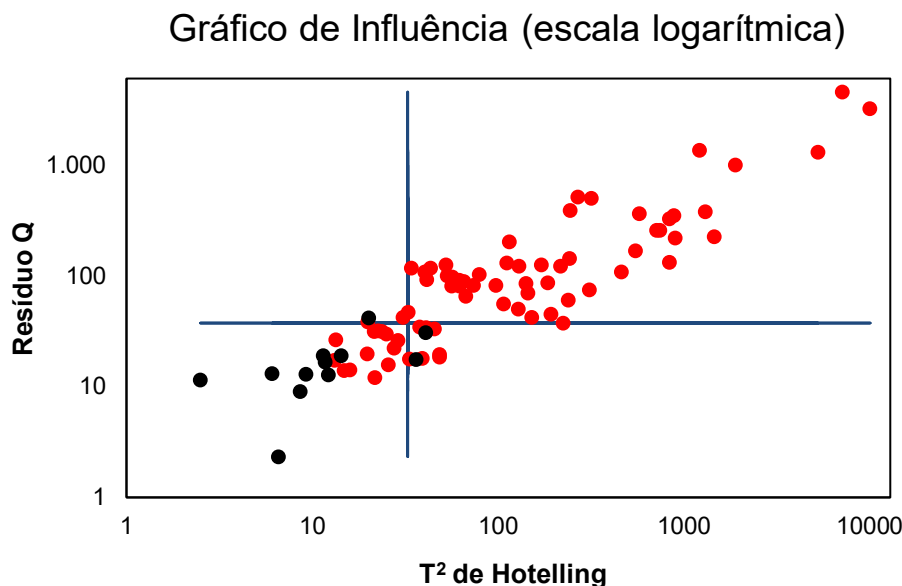
Os *outliers* são comuns em dados de qualidade de água e podem ser indicativos de condições extremas de poluição das águas. Resultados semelhantes foram encontrados por Sancho et al. (2016), que aplicaram cartas de controle univariadas e análise funcional de dados para a detecção de *outliers*, caracterizados pelos autores como possíveis episódios de poluição.

Na Figura 5.7 (b), as observações também estão dispostas em ordem cronológica, e a carta de controle Q detectou 72,97% das amostras classificadas como críticas e gerou um alarme falso. A carta de controle Q é utilizada para analisar as amostras que não são descritas pelo modelo de ACP.

Na Figura 5.8, a estatística Q foi plotada em relação à estatística T^2 , gráfico conhecido como gráfico de influência (BRO; SMILDE, 2014). Foi adotada a escala logarítmica para facilitar a visualização das variações registradas. As duas linhas azuis são os níveis de confiança de 95%. As observações na cor preta são as amostras de referência e aquelas na cor vermelha referem-se às amostras críticas. É possível perceber que 75% das amostras de referência permanecem dentro dos dois limites de confiança. As observações críticas apresentam um comportamento aproximadamente linear, com tendência crescente, onde 70,27% das amostras se destacam com valores de T^2 de Hotelling e Q elevados, acima dos dois limites de controle.

Em termos práticos, a partir da Figura 5.8, observa-se que, se limites de controle mais restritivos fossem adotados, um maior número de amostras críticas estaria acima dos limites de controle e, conseqüentemente, os percentuais de detecção das amostras críticas poderiam ser mais elevados.

Figura 5.8 - Gráfico de influência dos dados de qualidade da água calculado a partir de um modelo de cinco componentes principais



Fonte: Próprio autor.

Na Tabela 5.2, é apresentado um resumo do número de detecções, não detecções, alarmes falsos e aceites verdadeiros das cartas de controle, determinados pelos limites de controle. Os aceites verdadeiros referem-se às amostras não críticas que deveriam estar abaixo do limite de controle e efetivamente estão.

Analisando em conjunto a estatística Q e a estatística T^2 , foram detectadas 61 observações críticas dentre as 74 analisadas; destas, 52 observações foram detectadas nas duas cartas de controle. Ressalta-se que 13 observações críticas não foram detectadas por nenhuma das duas cartas de controle.

Esses resultados indicam que as métricas de monitoramento T^2 e Q são sensíveis às mudanças nos parâmetros de qualidade da água e, portanto, podem ser utilizadas para indicar episódios de poluição.

Tabela 5.2 - Resumo dos resultados apresentados pelas cartas de controle T^2 de Hotelling e de resíduos Q

Tipo de observação	T^2 de Hotelling	Resíduo Q	Número de observações
Observações críticas (total = 74)	Detecção	Não detecção	7
	Não detecção	Detecção	2
	Não detecção	Não detecção	13
	Detecção	Detecção	52
Observações não críticas (total = 12)	Aceites verdadeiros	Aceites verdadeiros	9
	Aceites verdadeiros	Alarme falso	1
	Alarme falso	Aceites verdadeiros	2
	Alarme falso	Alarme falso	0

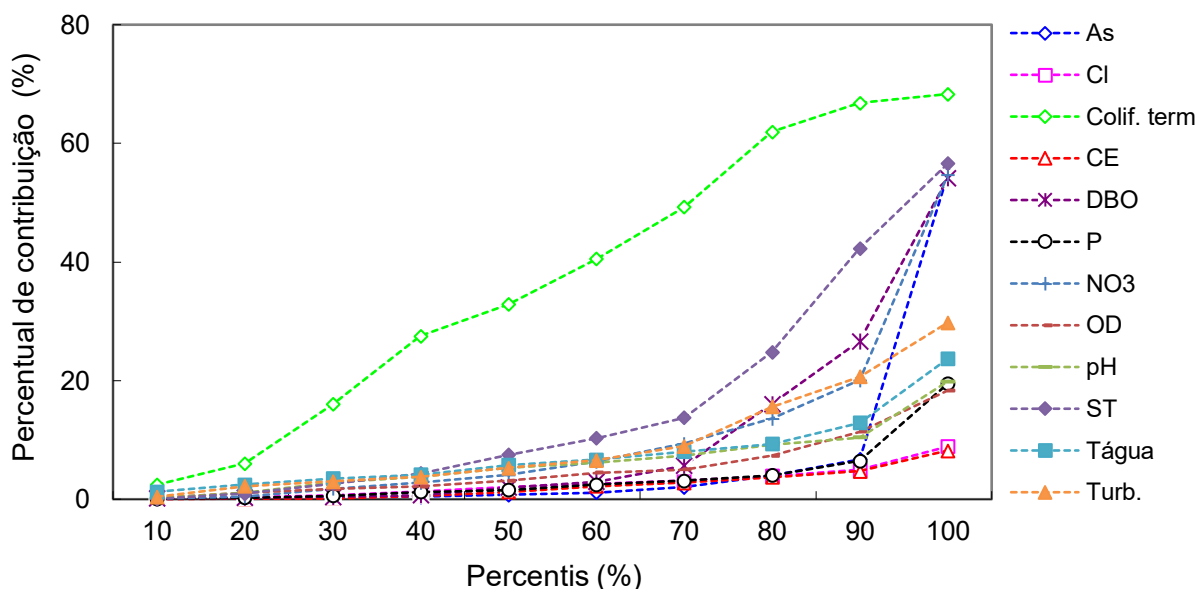
Fonte: Próprio autor.

As observações com valores extremos de Q foram analisadas com o suporte dos gráficos de contribuição (ferramenta de diagnóstico), para identificação dos parâmetros de qualidade de água com os maiores percentuais de contribuição. Dessa forma, é possível intervir de maneira prática no processo, subsidiando gestores e analistas com informações importantes para a tomada de decisão, fornecendo um direcionamento para a implementação de ações.

Os percentuais de contribuição dos parâmetros de qualidade de água na estatística Q , para cada amostra, foram cálculos a partir da Equação (2.15). Com base nos resultados, com o objetivo de identificar os parâmetros com os maiores percentuais de contribuição, considerando o conjunto de amostras críticas detectadas pela carta de controle Q , foram calculados os percentis apresentados na Figura 5.9.

Os resultados mostraram que os parâmetros com os maiores percentuais de contribuição para essas observações foram, em geral, coliformes termotolerantes, sólidos totais, turbidez, DBO e nitrato. Na Figura 5.9, observa-se que o parâmetro coliformes termotolerantes se destaca, apresentando em 30% das amostras um percentual de contribuição superior a 49%.

Figura 5.9 - Percentual de contribuição dos parâmetros de qualidade de água na estatística Q , considerando as amostras críticas detectadas



Fonte: Próprio autor.

O gráfico de contribuição identifica o grupo de parâmetros que fazem com que T^2 e Q apresentem valores extremos e não sejam consistentes com as condições de

referência. Nesse caso, a experiência dos analistas da área sobre o processo em questão deve ser utilizada para realizar inferências das causas que provocaram as alterações nos parâmetros identificados.

Os resultados encontrados são coerentes, visto que são compatíveis com as fontes de poluição existentes na região e com os resultados de outros estudos. A estação BV137 está localizada a jusante de municípios da Região Metropolitana de Belo Horizonte (RMBH), e recebe influência dos lançamentos de esgotos de grandes centros urbanos como, por exemplo, Sabará, Belo Horizonte, Contagem, Santa Luzia e Lagoa Santa, sendo um trecho crítico em termos de poluição (IGAM, 2010, 2018).

No estudo de Calazans et al. (2018b), os autores aplicaram análises estatísticas multivariadas (Análise de Cluster e Análise Fatorial), e identificaram que essa estação apresentou os dois primeiros fatores da AF correlacionados com parâmetros de qualidade da água característicos de atividades minerárias e matéria orgânica.

Pinto, Almeida e Oliveira (2018) também analisaram as estações de monitoramento da calha do rio das Velhas, e a estação BV137 estava dentre as estações classificadas como críticas. Trindade et al. (2017) encontraram resultados análogos na avaliação das estações de monitoramento da sub-bacia do rio das Velhas, que verificou que as estações localizadas no limite entre o Alto e Médio rio das Velhas apresentaram valores medianos do IQA entre 30 e 48.

As análises de dados estatísticos podem representar uma ferramenta confiável para os tomadores de decisão no planejamento de bacias hidrográficas, fornecendo uma visão geral das políticas de gestão dos cursos de água (DIAMANTINI et al., 2018). Após aplicação das cartas de controle das estatísticas T^2 e Q no estudo de caso da estação de monitoramento BV137, fica comprovado que essa metodologia oferece resultados satisfatórios para a identificação de eventos críticos de poluição, além de possibilitar a identificação dos parâmetros de qualidade de água que contribuem para a degradação da qualidade da água. Neste estudo, a metodologia foi aplicada para o monitoramento *off-line* (análise retrospectiva) de uma estação de monitoramento, podendo ser aplicada para o monitoramento de qualidade de água de outras estações em diferentes bacias hidrográficas. Além disso, ela pode ser utilizada para detecção *on-line* de mudanças abruptas na evolução da variável de interesse, principalmente em caso de frequências mais altas de coleta de dados.

As águas superficiais são sistemas complexos, com variações nos parâmetros de qualidade de água influenciadas por muitos fatores (climáticos, condições hidrogeológicas e interferências humanas), e com uma dinâmica temporal dos parâmetros inevitavelmente mascarada pelas frequências de amostragem tradicionais. Entretanto, os dados de séries temporais de alta frequência coletados pelos sensores de monitoramento automático podem refletir nessa dinâmica temporal complexa, e então fornecer uma nova visão sobre o funcionamento interno das bacias hidrográficas e cursos de água (RODE et al., 2016; SHI et al., 2018).

Neste século XXI, o monitoramento da qualidade da água ganhou mais interesse entre os pesquisadores (CHOWDURY et al., 2019). Nos últimos anos, com a ampliação da tecnologia de sensores, aumentou o número de locais nos quais são realizadas medições *in situ*, o que transformou as ciências das bacias hidrográficas (RODE et al., 2016). Parâmetros amplamente utilizados, como oxigênio dissolvido, turbidez, pH, condutividade específica, ortofosfato, matéria orgânica dissolvida, clorofila-*a*, nitrito e nitrato, que especialistas na área consideram relevantes, já são possíveis de serem medidos em tempo real e com alta resolução espacial ou temporal (RODE et al., 2016; SHI et al., 2018).

No estudo de Chowdury et al. (2019), por exemplo, os autores propõem um sistema de monitoramento da qualidade da água usando uma rede de sensores baseada em Internet das coisas (*Internet of Things – IoT*). A ideia é fornecer uma nova abordagem para aquisição, transmissão e processamento de dados em tempo real.

Rode et al. (2016) sintetizam os *insights* de tecnologias emergentes para medições de qualidade de água apresentando um conjunto de aplicações, e apontam avanços futuros, possibilitados por sensores, em nossa capacidade de entender, prever e restaurar sistemas de bacias hidrográficas e cursos de água. Segundo esses autores, o uso de sensores de alta frequência avançou do campo da pesquisa puramente acadêmica, e esses sensores são hoje empregados por vários órgãos ambientais nacionais, estaduais e municipais.

As estações de monitoramento de qualidade da água de alta frequência estão presentes em vários países. Nos Estados Unidos, por exemplo, a USGS (*United States Geological Survey*, uma agência de ciências) utiliza sensores ópticos em uma variedade de programas de pesquisa e monitoramento para entender melhor a qualidade da água *in situ* e em tempo real (PELLERIN, 2015). Na Espanha, foi

implantado pelo governo um Sistema Automatizado de Informação da Qualidade da Água (*Automated Water Quality Information System - AWQIS*), que consiste em aproximadamente 200 estações automáticas de alerta, espalhadas pelos oito distritos da bacia hidrográfica espanhola, com informações registradas a cada cinco minutos (DI BLASI et al., 2013). Na Alemanha, as autoridades hídricas estaduais usam cada vez mais o monitoramento de alta frequência, por exemplo, para quantificar fluxos de matéria (especialmente na saída de grandes rios) (RODE et al., 2016). No Brasil, em decorrência do desastre ambiental causado pelo rompimento de uma barragem de minério de ferro que atingiu um importante curso de água – o rio Doce – foram implantadas 22 estações automáticas no curso de água que fazem análises em tempo real, com registro de dados a cada 30 minutos. Em algumas dessas estações, são realizadas medições de turbidez, acidez (pH), oxigênio dissolvido, condutividade, temperatura da água e presença de microrganismos (FUNDAÇÃO RENOVA, 2020).

A geração de dados de monitoramento ambiental de ambientes aquáticos tem aumentado rapidamente, consequência de tecnologias aprimoradas para sensores de campo e crescentes pressões ambientais do desenvolvimento humano e das mudanças climáticas (CARLSON, 2016). Zimmer et al. (2019) destacam que o recente uso generalizado de sensores de alta frequência permitiram um aumento nos estudos, e conseqüentemente melhor conhecimento dos eventos hidrológicos e processos químicos em bacias hidrográficas.

Portanto, na era do “big data”, o gerenciamento ambiental inteligente e preciso é essencial (SHI et al., 2018). O modelo CEP-ACP desenvolvido neste estudo pode ser uma técnica eficaz para a detecção de anomalias na qualidade de água em locais específicos (abordagem local), com base em variáveis de qualidade da água com amostragem em alta frequência. O monitoramento *on-line* desses dados, a partir dos gráficos multivariados de controle estatístico, poderia ser a fonte para detectar tendências e gerar alertas sobre a qualidade da água em tempo real.

5.4 CONCLUSÃO

Este estudo apresenta uma nova abordagem para a detecção de condições críticas de degradação da qualidade da água, propiciando um processo de tomada de decisão mais racional e rápido em relação ao gerenciamento da qualidade das águas

superficiais. Um modelo ACP integrado ao Controle Estatístico de Processo Multivariado (CEPM-ACP) foi proposto e aplicado a um estudo de caso na bacia do rio das Velhas (Minas Gerais, Brasil). Neste estudo, a metodologia foi aplicada para o monitoramento *off-line* e os resultados das avaliações demonstraram que a ferramenta CEPM-ACP foi capaz de identificar as amostras críticas, indicando, efetivamente, os picos acentuados nas variações da qualidade da água.

Dessa forma, essa metodologia pode ser aplicada para o monitoramento de variáveis de qualidade da água em outros locais, mas sempre considerando as especificidades de cada um deles na interpretação dos resultados. Além disso, esse procedimento pode ser utilizado para monitoramento *on-line* nos casos em que o monitoramento é realizado com alta frequência, para detecção *on-line* de mudanças na evolução temporal dos parâmetros de qualidade de água. Assim, CEPM-ACP é uma ferramenta com grande potencial de uso para o monitoramento de qualidade de água superficial, podendo servir como sinal de alerta sobre desvios em relação à condição usual de qualidade de água, e então propiciar um processo de tomada de decisão mais racional em relação à redução e prevenção de fontes de poluição em geral.

**CAPÍTULO 6:
MONITORAMENTO E AVALIAÇÃO
ESTATÍSTICA DAS VARIAÇÕES
TEMPORAIS NA QUALIDADE DA
ÁGUA SUPERFICIAL: ESTUDO DE
CASO DO RIO DAS VELHAS**

RESUMO

Como o gerenciamento e a proteção eficazes dos recursos hídricos dependem da compreensão das variações temporais da qualidade das águas superficiais, o presente estudo focou, principalmente, em uma metodologia para o estudo da variabilidade temporal da qualidade das águas superficiais. Como estudo de caso, foi utilizado um conjunto de dados de qualidade de água de 21 anos (1997-2018), contendo 13 parâmetros físico-químicos, microbiológicos e hidrológicos, de um local de amostragem localizado no rio das Velhas, Brasil. Os períodos climáticos, seco e chuvoso, foram definidos a partir do estudo da precipitação em 15 estações pluviométricas. As análises foram realizadas com base na utilização do coeficiente de correlação de Spearman, teste U de Mann-Whitney e Análise por Componentes Principais (ACP). Os parâmetros associados à variação temporal identificados pela correlação de Spearman e o teste U de Mann-Whitney foram cloreto total, condutividade elétrica, fósforo total, sólidos totais, temperatura da água, turbidez e vazão. A ACP identificou três componentes principais (CP), as quais são responsáveis pela estrutura dos dados, explicando 62,47% da variância total do conjunto de dados. Cada componente foi associada a uma fonte de poluição: CP1 representou a contribuição da poluição difusa, CP2 foi associada à poluição orgânica, e CP3 à contaminação por atividades de mineração. A partir do estudo da contribuição/importância dos meses de monitoramento, baseada nas CPs, verificou-se que os meses de janeiro, julho e outubro foram os que mais contribuíram para a explicação da variabilidade dos dados e que nos meses de junho e setembro, os percentuais de contribuições para as três componentes foram próximos, indicando uma menor influência da sazonalidade. Os *scores* obtidos também foram analisados e observou-se um padrão espacial considerando os dois períodos climáticos: as amostras coletadas no período chuvoso apresentaram maior influência, principalmente de CP1 (poluição difusa), e as amostras do período seco, maior influência de CP2 (poluição pontual). Este estudo apresentou a utilidade de diferentes técnicas estatísticas para avaliação da variabilidade temporal da qualidade da água e a importância dos dados pluviométricos e de vazão nas análises.

Palavras-chave: Qualidade da água; Variabilidade temporal; Variáveis físico-química; ACP; Vazão.

6.1 INTRODUÇÃO

Uma condição necessária para alcançar os objetivos de desenvolvimento sustentável em saúde, segurança hídrica e alimentar é ter água de boa qualidade e em quantidade adequada. Entretanto, desde os anos 90, na maioria dos rios da América Latina, África e Ásia a poluição da água tem se intensificado. Nessas regiões, um terço de todos os trechos de rios está poluído por patógenos e um sétimo por material orgânico (UNEP, 2016).

Diversos fatores influenciam a qualidade da água superficial em uma região: fatores naturais (geologia, precipitação, clima, topografia, lixiviação natural de matéria orgânica e nutrientes do solo, etc) e fatores antropogênicos (urbanização, atividades industriais e agrícolas, mudanças no uso e cobertura do solo, etc) (BARTRAM et al., 1996; KHATRI; TYAGI, 2015). Os fatores antropogênicos, como as descargas de efluentes domésticos e industriais, podem ser considerados uma fonte poluidora constante, enquanto fatores naturais, por exemplo, o escoamento superficial, são fenômenos sazonais amplamente afetados pelo clima (VEGA et al., 1998). Como consequência, os dados ambientais são geralmente caracterizados por alta variabilidade (SIMEONOV et al., 2002).

Em relação às variações sazonais, a precipitação e o escoamento superficial possuem um forte efeito sobre as taxas de fluxo nos rios, o que pode afetar a concentração de poluentes nos corpos d'água (VEGA et al., 1998). De acordo com Chapman et al. (1992), as variações temporais dos valores dos parâmetros podem ser determinadas com uma baixa frequência de amostragem por um longo período de tempo, ou com uma alta frequência por um período mínimo de um ano (que corresponde a um ciclo hidrológico), e essa variação temporal pode ser descrita a partir do estudo das concentrações ou cargas medidas. Dentre os principais tipos de variabilidade temporal, pode-se citar: variabilidade de dias a meses, sazonalidade dos ciclos hidrológicos e biológicos e tendências ano a ano (CHAPMAN et al., 1992).

Como a variação na qualidade da água ao longo do tempo e do espaço é complexa, para uma gestão eficaz da qualidade da água de um rio, dois tipos principais de informações são necessárias: características espaciais e temporais dos poluentes e informações sobre os fatores determinantes que influenciam a qualidade

da água (LIU et al., 2016). Essas informações são obtidas por meio dos programas de monitoramento da qualidade da água (PMQA).

Assim, tendo em vista que é necessária a análise de vários parâmetros e fatores que afetam a qualidade da água, simultaneamente, para facilitar a interpretação dos dados obtidos dos PMQA, diferentes abordagens estatísticas multivariadas têm sido aplicadas (GHOLIZADEH; MELESSE; REDDI, 2016; JABBAR; GROTE, 2019; SHRESTHA; KAZAMA, 2007; SINGH et al., 2004; WUNDERLIN et al., 2001; ZEINALZADEH; REZAEI, 2017). A Análise por Componentes Principais, uma dessas técnicas multivariadas, é a mais comumente utilizada em estudos de bacias hidrográficas (OLSEN; CHAPPELL; LOFTIS, 2012).

Nesse contexto, considerando a importância do estudo das variações temporais para investigar e avaliar a qualidade das águas superficiais, o objetivo deste estudo foi identificar os parâmetros responsáveis pelas variações temporais da qualidade da água no trecho de um curso de água, considerando os períodos climáticos seco e chuvoso de monitoramento, e analisar as fontes de poluição. Na metodologia proposta, a avaliação completa do banco de dados de qualidade da água foi realizada utilizando a correlação de Spearman, o teste U de Mann-Whitney e a Análise por Componentes Principais, além do estudo de dados pluviométricos e de vazão.

6.2 METODOLOGIA

6.2.1 Área de estudo

A área de estudo está detalhadamente descrita no Capítulo 3, item 3.1

6.2.2 Banco de dados

6.2.2.1 Dados de monitoramento de qualidade da água

As informações sobre o banco de dados da estação de monitoramento de qualidade da água localizada na área de estudo, as análises realizadas para a definição do banco de dados disponibilizado pelo Igam, e o pré-processamento dos dados estão descritos no Capítulo 3, nos itens 3.2 e 3.3.

6.2.2.2 Dados de monitoramento pluviométrico e fluviométrico

As informações sobre as estações de monitoramento pluviométrico e fluviométrico, localizadas na região da área de estudo, estão descritas no Capítulo 3, item 3.1.5.

6.2.3 Dados de vazão obtidos a partir do monitoramento fluviométrico

A primeira etapa para realizar o estudo das variações sazonais na qualidade da água na estação BV137 consistiu na obtenção dos dados de vazão.

Como na estação de monitoramento BV137 não há medição simultânea de vazão, foi necessário realizar um procedimento de transferência espacial de informação de uma estação de monitoramento fluviométrico para a estação BV137. Visto que o local de interesse (estação de qualidade de água BV137) encontra-se localizado a 4,8 km da estação fluviométrica 41340000, foi possível realizar o procedimento de transferência de informações por meio do estabelecimento de uma relação linear entre vazão e a razão entre as áreas de drenagem do ponto de coleta de qualidade da água e da estação fluviométrica (Equação 6.1) (COLLISCHONN; TASSI, 2008).

$$Q_{QA} = Q_{Flu} \times \frac{A_{QA}}{A_{Flu}} \quad (6.1)$$

Onde,

Q_{QA} = vazão na estação de monitoramento de qualidade da água

Q_{Flu} = vazão na estação fluviométrica

A_{QA} = área de drenagem da estação de monitoramento de qualidade da água

A_{Flu} = área de drenagem da estação fluviométrica

Esse procedimento também foi aplicado em outros estudos (OLIVEIRA; MAILLARD; PINTO, 2016, 2017; PINTO et al., 2017).

Para realizar a transferência da vazão proporcionalmente à respectiva área de drenagem, três recomendações são sugeridas no estudo conjunto da Fundação Rural Mineira (RURALMINAS), Universidade Federal de Viçosa (UFV) e Igam (2007): a diferença (valor absoluto) entre as áreas das bacias de drenagem do local de interesse e a da estação fluviométrica não deve superar 30% da área de influência da estação, as estações devem, preferencialmente, estar inseridas na mesma calha fluvial, e em uma região de comportamento hidrológico semelhante. Essas recomendações foram adotadas nesta pesquisa e verificou-se que a relação entre as áreas das bacias de drenagem da estação fluviométrica (ADFLU) 41340000 e da estação de qualidade da água (ADQA) BV137 é inferior a 30% da área de drenagem da estação fluviométrica, e as duas estações estão localizadas no mesmo curso de água. Além disso, as duas estações estão localizadas em uma região hidrológicamente homogênea (RURALMINAS; UFV; IGAM, 2021). Dessa forma, todas as recomendações foram satisfeitas, tornando pertinente a aplicação do procedimento descrito anteriormente.

Dado que na estação fluviométrica a medição da vazão é diária, os dados utilizados para a transferência foram selecionados de acordo com as datas das coletas realizadas na estação de qualidade da água. Assim, a estimativa da vazão na estação BV137 foi realizada utilizando a Equação 6.1.

Os dados de vazão foram, então, adicionados ao banco de dados de qualidade de água da estação BV137, formando um conjunto de observações que contém 12 (doze) parâmetros físico-químicos e microbiológicos de qualidade da água (Tabela 3.3) e 1 (um) parâmetro hidrológico – a vazão, monitorados de 1997 a 2018 (monitoramento inicialmente trimestral e a partir de 2008 mensal), formando uma

matriz de dados $X_{[n=146, p=13]}$, onde n é o número de observações por variável e p é o número de variáveis.

6.2.4 Avaliação dos dados de monitoramento pluviométrico

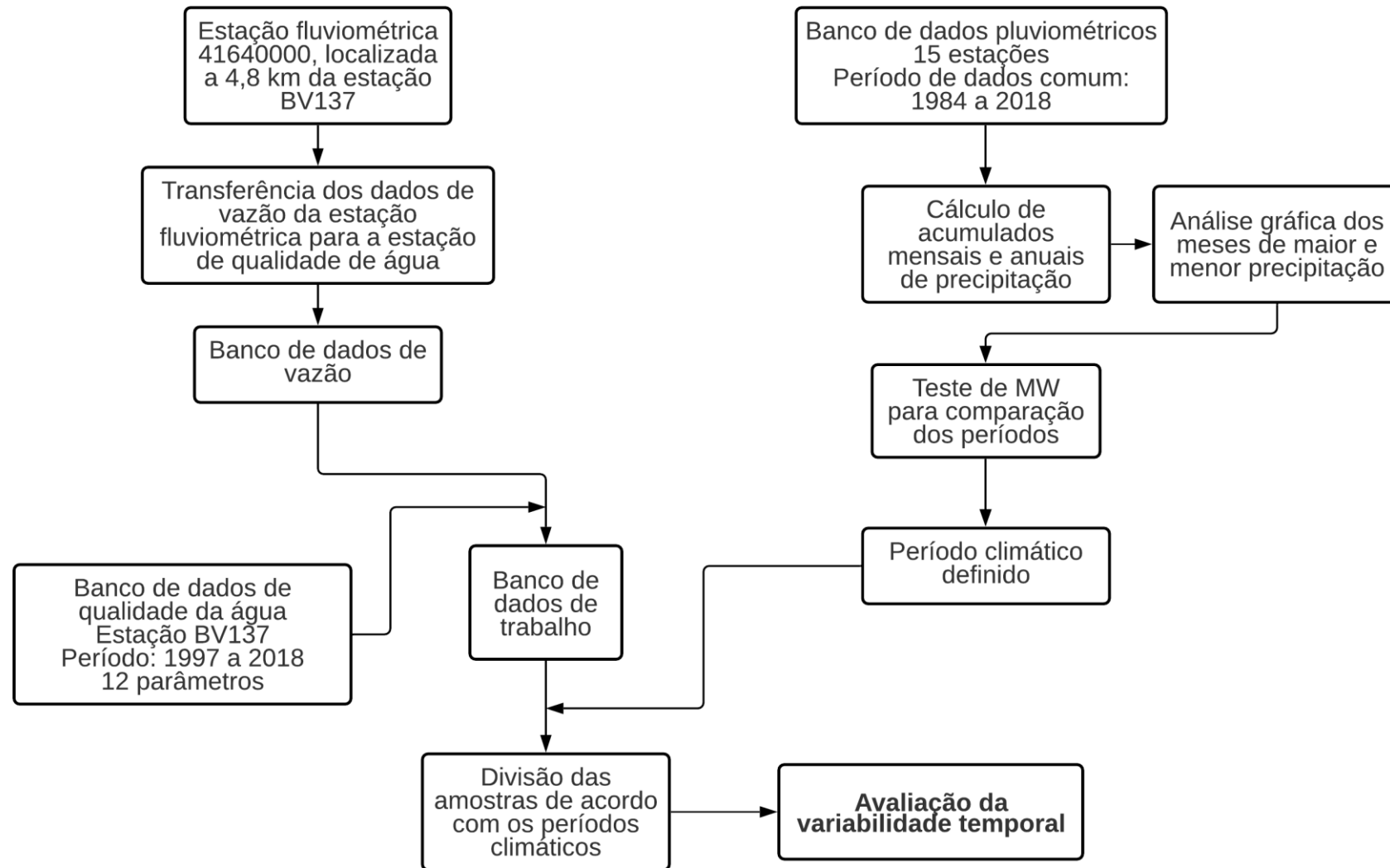
A segunda etapa para realizar a avaliação da variabilidade temporal da qualidade da água foi a definição dos períodos seco e chuvoso, realizada a partir dos dados pluviométricos.

Na região da área de drenagem da estação BV137, existem 15 (quinze) estações pluviométricas (Figura 3.8), que foram utilizadas para caracterizar o regime de precipitação na região. Para uma caracterização adequada, utilizou-se a série de dados disponível, considerando um período comum a todas as estações. Após avaliação do banco de dados obtido no sistema Hidroweb do SNIRH, foram utilizados os dados do período de 1984 a 2018 (1º semestre). Os acumulados mensais e anuais de precipitação e os respectivos valores médios foram calculados e analisados.

A partir dos resultados obtidos das principais características sazonais usuais da precipitação na região, foram identificados os meses do ano de menor e maior precipitação (meses secos e chuvosos, respectivamente). Para confirmar os resultados, também foi aplicado um teste não paramétrico para comparação entre dois grupos - o teste U de Mann-Whitney, ao nível de significância (α) de 5%. Assim, verificou-se a existência de diferenças significativas entre os períodos definidos, considerando os acumulados mensais de precipitação.

As amostras de qualidade de água foram, então, agrupadas considerando esses dois períodos climáticos e as análises descritas abaixo foram realizadas considerando cada um deles separadamente. O fluxograma da metodologia utilizada para a avaliação das variações temporais da qualidade da água é apresentado na Figura 6.1.

Figura 6.1 - Fluxograma da metodologia adotada para a avaliação das variações temporais da qualidade da água superficial



Fonte: Próprio autor.

6.2.5 Avaliação das variações temporais da qualidade da água

6.2.5.1 Análises estatísticas preliminares do banco de dados

A princípio, as amostras do banco de dados foram divididas em dois grupos, período seco e período chuvoso, baseada nos resultados da análise pluviométrica, conforme descrito anteriormente.

A estatística descritiva foi calculada para todos os parâmetros de qualidade de água e de vazão analisados, considerando os dois períodos climáticos, e consistiu em apresentar os valores mínimo e máximo, a média, o desvio-padrão, os percentis (25, 50 - mediana e 75) e o coeficiente de assimetria.

Para identificar quais parâmetros apresentaram diferenças significativas entre os períodos, foi utilizado o teste U de Mann-Whitney, ao nível de significância (α) de 5%. Os resultados dos parâmetros com diferença significativa entre os períodos foram apresentados em gráficos *box-plot*.

Em seguida, foi utilizado o coeficiente de correlação de Spearman (ρ), que é uma medida não paramétrica de correlação, no qual os dados são convertidos em *ranks* antes de se calcular o coeficiente de correlação (REIMANN et al., 2008). Para realizar a avaliação do coeficiente de correlação de Spearman, cada período climático foi transformado em um valor numérico no arquivo de dados (chuvoso = 1 e seco = 2), e essa variável numérica foi então correlacionada (par a par) com todos os parâmetros medidos (KRISHNARAJ; DEKA, 2020; SINGH et al., 2004; WUNDERLIN et al., 2001).

A partir dessas análises, os parâmetros que mais se destacaram foram avaliados separadamente.

6.2.5.2 Análise por Componentes Principais

A ACP foi aplicada ao conjunto de dados completo e a análise sazonal foi realizada com base nas componentes principais. Inicialmente, o banco de dados foi padronizado em escala-Z com média 0 e desvio-padrão 1. A matriz de covariâncias foi calculada e a partir dela foram obtidos o vetor de autovalores ($\hat{\lambda}_{[p \times 1]}$) e a matriz de autovetores associados ($W_{[p \times p]}$), que descrevem a estrutura das variáveis originais.

Na sequência, foi utilizado o critério de Kaiser (1958), com exemplos de aplicação em Kowalkowski et al. (2006) e Simeonov et al. (2003), para a definição do

número k de componentes principais (CPs) a se reter (componentes com autovalores menores do que 1 (um) foram excluídos). Essa abordagem de seleção de CP é comumente usada na área de análise de qualidade de água. Baseado nesse critério, foram selecionadas k CPs, obtendo-se um modelo com o vetor reduzido de autovalores ($\lambda_{[k \times 1]}$) e a matriz reduzida de autovetores $W_{[p \times k]}$.

A partir das cargas (*loadings*), foram identificados os principais parâmetros responsáveis por explicar a maior variabilidade da qualidade da água e as diferentes fontes de poluição atuantes em cada CP retido. Para a seleção das cargas significativas, foi adotada a definição de Liu; Lin e Kuo (2003), que consideraram que cargas maiores que 0,75 são fortes, entre 0,75-0,50 são moderadas, e entre 0,50-0,30 são fracas. Em outras palavras, quanto maiores os valores de carga, mais importante é a variável para explicar a componente. Dessa forma, nessa pesquisa, as cargas fortes e moderadas foram utilizadas para indicar o tipo de poluição que mais afeta a qualidade da água monitorada em cada componente principal.

Após a identificação das componentes principais e sua associação a fontes de poluição, os valores dos *scores* foram analisados. A importância de uma observação para uma componente pode ser obtida pela razão entre o quadrado do *score*, de uma determinada observação, e a soma dos quadrados dos *scores* de todas as observações na componente. Para uma dada componente, a soma das contribuições de todas as observações é igual a 1 (um) (ABDI; WILLIAMS, 2010).

Portanto, quanto maior o valor da contribuição, mais a observação contribui para explicar o componente. A contribuição por mês de monitoramento foi calculada como a soma das contribuições de todas as observações naquele mês. Assim, foi possível analisar a contribuição/importância dos meses de monitoramento na explicação da variabilidade temporal da qualidade da água.

Ademais, foi apresentado o gráfico de *scores* para as duas primeiras componentes principais, para verificar a dispersão das amostras coletadas nos dois períodos climáticos.

Todas as análises estatísticas foram realizadas utilizando a linguagem de programação R (<https://rstudio.com/>), e foi utilizado também o Microsoft Excel (2019).

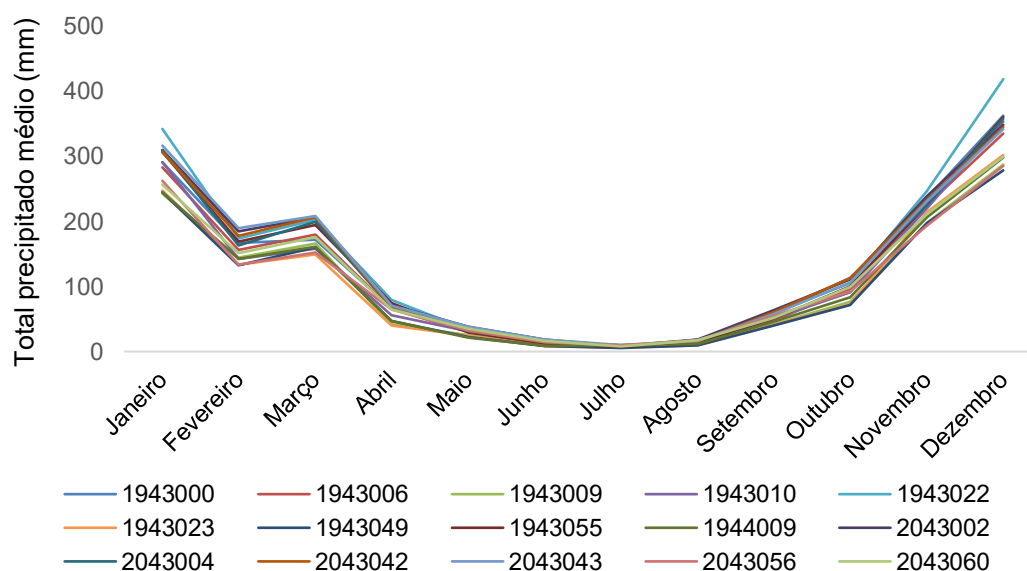
6.3 RESULTADOS E DISCUSSÃO

6.3.1 Avaliação dos dados de monitoramento pluviométrico

A partir dos dados da Rede Hidrometeorológica Nacional, 15 estações pluviométricas foram utilizadas neste estudo. A partir do mapa da Figura 3.8, verifica-se que as estações estão distribuídas de forma abrangente em toda a bacia de drenagem da estação BV137.

Considerando todas as estações pluviométricas, foi possível analisar a precipitação média mensal e anual na região, e então definir os períodos com maiores e menores acumulados de chuva, caracterizando, respectivamente, os períodos de chuva e estiagem. A Figura 6.2 apresenta a precipitação média mensal para as estações pluviométricas da ANA localizadas na área de estudo. Observa-se que o regime pluviométrico entre as estações é muito semelhante e que o período chuvoso se inicia no mês de outubro e segue até o mês de março, com poucas chuvas ocorrendo nos meses de abril e setembro. Considerando todas as estações, a precipitação média anual variou entre 1.217 mm e 1.698 mm, sendo que 85% a 90% do total anual de chuvas ocorre nos meses de outubro a março.

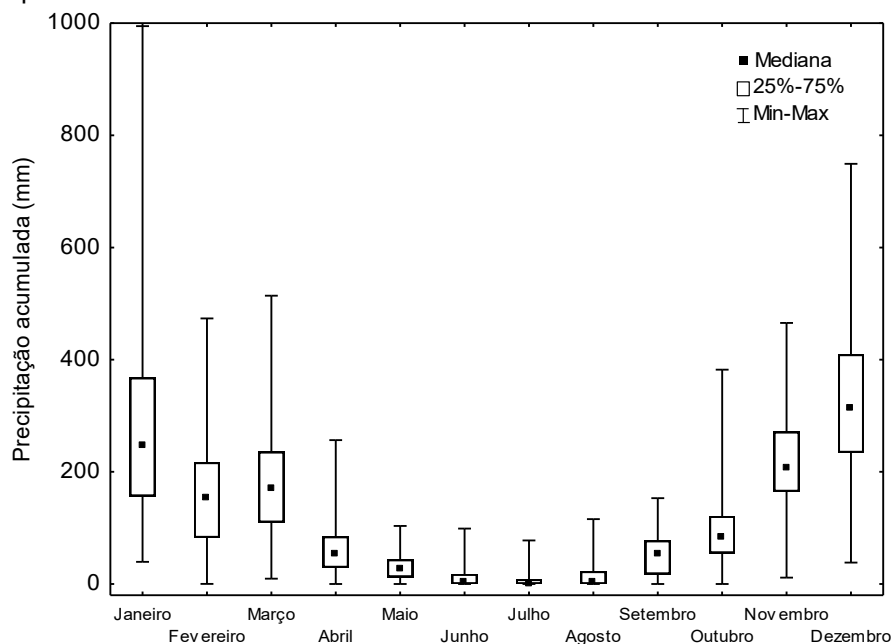
Figura 6.2 - Precipitação média mensal (em mm), na bacia de drenagem da estação BV137, de 1984 a 2018 (1º semestre)



Fonte: Próprio autor.

A Figura 6.3 apresenta o *box-plot* da precipitação acumulada mensal para todas as estações pluviométricas.

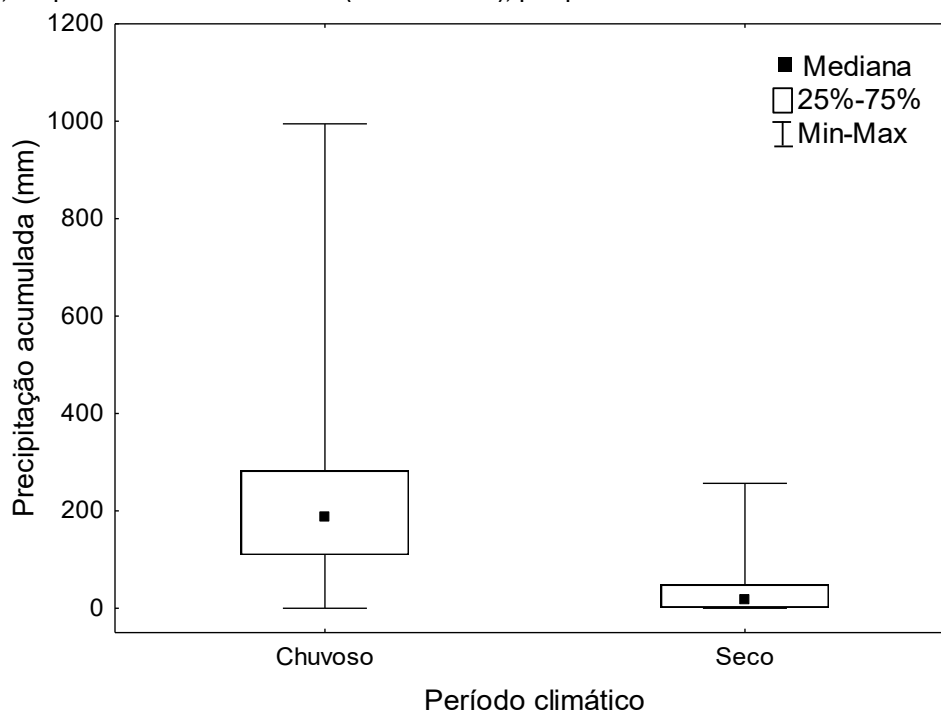
Figura 6.3 - *Box-plot* da precipitação acumulada, considerando o conjunto de estações pluviométricas analisadas, no período de 1984 a 2018



Fonte: Próprio autor.

O teste U de Mann-Whitney apresentou diferenças significativas (valor- $p < 0,05$) para os acumulados de chuvas no período analisado (Figura 6.4).

Figura 6.4 - *Box-plot* da precipitação acumulada, considerando o conjunto de estações pluviométricas analisadas, no período de 1984 a 2018 (1° semestre), por período climático



Fonte: Próprio autor.

Assim, considerou-se o período de chuva de outubro a março e o período seco de abril a setembro, que corrobora os períodos historicamente reportados na literatura. O estudo de Ferreira et al. (2020) apresenta resultados semelhantes aos obtidos.

6.3.2 Avaliação das variações temporais da qualidade da água

6.3.2.1 Análises estatísticas preliminares do banco de dados

As variações sazonais nos ambientes aquáticos geralmente estão associadas à ocorrência de chuvas, dessa forma, as amostras de qualidade de água dos períodos chuvoso e seco foram analisadas separadamente.

A estatística descritiva dos parâmetros estudados, na estação BV137, no período chuvoso (outubro a março) e no período seco (abril a setembro), é apresentada na Tabela 6.1. Os resultados mostraram que existem parâmetros que apresentam mudanças evidentes entre os períodos, indicando a variabilidade temporal na composição química da água. Comparando as medianas dos parâmetros de qualidade da água em cada período, é possível identificar que turbidez e vazão apresentam valores mais elevados no período chuvoso, enquanto as concentrações de cloreto total e fósforo total são mais elevadas no período seco.

Tabela 6.1 - Estatística descritiva dos dados amostrados na estação BV137 nos períodos chuvoso e seco

	Parâmetros	As _T (mg/L)	Cl ⁻ (mg/L)	Colif. term. (NMP ^a /100 mL)	CE (µS/cm)	DBO (mg/L)	P _T (mg/L)	N-NO ₃ ⁻ (mg/L)	OD (mg/L)	pH	ST (mg/L)	T _{água} (°C)	Turb. (UNT)	Vazão (m ³ /s)
Período chuvoso	N° de observações	74	74	74	74	74	74	74	74	74	74	74	74	74
	Mínimo	0,001	3,27	130	96	2,00	0,010	0,020	0,50	6,00	148	22,20	5	14,92
	Máximo	0,138	46,10	160.000	524	43,00	0,960	6,660	8,10	8,00	3.622	29,00	3.100	318,33
	Percentil 25	0,018	7,08	2.954	150	5,30	0,200	0,265	2,73	6,90	223	25,43	34	38,01
	Mediana	0,030	10,30	15.164	206	6,95	0,275	0,755	3,60	7,20	264	26,70	103	59,76
	Percentil 75	0,043	17,48	50.000	275	9,00	0,418	1,368	4,38	7,30	661	27,58	291	113,69
	Média	0,035	13,14	45.781	228	7,63	0,321	1,008	3,59	7,11	545	26,43	289	83,09
	Desvio-padrão	0,025	8,38	59.654	98	5,10	0,208	1,060	1,51	0,40	595	1,57	504	65,35
	CV ^b	70,9%	63,3%	129,4%	42,7%	66,4%	64,4%	104,5%	41,6%	5,5%	108,3%	5,9%	173,0%	78,1%
	Assimetria	1,79	1,47	1,20	1,17	4,63	1,06	2,60	0,37	-0,41	2,85	-0,50	3,37	1,50
Período seco	N° de observações	72	72	72	72	72	72	72	72	72	72	72	72	72
	Mínimo	0,006	3,90	50	129	2,000	0,030	0,010	0,50	6,20	130	18,00	4	12,86
	Máximo	0,077	40,50	160.000	512	131,00	1,610	4,000	5,70	7,80	1.157	31,70	486	117,95
	Percentil 25	0,022	12,33	4.089	212	4,58	0,280	0,188	2,50	6,90	184	21,50	12	28,97
	Mediana	0,028	17,05	11.000	282	6,15	0,490	0,540	3,35	7,20	201	22,80	23	36,56
	Percentil 75	0,035	21,70	35.000	334	10,20	0,690	0,995	3,90	7,40	235	24,30	46	49,95
	Média	0,031	17,48	37.385	282	10,99	0,505	0,775	3,16	7,13	240	23,13	56	42,06
	Desvio-padrão	0,015	7,27	57.154	89	16,57	0,289	0,816	1,20	0,33	143	2,25	94	20,03
	CV ^a	48,1%	41,3%	151,8%	31,3%	149,6%	56,9%	104,5%	37,7%	4,6%	59,3%	9,7%	168,1%	47,3%
	Assimetria	1,19	0,65	1,58	0,41	5,58	0,89	1,86	-0,39	-0,48	4,36	0,84	3,20	1,56

^a NMP – Número mais provável

^b CV – Coeficiente de variação

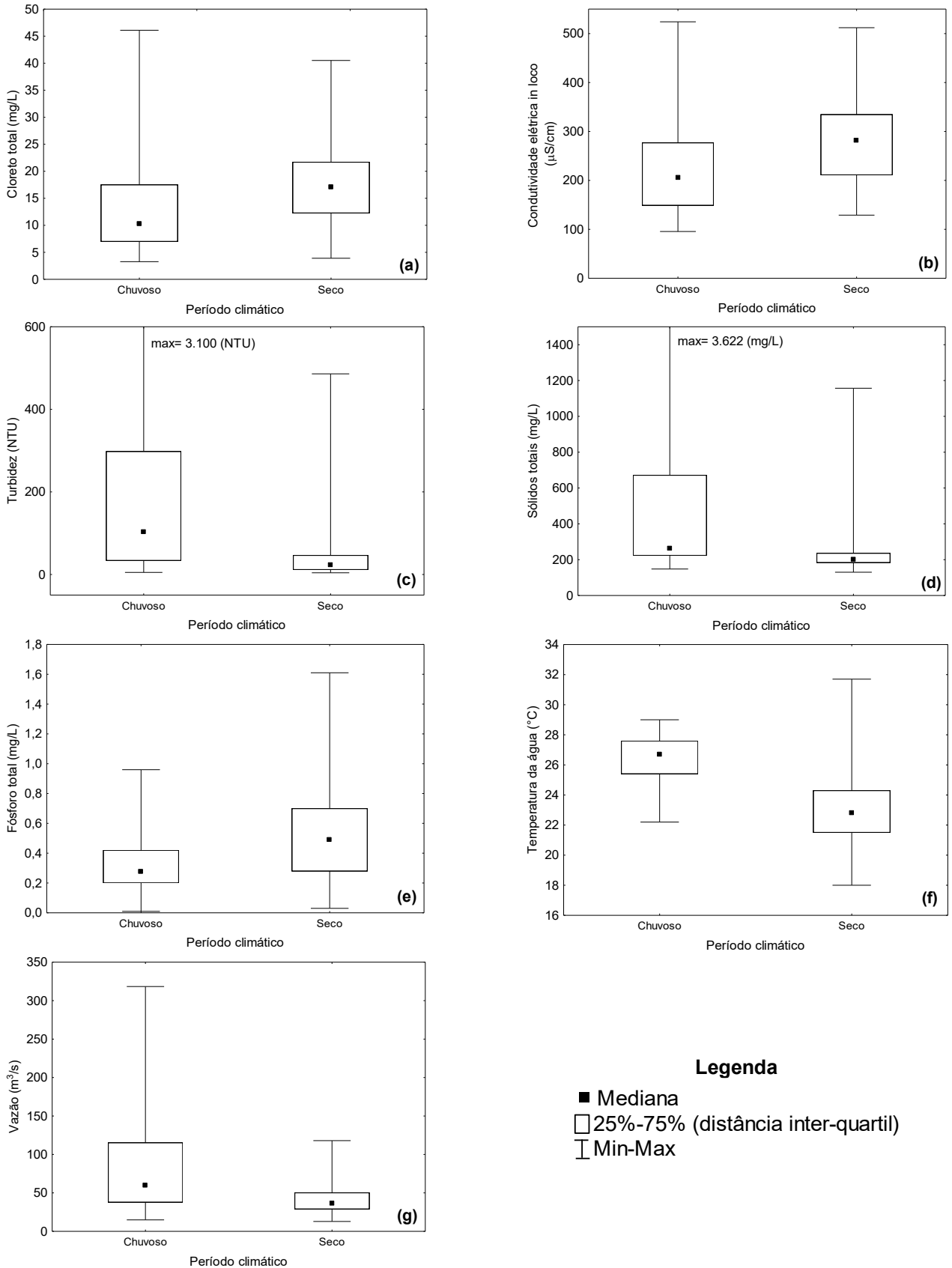
Os resultados do teste estatístico de Mann-Whitney, ao nível de significância de 5%, indicaram diferença significativa entre os valores medidos nos períodos seco e chuvoso para sete parâmetros, a saber: cloreto total, CE, fósforo total, sólidos totais, temperatura da água, turbidez e vazão. Os valores foram significativamente maiores no período chuvoso para a temperatura da água, sólidos totais, turbidez e vazão; e para os parâmetros cloreto total, fósforo total e CE as concentrações apresentaram valores mais elevados no período seco. Os demais parâmetros não apresentaram diferença significativa entre as estações seca e chuvosa.

A variabilidade temporal dos dados também foi avaliada utilizando o coeficiente de correlação de Spearman (ρ). O objetivo foi examinar a existência de uma correlação entre as variações nos valores dos parâmetros e o período climático. Os valores de correlação entre o período climático e cada parâmetro separadamente foram analisados e os resultados mostraram que houve correlações significativas (valor- $p < 0,05$) com o período climático para os parâmetros cloreto total ($\rho = 0,327$), CE ($\rho = 0,329$), fósforo total ($\rho = 0,340$), sólidos totais ($\rho = -0,483$), temperatura da água ($\rho = -0,673$), turbidez ($\rho = -0,415$) e vazão ($\rho = -0,366$). Os parâmetros relacionados ao período climático podem ser considerados a principal fonte de variações temporais na qualidade da água (MONICA; CHOI, 2016; SHRESTHA; KAZAMA, 2007; SINGH; MALIK; SINHA, 2005). Os valores de correlação de Spearman significativos confirmaram os resultados do teste de Mann-Whitney.

A seguir, são apresentados os gráficos *box-plot* na Figura 6.5 e um detalhamento dos resultados para os parâmetros que apresentaram diferença significativa entre os períodos seco e chuvoso.

No ponto de monitoramento em estudo, as concentrações de cloreto variaram entre 3,27 e 46,10 mg/L no período chuvoso, e entre 3,90 e 40,50 mg/L no período seco, com valor mediano mais elevado no período seco (Figura 6.5 (a)). A presença de cloreto nas águas superficiais pode estar relacionada às descargas de esgotos sanitários e aos efluentes industriais que apresentam concentrações de cloreto elevadas como os da indústria do petróleo, algumas indústrias farmacêuticas, curtumes etc (CETESB, 2020).

Figura 6.5 - *Box-plot* para os parâmetros (a) cloreto total, (b) condutividade elétrica, (c) turbidez, (d) sólidos totais, (e) fósforo total, (f) temperatura da água e (g) vazão, considerando os períodos chuvoso e seco



Fonte: Próprio autor.

O parâmetro condutividade elétrica é a medida da habilidade de uma solução aquosa em conduzir corrente elétrica e é influenciada pela presença de diferentes íons, sendo um deles o ânion cloreto (APHA; AWWA; WEF, 2017). Dessa forma, como a condutividade indica a quantidade de sais existentes na coluna d'água, ela representa uma medida indireta da concentração de poluentes. Na legislação estadual DN COPAM/CERH-MG n° 01/2008, não há limite estabelecido para esse parâmetro, mas, geralmente, níveis superiores a 100 $\mu\text{S}/\text{cm}$ indicam ambientes impactados (CETESB, 2020). O gráfico *box-plot* dos valores de CE nos dois períodos climáticos é apresentado na Figura 6.5(b). No período chuvoso, os valores variaram de 96 a 524 $\mu\text{S}/\text{cm}$, e no período seco os resultados ficaram entre 129 e 512 $\mu\text{S}/\text{cm}$ (Tabela 6.1). O valor mediano significativamente mais baixo observado na estação chuvosa (206 $\mu\text{S}/\text{cm}$), comparado à estação seca (282 $\mu\text{S}/\text{cm}$), pode estar relacionado à diluição (TIBEBE et al., 2019).

No período analisado, os valores de turbidez variaram de 5 a 3.100 UNT na estação chuvosa, e de 4 a 486 UNT durante a estação seca (Tabela 6.1), sendo os valores significativamente superiores no período chuvoso (Figura 6.5 (c)). Observa-se que o valor mediano de turbidez no período chuvoso é de 103 UNT (Tabela 6.1), resultado que se encontra acima do máximo permitido na legislação estadual DN COPAM/CERH-MG n° 01/2008 que é de 100 UNT, considerando classe 3 de enquadramento.

Os valores de sólidos totais também são significativamente superiores na estação chuvosa (Figura 6.5 (d)), variando entre 148 e 3.622 mg/L. Na estação seca, a concentração de sólidos totais esteve entre 130 e 1.157 mg/L (Tabela 6.1). No período chuvoso, há a ocorrência dos maiores índices pluviométricos, e conseqüentemente, valores mais elevados de vazão (Tabela 6.1). Nesse período, ocorre uma piora dos níveis de qualidade do curso de água em relação à turbidez e sólidos totais, resultante da contribuição do escoamento superficial para o aporte de sólidos (poluição difusa).

No mapa de uso e ocupação do solo da região (Figura 3.2), é possível verificar a presença de atividades minerárias (alto curso da bacia do rio das Velhas) e atividades agropecuárias (concentradas no médio e baixo curso), que demandam, para o seu funcionamento, grande remoção de cobertura vegetal. Isso contribui para intensificar os processos erosivos, que, com a ação do escoamento pluvial, carregam

componentes dos solos expostos para dentro dos ambientes aquáticos (IGAM, 2010). Esses resultados demonstram o quão complexos são os processos de poluição do ecossistema aquático, ocorrendo inter-relações ou transferência de problemas de um compartimento (água, ar ou solo) para outro (CETESB, 2020).

Em águas naturais, o fósforo é encontrado, principalmente, devido às descargas de esgotos sanitários, cuja principal fonte é a matéria orgânica fecal e os detergentes em pó empregados em larga escala domesticamente. Outras fontes são os efluentes industriais (por exemplo de indústrias de fertilizantes, agrotóxicos, abatedouros, laticínios, entre outras) e as águas drenadas em áreas agrícolas e urbanas (CETESB, 2020).

Na estação chuvosa, as concentrações de P_T variaram entre 0,010 e 0,960 mg/L, e na estação seca, essa variação foi de 0,030 a 1,610 mg/L. Os maiores valores de concentração de fósforo total ocorreram no período seco (Figura 6.5 (e)), indicando que neste ponto da bacia os lançamentos de esgotos domésticos dos municípios da RMBH, localizados a montante da estação BV137, têm grande contribuição no incremento desse nutriente nas águas do rio das Velhas. Durante o período de seca, nessa área de estudo, fatores como aumento da luminosidade, diminuição da vazão e aumento da concentração dos nutrientes, favorecem o crescimento do fitoplâncton. Como consequência, o processo de eutrofização é favorecido (IGAM, 2010). Von Sperling e Jardim (2009) relataram a influência marcante das condições climatológicas no início da proliferação de algas em cursos d'água tropicais, apresentando como estudo de caso as florações de cianobactérias no rio das Velhas em 2007.

A temperatura da água é um parâmetro importante tanto para o status da qualidade da água quanto para o ecossistema, pois afeta as reações físicas, químicas e biológicas na água, a fotossíntese das plantas aquáticas, e controla os níveis de oxigênio dissolvido na água (CHANG; IMEN; VANNAH, 2015; CHAPMAN et al., 1992). No ponto monitorado, a temperatura da água apresentou valores entre 22,20 e 29,00°C na estação chuvosa, e entre 18,00 e 31,70°C na estação seca (Tabela 6.1). O comportamento observado mostrou valor mediano significativamente mais elevado na estação chuvosa (26,70 °C) se comparado à estação seca (22,80 °C) (Figura 6.5 (f)). A temperatura é influenciada por diferentes fatores (por exemplo, latitude, altitude, estação do ano, período do dia, taxa de fluxo e profundidade) e as variações observadas são parte das oscilações climáticas normais (CHAPMAN et al., 1992).

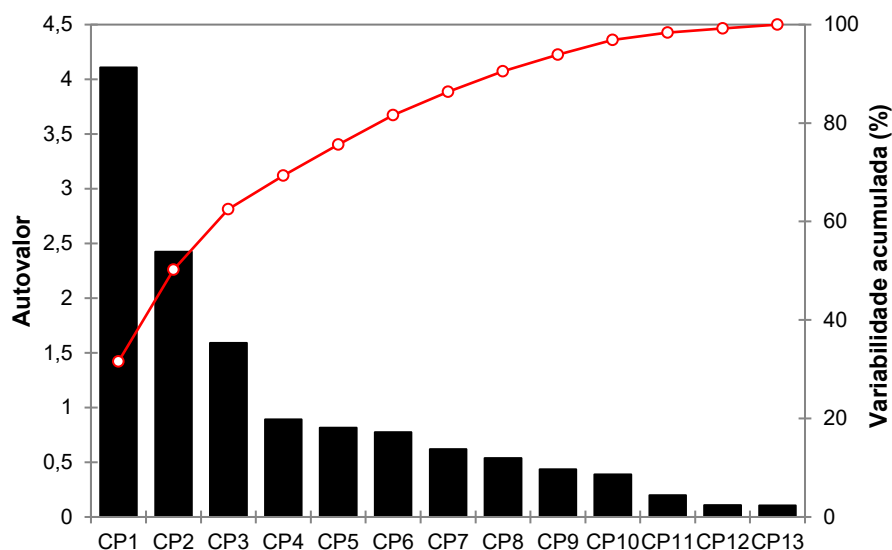
Em relação ao parâmetro vazão, conforme esperado, valores mais elevados foram registrados na estação chuvosa (Figura 6.5 (g)). A vazão apresentou valores entre 14,92 e 318,33 m³/s na estação chuvosa, e entre 12,86 e 117,95 m³/s na estação seca (Tabela 6.1).

6.3.2.2 Análise por Componentes Principais

A ACP teve como objetivo identificar quais foram os parâmetros mais importantes para caracterizar a qualidade da água na estação BV137, bem como obter um pequeno número de variáveis não correlacionadas que representasse toda a estrutura dos dados originais e a maior parte da variância encontrada nos mesmos.

A Figura 6.6 apresenta o gráfico *scree*, com os autovalores em ordem decrescente, e inversamente proporcional a eles estão indicadas as porcentagens da variabilidade acumulada, representadas pela linha em vermelho. Nesse gráfico, observa-se que três componentes principais apresentam autovalores maiores que um, e explicam juntos 62,47% da variabilidade total da qualidade da água da estação BV137. Então, as 13 variáveis analisadas foram reduzidas a três componentes não correlacionadas.

Figura 6.6 - Gráfico *scree* com os autovalores e a variabilidade acumulada das componentes principais obtidos na ACP



Fonte: Próprio autor.

As cargas (*loadings*) de cada parâmetro, para cada componente retida, estão apresentadas na Tabela 6.2. Em negrito, estão indicadas as maiores correlações entre

a variável e a CP (>0,75), e sublinhados estão os valores de carga que foram consideradas como moderadas (entre 0,75 e 0,50).

Tabela 6.2 - Resultado da ACP com as cargas obtidas para cada componente principal

Parâmetros	CP1	CP2	CP3
Arsênio total	0,364	0,485	<u>0,515</u>
Cloreto total	-0,793	0,236	0,374
Coliformes termotolerantes	0,433	<u>0,619</u>	-0,056
Condutividade elétrica in loco	-0,807	0,270	0,393
Demanda Bioquímica de Oxigênio	-0,005	<u>0,652</u>	0,042
Fósforo total	-0,479	0,454	0,179
Nitrato	-0,176	<u>-0,570</u>	0,310
Oxigênio dissolvido	-0,263	<u>-0,714</u>	0,382
pH in loco	-0,447	-0,060	<u>0,510</u>
Sólidos totais	0,763	0,049	0,459
Temperatura da água	0,257	-0,409	0,180
Turbidez	0,797	0,095	0,464
Vazão	0,833	-0,171	0,184
Autovalor	4,11	2,42	1,59
Variabilidade (%)	31,59	18,65	12,23
% acumulada	31,59	50,24	62,47

Negrito - Correlações fortes (valor absoluto maior que 0,75)

Sublinhado - Correlações moderadas (valor absoluto entre 0,50 - 0,75)

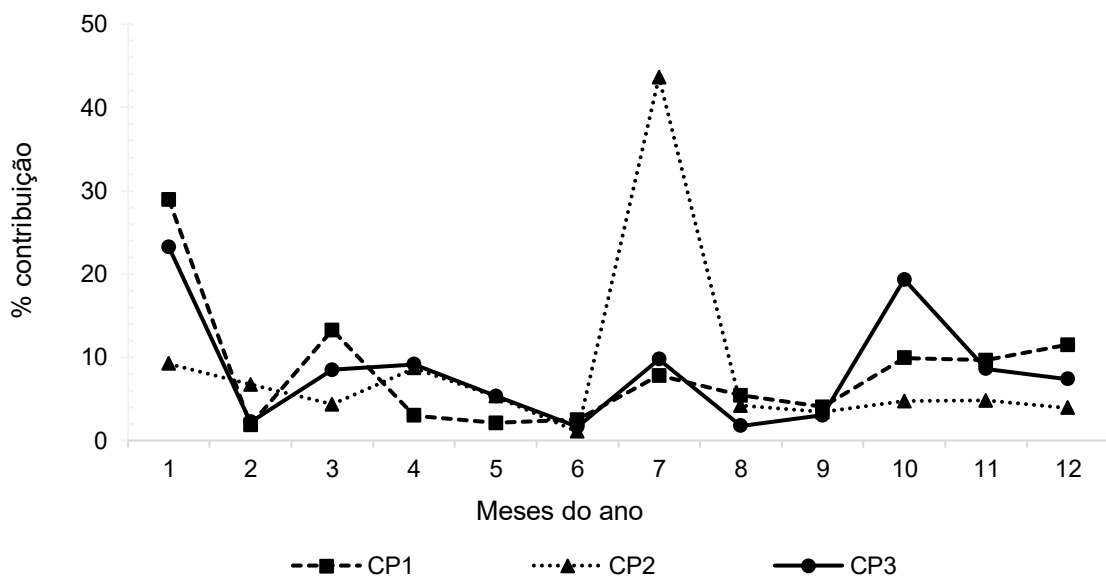
A primeira componente, que explica a maior variabilidade da qualidade da água (31,59%), está correlacionada positivamente com os parâmetros sólidos totais, turbidez e vazão, e negativamente com o cloreto total e a condutividade elétrica. Esse fator pode representar a contribuição de fontes de poluição difusa causada pelo carreamento de sólidos para o corpo d'água por meio do escoamento superficial, principalmente em locais de solo exposto, conforme discutido anteriormente.

A componente principal 2, que explica 18,65% da variabilidade total dos dados da qualidade da água, possui somente cargas moderadas e está correlacionada negativamente com os parâmetros nitrato e oxigênio dissolvido, e positivamente com a DBO e os coliformes termotolerantes. A alta correlação negativa com oxigênio dissolvido indica o alto grau de autodepuração do curso d'água. Essa componente representa a poluição orgânica originada, principalmente, dos lançamentos de esgoto doméstico. Como a estação BV137 está localizada a jusante de grandes centros urbanos, recebe a influência do lançamento de efluentes domésticos não tratados ou tratados de forma insuficiente, sendo um trecho crítico em termos de poluição (IGAM, 2010, 2018).

A terceira componente explica 12,23% da variabilidade da qualidade da água e está correlacionada positivamente com pH e arsênio total, indicando a poluição, principalmente, por atividades de mineração. O arsênio é um semimetal presente naturalmente no solo da bacia do rio das Velhas, mas o aumento de sua concentração nas águas superficiais é agravado por atividades de mineração desenvolvidas, sobretudo na região do Alto Velhas. Além disso, as fontes de arsênio podem estar relacionadas ao passivo ambiental (solo contaminado, pilhas de rejeitos antigas, minas antigas) e à presença de sedimentos enriquecidos com esse elemento nos cursos de água (BORBA; FIGUEIREDO; MATSCHULLAT, 2003; CHRISTOFARO; LEÃO, 2009; COSTA et al., 2015; DESCHAMPS; MATSCHULLAT, 2007; IGAM, 2010).

Para verificar a importância dos meses de monitoramento na explicação da variabilidade temporal da qualidade da água, a contribuição percentual que cada observação possui nas componentes retidas foi calculada. A Figura 6.7 mostra as contribuições, por mês de monitoramento, durante todo o período analisado para as três primeiras componentes. A contribuição total de uma componente, considerando o período de 12 meses, é de 100%.

Figura 6.7 - Variabilidade temporal das componentes principais, no período de monitoramento de 1997 a 2018, considerando o percentual de contribuição mensal



Fonte: Próprio autor.

A componente CP1, que representa a poluição difusa, apresentou os maiores percentuais de contribuição nos meses de janeiro (28,92%), março (13,27%) e dezembro (11,51%), que são meses de maior pluviosidade. Durante os meses do período seco (abril a setembro), essa componente apresenta contribuições mais baixas (inferiores a 10%), indicando uma menor influência do período.

Para a segunda componente, a sazonalidade não mostrou grande influência nos percentuais de contribuição, à exceção do mês de julho. A principal fonte de poluição associada a essa componente foram os lançamentos de esgoto doméstico, que são fontes pontuais e contínuas de poluição, o que pode explicar a menor influência dos meses no monitoramento. Destaca-se o percentual de contribuição obtido no mês de julho para essa componente (43,62%). De todas as amostras coletadas no mês de julho, a que apresentou maior percentual de contribuição foi a obtida no dia 20 de julho de 2001 (15,81%). Essa amostra foi coletada no mês que apresenta a menor precipitação acumulada (mediana de 1,2 mm) (Figura 6.3), além disso, no ano de 2001, ocorreu seca sazonal intensa. Essa amostra apresentou valores de concentração elevados para diversos parâmetros e vazão de 24,37 m³/s, que podem justificar esse percentual encontrado.

A terceira componente, que representa a poluição de atividades minerárias, assemelha-se, de forma geral, ao comportamento apresentado pelo primeiro componente. Os maiores percentuais de contribuição foram em meses de maior pluviosidade: janeiro (23,23%), outubro (19,32%) e novembro (8,61%). Os resultados indicam uma menor influência dos meses de amostragem do período seco na poluição por mineração na região.

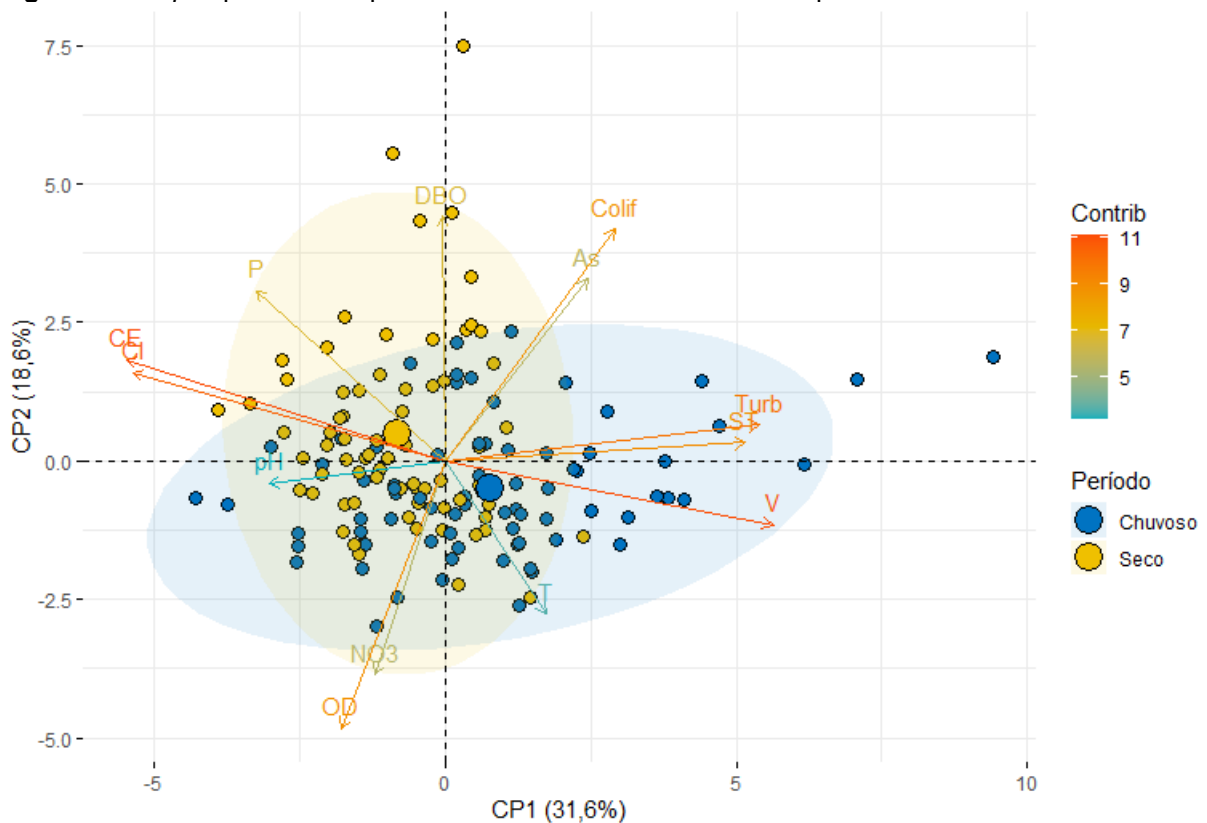
Em termos gerais, as variações nos percentuais de contribuição ao longo dos meses indicaram que as componentes estão sujeitas à variação temporal e que contribuições maiores indicam uma maior influência da sazonalidade. Os meses de janeiro, julho e outubro foram os que mais contribuíram para a explicação da variabilidade dos dados, considerando essas três componentes principais. Observa-se que nos meses de junho e setembro, os percentuais de contribuições para as três componentes são próximos.

A relação entre as componentes principais, tanto com as variáveis originais (colunas de uma matriz) quanto com as amostras originais (linhas de uma matriz), pode ser vista adicionalmente em *biplots*. O *biplot* fornece uma ferramenta útil de

análise de dados e permite a avaliação visual da estrutura de grandes matrizes de dados (GABRIEL, 1971).

A Figura 6.8 apresenta o gráfico *biplot* para as componentes CP1 e CP2 do conjunto de dados. Os pontos representam as observações projetadas no novo sistema de referência (CP1-CP2) e cada amostra é colorida com base no período climático em que foi coletada. As primeiras duas CPs respondem por 31,6% e 18,6%, respectivamente, da variação total no conjunto de dados. Por definição este é o melhor gráfico bidimensional com preservação da variância dos dados, representando 50,25% da variação total. As elipses de concentração (ou elipses de dados) delimitam a região de confiança de 95% em torno do valor central dos pontos analisados.

Figura 6.8 - *Biplot* para as componentes CP1 e CP2 considerando o período climático seco e chuvoso



Fonte: Próprio autor.

Embora haja uma evidente sobreposição dos períodos, o que é comum em dados de qualidade de água devido à variabilidade natural, o gráfico destaca uma tendência potencial de separação. Observa-se que as amostras coletadas no período chuvoso possuem maior influência, principalmente, da primeira componente CP1 (associado à poluição difusa), sendo os parâmetros sólidos totais, turbidez, vazão,

condutividade elétrica e cloreto os que mais contribuíram para a distinção das amostras. Em contrapartida, as amostras do período seco estão mais dispersas ao longo do eixo CP2 (associado à poluição orgânica), apresentando maior contribuição dos parâmetros DBO, OD, nitrato e coliformes termotolerantes.

6.4 CONCLUSÃO

As análises temporais da qualidade da água para a bacia de drenagem de uma estação localizada na calha do rio das Velhas foram realizadas usando estatísticas descritivas, teste U de Mann-Whitney, correlação de Spearman, e técnica estatística multivariada - a Análise por Componentes Principais. Os dados usados neste estudo compreendem 13 parâmetros de qualidade e quantidade da água, monitorados para o período de 1997 a 2018, além de dados pluviométricos da região para caracterização do período climático. As principais conclusões do estudo foram:

- A utilização dos dados pluviométricos e de vazão foram imprescindíveis para a análise da sazonalidade da qualidade da água;
- A análise dos dados pluviométricos mostrou que a região apresenta dois períodos climáticos bem definidos: um de maior precipitação (outubro a março) e outro de menor precipitação (abril a setembro), que apresentaram diferença significativa considerando os acumulados de chuvas.
- As estatísticas básicas calculadas para os parâmetros de qualidade da água considerando cada período climático, expressaram mudanças evidentes entre eles, indicando a variabilidade temporal na composição química da água.
- Os resultados do teste de Mann-Whitney e da correlação de Spearman indicaram que, dentre os parâmetros analisados, aqueles que apresentaram diferença significativa (valor- $p < 0,05$) entre os períodos climáticos foram: cloreto total, condutividade elétrica, fósforo total, sólidos totais, temperatura da água, turbidez e vazão.
- Os resultados da ACP indicaram que três CPs são responsáveis por explicar 62,47% da variabilidade total da qualidade da água. Cada componente foi associada a um tipo de fonte de poluição: CP1 representa a poluição difusa, CP2 a poluição orgânica, e CP3 a contaminação por atividades de mineração.

- As principais fontes de poluição identificadas na região foram as atividades minerárias e agropecuárias, além dos lançamentos de esgotos domésticos não tratados ou tratados de forma insuficiente.
- Os resultados do cálculo da contribuição percentual que cada observação possui nas componentes retidas indicou que os meses de janeiro, julho e outubro foram os que mais contribuíram para a explicação da variabilidade dos dados, e que nos meses de junho e setembro, a sazonalidade apresentou menor influência nos resultados de qualidade de água.
- O gráfico *biplot* destacou uma tendência potencial de separação entre as amostras coletadas em cada período climático, e indicou que as amostras coletadas no período chuvoso apresentaram maior influência, principalmente, de CP1 (poluição não pontual), e as amostras do período seco, maior influência de CP2 (poluição pontual).
- As análises realizadas para uma única estação de monitoramento demonstraram a importância de estudos locais para uma caracterização mais detalhada das características dos recursos hídricos em um trecho de um curso de água.

CAPÍTULO 7: CONCLUSÕES GERAIS

Neste capítulo, são apresentadas as principais conclusões obtidas a partir da aplicação da Análise por Componentes Principais a dados de monitoramento de qualidade de água superficial, de acordo com as abordagens propostas.

- As metodologias propostas e aplicadas a uma única estação de monitoramento (abordagem local) demonstraram a importância de estudos locais para a descrição detalhada das condições de qualidade de água do trecho do curso de água avaliado. Na prática, a combinação de abordagens globais, comumente realizadas, e locais podem levar a um melhor entendimento sobre as condições de qualidade da água.
- A Análise por Componentes Principais (ACP), associada ao filtro de Hampel, demonstrou o potencial da técnica ACP para a identificação de padrões de comportamento e classificação de cenários críticos de poluição. Nessa aplicação, os gráficos de *scores* da ACP (*score plots*) foram utilizados para a identificação de eventos críticos e associados às fontes de poluição características da área de estudo.
- Os resultados obtidos da aplicação de ACP, associada ao filtro de Hampel, podem ser utilizados como sinal de alerta sobre desvios em relação a uma condição de referência de qualidade de água, e para diagnóstico, por ser possível estabelecer, *a priori*, associações entre essas regiões espaciais e fontes de poluição.
- A abordagem combinada integrando Controle Estatístico de Processo Multivariado e Análise por Componentes Principais (CEPM-ACP) mostrou-se promissora para a detecção de condições críticas de poluição da água e o fornecimento de alerta, de acordo com os dados históricos de observação. Nas análises, foram utilizadas duas métricas de monitoramento, as estatísticas T^2 e Q , que foram capazes de identificar 79,73% e 72,97% dos dados extremos.
- Na aplicação de ACP para análise da variabilidade temporal da qualidade da água, foi obtido um número de componentes principais reduzido, que foram capazes de explicar a maior parte da variância dos dados originais. As análises dos *loadings* permitiu associar cada componente principal retida a uma fonte de poluição. Foram utilizados também os resultados dos percentuais de contribuição que cada observação possui nas componentes principais, sendo possível a identificação dos meses do ano em que a sazonalidade apresentou maior e menor influência nos resultados de qualidade de água. Por fim, a análise dos *scores* destacou a tendência potencial de separação entre as amostras coletadas em cada período climático

analisado.

- As metodologias utilizadas neste estudo podem ser aplicadas a outras estações de monitoramento.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, 2010.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **ABNT NBR ISO/IEC 17025**: Requisitos gerais para a competência de laboratórios de ensaio e calibração. Rio de Janeiro: ABNT, 2017.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Novo Portal vai disponibilizar informações sobre águas de superfície no Brasil**, 2010. Disponível em: <<https://www.gov.br/mma/pt-br/noticias/novo-portal-vai-disponibilizar-informacoes-sobre-aguas-de-superficie-no-brasil>>. Acesso em: 12 jan. 2021.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Inventário das estações fluviométricas**. Brasília: Agência Nacional de Águas, Superintendência de Gestão da Rede Hidrometeorológica, 2009. 196p.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Cuidando das águas: soluções para melhorar a qualidade dos recursos hídricos**. Brasília: Agência Nacional de Águas, 2011. 154 p.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Relatório de Esgotamento Sanitário Municipal - Atlas Esgotos - Despoluição de Bacias Hidrográficas**. Brasília - DF: Agência Nacional de Águas, 2017a. Disponível em: <http://portal1.snirh.gov.br/arquivos/Atlas_Esgoto/Minas_Gerais/Relatorio_Geral/Lag oa_Santa.pdf>.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Atlas esgotos: despoluição de bacias hidrográficas**. Brasília: Agência Nacional de Águas, Secretaria Nacional de Saneamento Ambiental, 2017b. 88p. Disponível em: <http://arquivos.ana.gov.br/imprensa/publicacoes/ATLASESGOTOSDespoluicaoodeBaciasHidrograficas-ResumoExecutivo_livro.pdf>.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO - ANA. **Conjuntura dos recursos hídricos no Brasil 2020: informe anual**. Brasília: Agência Nacional de Águas e Saneamento Básico, 2020a. 118 p.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO - ANA. **Portal HidroWeb v3.1.1**. Disponível em: <<http://www.snirh.gov.br/hidroweb/apresentacao>>. Acesso em: 19 ago. 2020b.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Atlas esgotos: atualização da base de dados de estações de tratamento de esgotos no Brasil**. Brasília: Agência Nacional de Águas. 2020c. 44 p.

ALCALA, C. F.; QIN, S. J. Analysis and generalization of fault diagnosis methods for process monitoring. **Journal of Process Control**, v. 21, n. 3, p. 322–330, 2011.

AMERICAN PUBLIC HEALTH ASSOCIATION - APHA; AMERICAN WATER WORKS ASSOCIATION - AWWA; WATER ENVIRONMENT FEDERATION - WEF. **Standard Methods for the Examination of Water and Wastewater**. 23. ed. Washington: American Public Health Association, 2017.

ARCISZEWSKI, T. J. et al. Developing and applying control charts to detect changes in water chemistry parameters measured in the Athabasca River near the oil sands: A tool for surveillance monitoring. **Environmental Toxicology and Chemistry**, v. 37, n. 9, p. 2296–2311, 2018.

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.

BARAKAT, A. et al. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. **International Soil and Water Conservation Research**, v. 4, n. 4, p. 284–292, 2016.

BARTRAM, J.; BALLANCE, R.; WORLD HEALTH ORGANIZATION - WHO; UNITED NATIONS ENVIRONMENT PROGRAMME - UNEP. **Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programs**. London: E & FN Spon, 1996. Disponível em: <<https://apps.who.int/iris/handle/10665/41851>>.

BEHMEL, S. et al. Water quality monitoring strategies — A review and future perspectives. **Science of The Total Environment**, v. 571, p. 1312–1329, 2016.

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.

BHASIN, S. et al. Control Chart Model for Assessment of Water Quality of a Tropical River-Kshipra Ujjain, India. **Haya: The Saudi Journal of Life Sciences**, v. 1, n. 1, p. 51–64, 2016.

BHAT, S. A. et al. Statistical Assessment of Water Quality Parameters for Pollution Source Identification in Sukhnag Stream: An Inflow Stream of Lake Wular (Ramsar Site), Kashmir Himalaya. **Journal of Ecosystems**, v. 2014, p. 1–18, 2014.

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.

BOHM, K.; SMIDT, E.; TINTNER, J. Application of Multivariate Data Analyses in Waste Management. In: **Multivariate Analysis in Management, Engineering and the Sciences**. [s.l.] InTech, 2013.

BOLLMANN, H. A.; EDWIGES, T. Avaliação da qualidade das águas do Rio Belém, Curitiba-PR, com o emprego de indicadores quantitativos e perceptivos. **Engenharia Sanitária e Ambiental**, v. 13, n. 4, p. 443–452, 2008.

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.

BRASIL. **Lei n° 9.984**, de 17 de julho de 2000. Dispõe sobre a criação da Agência Nacional de Águas e Saneamento Básico (ANA), entidade federal de implementação da Política Nacional de Recursos Hídricos, integrante do Sistema Nacional de Gerenciamento de Recursos Hídricos (Singreh) e responsável pela instituição de normas de referência para a regulação dos serviços públicos de saneamento básico.

BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, n. 9, p. 2812–2831, 2014.

BU, H. et al. Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. **Ecotoxicology and Environmental Safety**, v. 73, n. 5, p. 907–913, 2010.

- CALAZANS, G. M. et al. The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. **Environmental Monitoring and Assessment**, v. 190, n. 8, p. 491, 2018a.
- CALAZANS, G. M. et al. 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, 2018b.
- CAMACHO, J. et al. PCA-based multivariate statistical network monitoring for anomaly detection. **Computers and Security**, v. 59, p. 118–137, 2016.
- CARLSON, N. Aquatic Sensor Workgroup - ASW. **Watersensors**, 2016. Disponível em: < <http://www.watersensors.org/about.html>>. Acesso em: 11 jan. 2020.
- CARVALHO, M. A. G. **Métodos Estatísticos Para Análise De Dados De Monitoração Ambiental**. 2003. 134 f. Tese (Doutorado em Ciências) – Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, São Paulo. 2003.
- CHANG, F. et al. Modeling water quality in an urban river using hydrological factors – Data driven approaches. **Journal of Environmental Management**, v. 151, p. 87–96, 2015.
- CHANG, N.-B.; IMEN, S.; VANNAH, B. Remote Sensing for Monitoring Surface Water Quality Status and Ecosystem State in Relation to the Nutrient Cycle: A 40-Year Perspective. **Critical Reviews in Environmental Science and Technology**, v. 45, n. 2, p. 101–166, 2015.
- CHANG, N.; CHEN, H. W.; NING, S. K. Identification of river water quality using the Fuzzy Synthetic Evaluation approach. **Journal of Environmental Management**, v. 63, n. 3, p. 293–305, 2001.
- CHAPMAN, D. V.; WORLD HEALTH ORGANIZATION - WHO; UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION - UNESCO; UNITED NATIONS ENVIRONMENT PROGRAMME - UNEP. **Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring**. 2 ed. London: E & FN Spon, 1996. Disponível em: <<https://apps.who.int/iris/handle/10665/41850>>.
- CHEN, K. H. **Data-Rich Multivariate Detection and Diagnosis Using Eigenspace Analysis**. 2001. 162 f. Tese (Doctor of Philosophy) – Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge. 2001.
- CHOWDURY, M. S. U. et al. IoT based real-time river water quality monitoring system. **Procedia Computer Science**, v. 155, p. 161–168, 2019.
- CHRISTOFARO, C.; LEÃO, M. Caracterização temporal do arsênio nos cursos d'água da bacia hidrográfica do Rio das Velhas, MG, Brasil, ao longo de uma década (1998 - 2007). **Ambiente e Água - An Interdisciplinary Journal of Applied Science**, v. 4, n. 3, p. 54–66, 2009.
- CHRISTOFARO, C.; LEÃO, M. M. D. Tratamento de dados censurados em estudos ambientais. **Química Nova**, v. 37, n. 1, p. 104–110, 2014.
- COLLISCHONN, W.; TASSI, R. **Introduzindo Hidrologia**. [s.l.] Instituto de Pesquisas Hidráulicas - UFRGS, 2008.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO DAS VELHAS - CBH VELHAS. **Plano Diretor de Recursos Hídricos da Bacia Hidrográfica do Rio das Velhas 2015: Resumo Executivo**. Belo Horizonte: Comitê da Bacia Hidrográfica do Rio das Velhas, 2015a. 233 p.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO DAS VELHAS - CBH VELHAS. **Plano Diretor de Recursos Hídricos da Bacia Hidrográfica do Rio das Velhas 2015: Plano Diretor Consolidado – Volume I**. Belo Horizonte: Comitê da Bacia Hidrográfica do Rio das Velhas, 2015b. 316p.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO DAS VELHAS - CBH VELHAS. **A bacia hidrográfica do rio das Velhas**. Disponível em: <<http://cbhvelhas.org.br/a-bacia-hidrografica-do-rio-das-velhas/>>. Acesso em: 25 fev. 2019.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO - CBHSF. **Resumo Executivo do Plano de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco 2016-2025**. Alagoas: Comitê da Bacia Hidrográfica do Rio São Francisco, 2016. 300 p.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO - CBHSF. **A Bacia**. Disponível em: <<https://cbhsaofrancisco.org.br/a-bacia/>>. Acesso em: 20 fev. 2020.

COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO - CETESB. **Qualidade das águas interiores no estado de São Paulo 2019**. São Paulo: Companhia Ambiental do Estado de São Paulo, 2020.

COMPANHIA DE PESQUISA DE RECURSOS MINERAL - CPRM. Projeto Geoquímica MultiUsos no Estado de Minas Gerais: **Altas Geoquímico da Bacia do Rio das Velhas**. Belo Horizonte: Companhia de Pesquisa de Recursos Mineral, 2010.

CONSERVATION ONTARIO. **Overview of Integrated Watershed Management in Ontario**. Canadá: Conservation Ontario, 2010.

CORBETT, C. J.; PAN, J. N. Evaluating environmental performance using statistical process control techniques. **European Journal of Operational Research**, v. 139, n. 1, p. 68–83, 2002.

CORDOVA, J. E. **Análise Integrada da Influência das Alterações do Uso e Ocupação do Solo Na Qualidade das Águas Superficiais das Bacias dos Rios Pará e das Velhas - MG**. 2020. Tese (Doutorado em Saneamento, Meio Ambiente e Recursos Hídricos) – Departamento de Engenharia Sanitária e Ambiental, Universidade Federal de Minas Gerais, Belo Horizonte. 2020.

COSTA, R. V. F. et al. Geochemical mapping of arsenic in surface waters and stream sediments of the Quadrilátero Ferrífero, Brazil. **Rem: Revista Escola de Minas**, v. 68, n. 1, p. 43–51, 2015.

COSTA, E. P. et al. 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.

COTHERN, C. R.; ROSS, N. P. **Environmental Statistics, Assessment, and Forecasting**. 1. ed. [s.l.] 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.

- CRUZ, D. V.; OLIVEIRA, M. R. G.; FILHO, M. C. Monitoring pH with Quality Control Based Geostatistics Methodology. **IEEE LATIN AMERICA TRANSACTIONS**, v. 14, n. 12, p. 4787–4791, 2016.
- CRUZ, M. A. S. et al. 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.
- DAVIES, L.; GATHER, U. The Identification of Multiple Outliers. **Journal of the American Statistical Association**, v. 88, n. 423, p. 782–792, 1993.
- DAVIS, J. C. **Statistics and Data Analysis in Geology**. 3. ed. New York: John Wiley & Sons, 2002.
- DESCHAMPS, E.; MATSCHULLAT, J. **Arsênio Antropogênico e Natural: Um Estudo em Regiões do Quadrilátero Ferrífero**. Belo Horizonte: Fundação Estadual do Meio Ambiente, 2007.
- DI BLASI, J. I. P. et al. Analysis and detection of outliers in water quality parameters from different automated monitoring stations in the Miño river basin (NW Spain). **Ecological Engineering**, v. 60, n. September 1993, p. 60–66, 2013.
- DIAMANTINI, E. et al. Driver detection of water quality trends in three large European river basins. **Science of the Total Environment**, v. 612, p. 49–62, 2018.
- DIAMANTOPOULOU, M. J.; ANTONOPOULOS, V. Z.; PAPAMICHAIL, D. M. Cascade Correlation Artificial Neural Networks for Estimating Missing Monthly Values of Water Quality Parameters in Rivers. **Water Resources Management**, v. 21, n. 3, p. 649–662, 2007.
- DU, X. et al. Water quality assessment with hierarchical cluster analysis based on Mahalanobis distance. **Environmental Monitoring and Assessment**, v. 189, n. 7, 2017.
- ENVIRONMENTAL PROTECTION AGENCY - EPA. **Data Quality Assessment: Statistical Methods for Practitioners**. Washington: United States Environmental Protection Agency, 2006.
- FERREIRA, D. B. et al. **Análise se séries temporais pluviométricas na porção mineira da bacia hidrográfica do rio São Francisco**. III Simpósio da Bacia Hidrográfica do Rio São Francisco. **Anais...**Belo Horizonte, 2020.
- FERRER, A. Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. **Quality Engineering**, v. 19, n. 4, p. 311–325, 2007.
- FOLLADOR, F. A. C. **Controle estatístico de processo aplicado à qualidade de águas superficiais**. 2010. 118 f. Tese (Doutorado em Engenharia Agrícola) – Programa de Pós-Graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel. 2010.
- FOLLADOR, F. A. C. et al. Controle de Qualidade da Água Medido Através de Cartas de Controle de Shewhart, Cusum e MMEP. **Engenharia Ambiental**, v. 9, n. 3, p. 183–197, 2012.
- FORIO, M. A. E. et al. Bayesian belief network models to analyse and predict ecological water quality in rivers. **Ecological Modelling**, v. 312, p. 222–238, 2015.
- FUNDAÇÃO RENOVA. **Programa de monitoramento qualitativo e quantitativo**

sistemático (PMQQS). Disponível em: <<https://www.fundacaorenova.org/programa/monitoramento-da-bacia-do-rio-doce/>>. Acesso em: 21 jan. 2020.

FUNDAÇÃO RURAL MINEIRA - RURALMINAS; UNIVERSIDADE FEDERAL DE VIÇOSA - UFV; INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS - IGAM. **Atlas Digital das Águas de Minas - Uma ferramenta para o planejamento e gestão dos recursos hídricos**. 2 ed. [s.l.] [s.n]. 2007. 78 p.

FUNDAÇÃO RURAL MINEIRA - RURALMINAS; UNIVERSIDADE FEDERAL DE VIÇOSA - UFV; INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS - IGAM. Consulta Espacial Georreferenciada: Regiões hidrologicamente homogêneas – bacia do rio das Velhas. Disponível em: <http://www.atlasdasaguas.ufv.br/velhas/Regioes_homogeneas/Regioes_homogenea_s.html>. Acesso em: 03 jan. 2021.

GABRIEL, K. R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. **Biometrika**, v. 58, n. 3, p. 453, 1971.

GALLAGHER, N. B. et al. Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness through Model Updating. **IFAC Proceedings Volumes**, v. 30, n. 9, p. 79–84, 1997.

GATICA, E. A. et al. 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.

GE, Z.; SONG, Z. **Multivariate Statistical Process Control**. 1. ed. London: Springer London, 2013.

GHOLIZADEH, M. H.; MELESSE, A. M.; REDDI, L. Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. **Science of The Total Environment**, v. 566–567, p. 1552–1567, 2016.

GILBERT, R. O. **Statistical Methods for Environmental Pollution Monitoring**. 1. ed. New York: John Wiley & Sons, 1987.

GOVERNMENT OF WESTERN AUSTRALIA. **Water quality monitoring program design - A guideline for field sampling for surface water quality monitoring programs**. Department of Water, 2009.

GRONEWOLD, A. D.; BORSUK, M. E. Improving water quality assessments through a hierarchical bayesian analysis of variability. **Environmental Science and Technology**, v. 44, n. 20, p. 7858–7864, 2010.

HAIR, J. F. et al. **Análise Multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HE, B. et al. Estimating monthly total nitrogen concentration in streams by using artificial neural network. **Journal of Environmental Management**, v. 92, p. 172–177, 2011.

HELSEL, D. R. **Statistics for Censored Environmental Data Using Minitab® and R**. 2. ed. Nova Jersey: John Wiley & Sons, Inc., 2011.

HELSEL, D. R.; HIRSCH, R. M. Techniques of Water Resources Investigations. In: **Statistical Methods in Water Resources**. Book 4 ed. [s.l.] U.S. Geological Survey, 2002. p. 522.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 6, p. 417–441, 1933.

ICAGA, Y. Fuzzy evaluation of water quality classification. **Ecological Indicators**, v. 7, n. 3, p. 710–718, 2007.

IGLESIAS, C. et al. Shewhart-type control charts and functional data analysis for water quality analysis based on a global indicator. **Desalination and Water Treatment**, v. 57, n. 6, p. 2669–2684, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Panorama - Lagoa Santa. Disponível em: <<https://cidades.ibge.gov.br/brasil/mg/lagoa-santa/panorama>>. Acesso em: 13 fev. 2021.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Monitoramento da qualidade das águas superficiais na sub-bacia do rio das Velhas em 2009**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2010. 206 p.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Identificação de municípios com condição crítica para a qualidade de água na bacia do rio das Velhas**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2013. 48p. Disponível em: <http://www.igam.mg.gov.br/images/stories/ARQUIVO_SANEAMENTO/estudo-saneamento-rio-das-velhas.pdf>.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Análise de Tendência da Qualidade das Águas na Bacia do rio das Velhas no período de 2000 a 2013**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2014. 72 p.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Avaliação da Qualidade das Águas Superficiais de Minas Gerais em 2017: Resumo Executivo Anual**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2018. 189 p. Disponível em: <<http://200.198.57.118:8080/jspui/handle/123456789/2336>>.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS - IGAM. **Avaliação da Qualidade das Águas Superficiais de Minas Gerais em 2018: Resumo Executivo Anual**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2019a. 327 p. Disponível em: <<http://200.198.57.118:8080/jspui/handle/123456789/3210>>.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS - IGAM. **Gestão e Situação das Águas de Minas Gerais 2019**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2019b. 160 p.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Monitoramento da Qualidade das Águas**. Disponível em: <<http://www.igam.mg.gov.br/component/content/article/16-duvidas/2438--acessar-dados-e-boletins-de-qualidade-das-aguas-em-minas-gerais-temporario>>. Acesso em: 25 dez. 2020.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS. **Ficha descritiva de amostragem**. (Documento solicitado ao Igam). Acesso em 21 jan. 2021.

JABBAR, F. K.; GROTE, K. Statistical assessment of nonpoint source pollution in agricultural watersheds in the Lower Grand River watershed, MO, USA. **Environmental Science and Pollution Research**, v. 26, n. 2, p. 1487–1506, 2019.

JACKSON, J. E. **A User's Guide to Principal Components**. New York: John Wiley & Sons, 2003.

JARDIM, F. A. et al. Fatores determinantes das florações de cianobactérias na água do Rio Doce, Minas Gerais, Brasil. **Engenharia Sanitaria e Ambiental**, v. 19, n. 3, p. 207–218, 2014.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6. ed. New Jersey: Prentice Hall, 2007.

JOLLIFFE, I. T. **Interpreting Principal components: Examples**. 2. ed. New York: Springer-Verlag, 2002.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 374, n. 2065, 2016.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, n. 3, p. 187–200, 1958.

KAZI, T. G. G. et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. **Ecotoxicology and Environmental Safety**, v. 72, n. 2, p. 301–309, 2009.

KHATRI, N.; TYAGI, S. Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. **Frontiers in Life Science**, v. 8, n. 1, p. 23–39, 2015.

KORZENOWSKI, A. L. **Premissas E Suposições Para Construção De Gráficos De Controle: Um Framework para Verificação**. 2009. 102 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2009.

KOURTI, T. Process analysis and abnormal situation detection: from theory to practice. **IEEE Control Systems**, v. 22, n. 5, p. 10–25, 2002.

KOURTI, T.; MACGREGOR, J. F. Multivariate SPC methods for process and product monitoring. **Journal of Quality Technology**, v. 28, n. 4, p. 409–428, 1996.

KOWALKOWSKI, T. et al. Application of chemometrics in river water classification. **Water Research**, v. 40, n. 4, p. 744–752, 2006.

KRISHNARAJ, A.; DEKA, P. C. Spatial and temporal variations in river water quality of the Middle Ganga Basin using unsupervised machine learning techniques. **Environmental Monitoring and Assessment**, v. 192, n. 12, p. 744, 2020.

LATTIN, J.; CARROLL, J. D.; GREEN, P. **Análise de dados multivariados**. 1º ed. São Paulo: Cengage Learning, 2011.

LEE, P. H. et al. The effect of the individual chart with variable control limits on the river pollution monitoring. **Quality and Quantity**, v. 47, n. 4, p. 1803–1812, 2013.

LEGENDRE, P.; LEGENDRE, L. **Numerical Ecology**. 3. ed. Amsterdam: Elsevier, 2012.

LERMONTOV, A. et al. River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. **Ecological Indicators**, v. 9, n. 6, p. 1188–1197, 2009.

LI, R.; ZOU, Z.; AN, Y. Water quality assessment in Qu River based on fuzzy water pollution index method. **Journal of Environmental Sciences**, v. 50, p. 87–92, 2016.

- LI, Y.; XU, L.; LI, S. Water Quality Analysis of the Songhua River Basin Using Multivariate Techniques. **Journal of Water Resource and Protection**, v. 01, n. 02, p. 110–121, 2009.
- LIN, B. et al. A systematic approach for soft sensor development. **Computers & Chemical Engineering**, v. 31, n. 5–6, p. 419–425, 2007.
- LIU, C.-W.; LIN, K.-H.; KUO, Y.-M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. **Science of the Total Environment**, v. 313, n. 1–3, p. 77–89, 2003.
- LIU, J. et al. Characterizing and explaining spatio-temporal variation of water quality in a highly disturbed river by multi-statistical techniques. **SpringerPlus**, v. 5, n. 1, p. 1171, 2016.
- MA, F. et al. A Bayesian method for comprehensive water quality evaluation of the Danjiangkou Reservoir water source area, for the middle route of the South-to-North Water Diversion Project in China. **Frontiers of Earth Science**, v. 8, n. 2, p. 242–250, 2014.
- MAGALHÃES JÚNIOR, H.; LOPES, F. A. L.; MACEDO, D. R. Diagnóstico multitemporal do uso e cobertura da terra e qualidade das águas na bacia do rio Jequitinhonha em Minas Gerais como subsídio à gestão dos recursos hídricos superficiais. **Revista Espinhaço**, v. 8, n. 2, p. 47–57, 2019.
- MAILLARD, P.; SANTOS, N. A. P. A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed - Brazil. **Journal of Environmental Management**, v. 86, n. 1, p. 158–170, 2008.
- MANLY, B. F. J. **Métodos Estatísticos Multivariados: Uma Introdução**. 3. ed. Porto Alegre: Bookman, 2008.
- MCBRIDE, G. B. **Using Statistical Methods for Water Quality Management**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005.
- MINAS GERAIS (Estado). **Deliberação Normativa Conjunta COPAM/CERH-MG nº 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.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005. 297 p.
- MINISTÉRIO DO MEIO AMBIENTE - MMA. **Conferência das Nações Unidas Sobre Meio Ambiente e Desenvolvimento Capítulo 8**. Disponível em: <<https://www.mma.gov.br/responsabilidade-socioambiental/agenda-21/agenda-21-global/item/670.html>>. Acesso em: 8 ago. 2019a.
- MINISTÉRIO DO MEIO AMBIENTE. **Áreas Prioritárias para Conservação da Biodiversidade Brasileira**, 2019b. Disponível em: <<http://areasprioritarias.mma.gov.br/faqs>>. Acesso em: 24 jun. 2020.
- MINISTÉRIO DO MEIO AMBIENTE. **Ações de revitalização na bacia hidrográfica do Rio São Francisco**. Brasília: Ministério do Meio Ambiente, Secretaria de Recursos Hídricos e Ambiente Urbano, 2009. 167 p.

- MOHAMED, I. et al. 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.
- MONICA, N.; CHOI, K. S. Temporal and spatial analysis of water quality in Saemangeum watershed using multivariate statistical techniques. **Paddy and Water Environment**, v. 14, n. 1, p. 3–17, 2016.
- MONTGOMERY, D. C. **Introdução ao Controle Estatístico de Qualidade**. 7. ed. Rio de Janeiro: LTC, 2016.
- MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 5. ed. Rio de Janeiro: LTC, 2015.
- MORRISON, L. W. The Use of Control Charts to Interpret Environmental Monitoring Data. **Natural Areas Journal**, v. 28, n. 1, p. 66–73, 2008.
- NOMIKOS, P. Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis. **ISA Transactions**, v. 35, n. 3, p. 259–266, 1996.
- NOORI, R. et al. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. **Desalination**, v. 260, n. 1–3, p. 129–136, 2010.
- NOSKIEVIČOVÁ, D. Complex Control Chart Interpretation. **International Journal of Engineering Business Management**, v. 5, p. 13, 2013.
- OCAMPO-DUQUE, W. et al. Assessing water quality in rivers with fuzzy inference systems: A case study. **Environment International**, v. 32, n. 6, p. 733–742, 2006.
- OCAMPO-DUQUE, W.; SCHUHMACHER, M.; DOMINGO, J. L. A neural-fuzzy approach to classify the ecological status in surface waters. **Environmental Pollution**, v. 148, n. 2, p. 634–641, 2007.
- OLIVEIRA, D. G. et al. Land use and its impacts on the water quality of the Cachoeirinha Invernada Watershed, Guarulhos (SP). **Ambiente e Água - An Interdisciplinary Journal of Applied Science**, v. 13, n. 1, p. 1, 2018.
- OLIVEIRA, J. C. et al. Spatial-temporal analysis of the surface water quality of the Pará River Basin through statistical techniques. **Ambiente e Água - An Interdisciplinary Journal of Applied Science**, v. 14, n. 1, p. 1, 2019.
- OLIVEIRA, L. M.; MAILLARD, P.; PINTO, É. J. A. Modeling the effect of land use/land cover on nitrogen, phosphorous and dissolved oxygen loads in the Velhas River using the concept of exclusive contribution area. **Environmental Monitoring and Assessment**, v. 188, n. 6, p. 333, 2016.
- OLIVEIRA, L. M.; MAILLARD, P.; PINTO, E. J. A. Application of a land cover pollution index to model non-point pollution sources in a Brazilian watershed. **CATENA**, v. 150, p. 124–132, 2017.
- OLIVEIRA, S. C. et al. Qualidade das águas superficiais do Médio São Francisco após a implantação dos perímetros irrigados de Gortuba/Lagoa Grande e Jaíba. **Engenharia Sanitária e Ambiental**, v. 22, n. 4, p. 711–721, 2017.

OLIVEIRA, S. C.; PINTO, C. C.; SOARES, A. L. C. O Programa Águas de Minas e sua importância para a avaliação da qualidade das águas do estado de Minas Gerais. In: **Compartilhando experiências das águas de Minas Gerais**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2018. v. 2, p. 51–56.

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.

ORDÓÑEZ, M. L. R. **Multivariate statistical process control and case-based reasoning for situation assessment of sequencing batch reactors**. 2008. 141 f. Tese (Doutorado em Tecnologias de la Informació) – Departament d'Enginyeria Elèctrica, Electrònica I Automàtica, Universitat de Girona, Girona, 2008.

PALANI, S.; LIONG, S.; TKALICH, P. An ANN application for water quality forecasting. **Marine Pollution Bulletin**, v. 56, n. 9, p. 1586–1597, 2008.

PARINET, B.; LHOTE, A.; LEGUBE, B. Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. **Ecological Modelling**, v. 178, n. 3–4, p. 295–311, 2004.

PARK, S.-Y. et al. Design of a water quality monitoring network in a large river system using the genetic algorithm. **Ecological Modelling**, v. 199, n. 3, p. 289–297, 2006.

PATI, S. et al. Assessment of water quality using multivariate statistical techniques in the coastal region of Visakhapatnam, India. **Environmental Monitoring and Assessment**, v. 186, n. 10, p. 6385–6402, 2014.

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.

PELLERIN, B. **Applications of optical sensors for high-frequency water-quality monitoring and research**. The SMART water grid International conference. Anais...Incheon, South Korea, 2015. Disponível em: <<https://pubs.er.usgs.gov/publication/70156710>>.

PEREIRA, M. A. S. et al. A multivariate statistical approach to the integration of different land-uses, seasons, and water quality as water resources management tool. **Environmental Monitoring and Assessment**, v. 191, n. 9, p. 539, 2019.

PERONA, E.; BONILLA, I.; MATEO, P. Spatial and temporal changes in water quality in a Spanish river. **Science of The Total Environment**, v. 241, n. 1–3, p. 75–90, 1999.

PESSOA, J. O.; ORRICO, S. R. M.; LORDÊLO, M. S. Qualidade da água de rios em cidades do Estado da Bahia. **Engenharia Sanitaria e Ambiental**, v. 23, n. 4, p. 687–696, 2018.

PHUNG, D. et al. 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.

PINTO, C. C. et al. Trend analysis of the load and concentration of physical, chemical and biological parameters in the Verde River Basin. **Revista Brasileira de Recursos Hídricos**, v. 22, 2017.

- PINTO, C. C.; ALMEIDA, K. D. B.; OLIVEIRA, S. C. Spatial evaluation of the water quality from the Velhas river channel, in the state of Minas Gerais. **Periódico Tchê Química**, v. 15, n. 30, 2018.
- PINTO, C. C.; CALAZANS, G. M.; OLIVEIRA, S. C. Assessment of spatial variations in the surface water quality of the Velhas River Basin, Brazil, using multivariate statistical analysis and nonparametric statistics. **Environmental Monitoring and Assessment**, v. 191, n. 3, p. 164, 2019.
- QIAN, S. S.; MILTNER, R. J. A continuous variable Bayesian networks model for water quality modeling: A case study of setting nitrogen criterion for small rivers and streams in Ohio, USA. **Environmental Modelling and Software**, v. 69, p. 14–22, 2015.
- QIN, S. J. Statistical process monitoring: basics and beyond. **Journal of Chemometrics**, v. 17, n. 8–9, p. 480–502, 2003.
- QIU, P. **Introduction to statistical process control**. New York: Taylor & Francis Group, LLC, 2014.
- RAZMKHAH, H.; ABRISHAMCHI, A.; TORKIAN, A. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). **Journal of Environmental Management**, v. 91, n. 4, p. 852–860, 2010.
- REID, R. D.; SANDERS, N. R. **Operations Management An Integrated Approach**. 5. ed. Nova Jersey: John Wiley & Sons, 2013.
- REIMANN, C. et al. **Statistical Data Analysis Explained: Applied Environmental Statistics with R**. Chichester, UK: John Wiley & Sons, Ltd, 2008.
- RIUS, A.; CALLAO, M. P.; RIUS, F. X. Multivariate Statistical Process Control Applied to Sulfate Determination by Sequential Injection Analysis. **The Analyst**, v. 122, n. 8, p. 737–741, 1997.
- ROCHA, J. L. S. **Mapeamento por SIG e random forest do impacto dos resíduos da suinocultura na qualidade das águas da bacia do rio Piranga-MG**. 2016. 140 f. Tese (Doctor Scientiae) – Programa de Pós Graduação em Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa. 2016.
- RODE, M. et al. Sensors in the Stream: The High-Frequency Wave of the Present. **Environmental Science and Technology**, v. 50, n. 19, p. 10297–10307, 2016.
- RUSSELL, E. L.; CHIANG, L. H.; BRAATZ, R. D. **Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes**. London: Springer London, 2000.
- SABINO, C. V. S.; LAGE, L. V.; DE NORONHA, C. V. Variação sazonal e temporal da qualidade das águas em um ponto do Córrego Gameleiras usando técnicas quimiométricas robustas. **Engenharia Sanitária e Ambiental**, v. 22, n. 5, p. 969–983, 2017.
- DANTAS, M. S. et al. Impact of fecal contamination on surface water quality in the São Francisco River hydrographic basin in Minas Gerais, Brazil. **Journal of Water and Health**, v. 18, n. 1, p. 48–59, 2020.
- SAMSUDIN, M. S. et al. River water quality assessment using APCS-MLR and statistical process control in Johor River Basin, Malaysia. **International Journal of Advanced and Applied Sciences**, v. 4, n. 8, p. 84–97, 2017.

SANCHO, J. et al. Study of Water Quality in a Spanish River Based on Statistical Process Control and Functional Data Analysis. **Mathematical Geosciences**, v. 48, n. 2, p. 163–186, 2016.

SCHUWIRTH, N. et al. Multi-criteria decision analysis for integrated water quality assessment and management support. **Water Research X**, v. 1, p. 1–11, 2018.

SERGEANT, C. J. et al. 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.

SHARMA, S. **Applied Multivariate Techniques**. 1. ed. New York: John Wiley & Sons, Inc., 1996.

SHAW, P. J. A. **Multivariate Statistics for the Environmental Sciences**. 1 ed. London: Hodder Education, 2003.

SHEATS, R. D.; PANKRATZ, V. S. Understanding distributions and data types. **Seminars in Orthodontics**, v. 8, n. 2, p. 62–66, 2002.

SHI, B. et al. Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. **Science of the Total Environment**, v. 610–611, p. 1390–1399, 2018.

SHRESTHA, S.; KAZAMA, F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. **Environmental Modelling & Software**, v. 22, n. 4, p. 464–475, 2007.

SILVA, M. M. A. P. M.; FARIA, S. D.; MOURA, P. M. Modelagem da qualidade da água na bacia hidrográfica do Rio Piracicaba (MG). **Engenharia Sanitaria e Ambiental**, v. 22, n. 1, p. 133–143, 2017.

SIMEONOV, V. et al. Danube river water data modelling by multivariate data analysis. **Mikrochimica Acta**, v. 137, n. 3–4, p. 243–248, 2001.

SIMEONOV, V. et al. Environmetric modeling and interpretation of river water monitoring data. **Analytical and Bioanalytical Chemistry**, v. 374, n. 5, p. 898–905, 2002.

SIMEONOV, V. et al. Assessment of the surface water quality in Northern Greece. **Water Research**, v. 37, n. 17, p. 4119–4124, 2003.

SINGH, G. et al. Assessment of spatial and temporal variations in water quality by the application of multivariate statistical methods in the Kali River, Uttar Pradesh, India. **Environmental Monitoring and Assessment**, v. 192, n. 6, p. 394, 2020.

SINGH, K. P. et al. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. **Water Research**, v. 38, n. 18, p. 3980–3992, 2004.

SINGH, K. P. et al. Artificial neural network modeling of the river water quality—A case study. **Ecological Modelling**, v. 220, n. 6, p. 888–895, 2009.

SINGH, K. P.; MALIK, A.; SINHA, S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. **Analytica Chimica Acta**, v. 538, n. 1–2, p. 355–374, 2005.

SOARES, A. L. C.; PINTO, C. C.; OLIVEIRA, S. C. Impacts of anthropogenic activities and calculation of the relative risk of violating surface water quality standards established by environmental legislation: a case study from the Piracicaba and Paraopeba river basins, Brazil. **Environmental Science and Pollution Research**, v. 27, n. 12, p. 14085–14099, 2020.

SONG, M. W. et al. Water quality of a tributary of the Pearl River, the Beijiang, Southern China: implications from multivariate statistical analyses. **Environmental Monitoring and Assessment**, v. 172, n. 1–4, p. 589–603, 2011.

SOTOMAYOR, G.; HAMPEL, H.; VÁZQUEZ, R. F. Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. **Water Research**, v. 130, p. 353–362, 2018.

SOUZA, M. M.; GASTALDINI, M. DO C. C. Avaliação da qualidade da água em bacias hidrográficas com diferentes impactos antrópicos. **Engenharia Sanitaria e Ambiental**, v. 19, n. 3, p. 263–274, 2014.

SPEARMAN, C. “General Intelligence,” Objectively Determined and Measured. **The American Journal of Psychology**, v. 15, n. 2, p. 201–293, 1904.

SPRAGUE, L. A. et al. Assessing water-quality changes in US rivers at multiple geographic scales using results from probabilistic and targeted monitoring. **Environmental Monitoring and Assessment**, v. 191, n. 6, p. 348, 2019.

TAOUFIK, G.; KHOUNI, I.; GHRABI, A. Assessment of physico-chemical and microbiological surface water quality using multivariate statistical techniques: a case study of the Wadi El-Bey River, Tunisia. **Arabian Journal of Geosciences**, v. 10, n. 7, p. 181, 2017.

TIAN, Y. et al. Using a water quality index to assess the water quality of the upper and middle streams of the Luanhe River, northern China. **Science of the Total Environment**, v. 667, p. 142–151, 2019.

TIBEBE, D. et al. Investigation of spatio-temporal variations of selected water quality parameters and trophic status of Lake Tana for sustainable management, Ethiopia. **Microchemical Journal**, v. 148, p. 374–384, 2019.

TOLEDO, L. G. DE; NICOLELLA, G. Índice de qualidade de água em microbacia sob uso agrícola e urbano. **Scientia Agricola**, v. 59, n. 1, p. 181–186, 2002.

TRINDADE, A. L. C. et al. Tendências temporais e espaciais da qualidade das águas superficiais da sub-bacia do Rio das Velhas, estado de Minas Gerais. **Engenharia Sanitaria e Ambiental**, v. 22, n. 1, p. 13–24, 2017.

UNITED NATIONS ENVIRONMENT PROGRAMME - UNEP. **Status report on the application of integrated approaches to water resources management**. United Nations Environment Programme, 2012.

UNITED NATIONS ENVIRONMENT PROGRAMME - UNEP. **A Snapshot of the World’s Water Quality: Towards a global assessment**. Nairobi, Kenya: United Nations Environment Programme, 2016.

UNITED NATIONS ENVIRONMENT PROGRAMME - UNEP; GLOBAL ENVIRONMENT MONITORING SYSTEM/WATER PROGRAMME - GEMS WATER. **Water Quality for Ecosystem and Human Health**. 2 ed. United Nations Environment Programme Global Environment Monitoring System/Water Programme, 2008.

- VALLE, S.; LI, W.; QIN, S. J. Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods. **Industrial & Engineering Chemistry Research**, v. 38, n. 11, p. 4389–4401, 1999.
- VAN DEN KERKHOFF, P. et al. Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control. **Chemical Engineering Science**, v. 104, p. 285–293, 2013.
- VEGA, M. et al. 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.
- VON SPERLING, E.; JARDIM, F. A. **Influence of climatic conditions on cyanobacteria blooms in a tropical water supply river (Rio das Velhas, Brazil)**. Proceedings of the 34th WEDC International Conference. **Anais...Addis Ababa, Ethiopia: WEDC, Loughborough University**, 2009.
- VON SPERLING, M.; VERBYLA, M. E.; OLIVEIRA, S. M. A. C. **Assessment of Treatment Plant Performance and Water Quality Data: A Guide for Students, Researchers and Practitioners**. 1. ed. [s.l.] IWA Publishing, 2020.
- 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.
- WESTERHUIS, J. A.; GURDEN, S. P.; SMILDE, A. K. Generalized contribution plots in multivariate statistical process monitoring. **Chemometrics and Intelligent Laboratory Systems**, v. 51, n. 1, p. 95–114, 2000.
- WISE, B. M. et al. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. **Journal of Chemometrics**, v. 13, n. 3–4, p. 379–396, 1999.
- WUNDERLIN, D. A. et al. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba–Argentina). **Water Research**, v. 35, n. 12, p. 2881–2894, 2001.
- YAN, H.; ZOU, Z.; WANG, H. Adaptive neuro fuzzy inference system for classification of water quality status. **Journal of Environmental Sciences**, v. 22, n. 12, p. 1891–1896, 2010.
- ZAVAREH, M.; MAGGIONI, V. Application of rough set theory to water quality analysis: A case study. **Data**, v. 3, n. 4, 2018.
- ZEINALZADEH, K.; REZAEI, E. Determining spatial and temporal changes of surface water quality using principal component analysis. **Journal of Hydrology: Regional Studies**, v. 13, p. 1–10, 2017.
- ZHANG, H. et al. Water quality assessment and pollution source apportionment using multi-statistic and APCS-MLR modeling techniques in Min River Basin, China. **Environmental Science and Pollution Research**, v. 27, n. 33, p. 41987–42000, 2020.
- ZHANG, S.; XIN, X. Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm. **Applied Water Science**, v. 7, n. 4, p. 1955–1963, 2017.

- ZHANG, X. et al. Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. **Environmental Monitoring and Assessment**, v. 173, n. 1–4, p. 17–27, 2011.
- ZHANG, Z. et al. 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.
- ZHAO, Y. et al. Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. **Procedia Environmental Sciences**, v. 13, n. 2011, p. 1213–1226, 2012.
- ZHOU, F.; LIU, Y.; GUO, H. Application of Multivariate Statistical Methods to Water Quality Assessment of the Watercourses in Northwestern New Territories, Hong Kong. **Environmental Monitoring and Assessment**, v. 132, n. 1–3, p. 1–13, 2007.
- ZHOU, W. et al. Application of water quality control charts to spring monitoring in karst terranes. **Environmental Geology**, v. 53, n. 6, p. 1311–1321, 2008.
- ZIMMER, M. A. et al. Temporal variability in nitrate-discharge relationships in large rivers as revealed by high-frequency data. **Water Resources Research**, v. 55, n. 2, p. 973–989, 2019.
- ZIMMERMAN, S. M. et al. 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.

ANEXO 1

No Estado de Minas Gerais, a legislação estadual para a qualidade da água é a Deliberação Normativa nº 01, de 05 de maio de 2008 (DN nº 01/2008), estabelecida pelo Conselho Estadual de Meio Ambiente e Conselho Estadual de Recursos Hídricos (COPAM/CERH-MG 2008) (Tabela A1).

Nesta legislação, os cursos de água interiores são classificados em cinco classes, cada uma com limites para os parâmetros de qualidade da água que devem ser alcançados para manter a qualidade desses cursos de água compatíveis com os usos locais (como consumo humano, irrigação, navegação ou diluição de efluentes).

Tabela A1 - Padrões de qualidade de água superficial definidos na Deliberação Normativa COPAM/CERH-MG nº 01/2008.

Parâmetro	Unidade	Classes			
		1	2	3	4
Arsênio total	mg.L ⁻¹ As	0,01	0,01	0,033	-
Cloreto	mg.L ⁻¹ Cl	250	250	250	-
Condutividade elétrica <i>in loco</i>	µS.cm ⁻¹	-	-	-	-
Coliformes termotolerantes	NMP ^a /100 ml	200	1000	4000	
Demanda bioquímica de oxigênio	mg.L ⁻¹ O ₂	3	5	10	
Fósforo total	mg.L ⁻¹ P	0,1	0,1	0,15	-
Nitrato	mg.L ⁻¹ N	10	10	10	
Oxigênio dissolvido	mg.L ⁻¹ O ₂	>6	>5	>4	>2
pH <i>in loco</i>	-	6 a 9	6 a 9	6 a 9	6 a 9
Sólidos total	mg.L ⁻¹	-	-	-	-
Temperatura da água	°C	-	-	-	-
Turbidez	UNT	40	100	100	-

^a Número mais provável.